THE UNIVERSITY OF MELBOURNE

Department of Economics

Working Paper Series

A Theory of Rational Jurisprudence

Scott Baker and Claudio Mezzetti

May 2012

Research Paper Number 1144

ISSN: 0819-2642
ISBN: 978 0 7340 4494 5

Department of Economics
The University of Melbourne
Parkville VIC 3010
www.economics.unimelb.edu.au

A Theory of Rational Jurisprudence*

Scott Baker † and Claudio Mezzetti ‡

† Washington University in St. Louis
‡ University of Melbourne

May 7, 2012

Abstract

We examine a dynamic model of up-or-down problem solving. A decision maker can either spend resources investigating a new problem before deciding what to do, or decide based on similarity with precedent problems. Over time, a decision making framework, or jurisprudence, develops. We focus on the model's application to judge-made law. We show that judges summarily apply precedent in some cases. The law may converge to efficient or inefficient rules. With positive probability, identical cases are treated differently. As the court learns over time, inconsistencies become less likely. We discuss the existing empirical evidence and the model's testable implications.

Keywords: Law and Economics, Incompleteness of Law, Judge-Made Law, Evolution of Legal Rules.

JEL Classification Numbers: K10, K40.

1 Introduction

We develop a dynamic model of up-or-down problem solving. In it, a decision maker must devise a method for dealing with a series of problems, project proposals, or cases, with an up-or-down, yes-or-no, answer. As matters arise, the decision maker has a choice: He can spend time and resources investigating the merits in depth and, based on the results, decide what to

*We would like to thank the associate editor and two anonymous referees; their insightful comments have led to a substantial improvement of the paper. We also thank participants at workshops at Tilburg University, the University of Chicago, the University of Illinois, Northwestern University, Boston University, USC, the University of Amsterdam, Washington University, AETW 2012 at UNSW, the 2010 American Law and Economics Association Meetings, and the 2009 Triangle Workshop on Law and Economics.

do. Alternatively, the decision maker can decide that a case or project is "close enough" to prior cases or projects that it can be decided similarly without additional thought. Reasoning in this way saves on decision costs but creates the chance of error. Over time the decision maker develops a decision making framework, a jurisprudence. The paper focuses on the properties of this jurisprudence: Will a practice of following precedent emerge? Will "like" cases or projects be treated "alike"? Under what conditions will the case law converge to efficiency – where a party wins or a project is approved if and only if the benefits outweigh the costs?

The problem solving we model is something many organizations must do. CEOs and venture capitalists must decide which projects to move ahead and which projects to forgo; at the same time, they must decide how to spend limited resources deciding which projects go in which bin. In handling customer complaints, firms must decide which complaints should be investigated and which ones should be declined or approved based solely on similarity to past complaints.

Although we believe there are many applications, our primary motivation is the study of judge-made law, a subject rarely studied by economists.[1] In common law countries, judicial decisions set the rules for property, govern the interpretation of contracts, and control negative externalities through the allocation of liability. Further, because ex-ante specification of statutes is costly, legislatures often use statutory language which includes broad phrases or terms. In antitrust, for example, Section 2 of the Sherman Act imposes liability on "every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize."[2] In intellectual property, a patent will not issue if the difference between the prior art and the claimed invention would have been "obvious" at the time of the invention.[3] Courts, then, fill in the law on a case by case basis.

In the United States, constitutional law is judge-made. As a result, judicial decisions set the ground rules for the political process and define the state's reach into private lives. Even in areas of law where agency regulation dominates (say, environmental law and occupational health and safety) judges play a significant oversight role. For example, judges review an agency's cost-benefit analysis and can, in some scenarios, decide whether the agency must do cost-benefit analysis at all (Sunstein, 2001).

Case by case adjudication differs from ex ante legislative or agency rule making in fundamental ways. Unlike other law-making bodies, judges don't announce policies. They are

[1] For a recent paper outlining several reasons why economists need to pay closer attention to judicial behavior, see Stephenson (2009). An enhanced focus on judicial decision-making and common law evolution is also critical in light of the empirical findings of the legal origins literature (see La Porta et al., 1998, and La Porta et al., 2004).

[2] 15 U.S.C. § 2.

[3] 35 U.S.C. § 103.

reactive. Generally speaking, judges don't respond unless an injury materializes.[4] The injury leads to a complaint or case, with each case providing grounds for revising the judge's belief about the benefits and costs of a particular judge-made law. In deciding cases, judges have to consider that careful examination of a case is costly. They need to select which cases are important to look at closely and which cases are not. One lesson from our model is that this triage process – an important act in judging – has consequences for the ultimate policy and the consistency of judicial decision making along the way.

Just because case by case decision making is different from ex-ante direct regulation of behavior doesn't make it important. But there is an interaction between the effectiveness of judicial decision making and the legislature's choice to delegate – via a vague statute – to the courts. Detailed statutes are expensive to draft and inflexible. If judicial decision making is likely to be poor or mistaken, the legislature should be more willing to spend resources setting the contours of the statute in the first place, and vice versa.

Judicial decision making is likely to be poor when the court must spend lots of resources figuring out the correct answer to the resolution of a case.[5] This might happen in complex areas of the law, such as tax. In addition, when judicial caseload is high, courts will react to additional delegation (passage of another statute with vague language) by reading prior cases "broadly." The judge will extrapolate a great deal from prior cases to new cases with different facts; new cases where the outcome of the prior case is not terribly informative. As a result, the court will make more errors, simply as a consequence of having to adjudicate all cases on its docket.

On the other hand, case by case adjudication is well suited to uncover unintended side-effects. By observing materialized injuries, judges learn the extent of the injury before setting the doctrine. Judges can thus incorporate this information in the doctrine, adjusting the rules over time as more information comes available.

To be concrete, zoning makes sure everyone knows what they can and cannot do with their property ex ante and saves on litigation costs. Regulation via nuisance claims, by contrast, allows for ex-post considerations and, thus, for better tailoring of the regulation to situation specific circumstances. Which is better or should we have a little of both? In short, the choice of the degree of ex-post versus ex-ante regulation turns on how good judicial decision making is likely to be. The answer to that question requires an understanding of how judges make decisions with scarce resources, a decision making exercise which forms the basis of our inquiry.

[4] For an informal discussion of the factors making legal intervention before the harm relatively more efficient than legal intervention after the harm, see Shavell (2004, 572-578).

[5] For a model relating judicial errors and the desirability of ex ante regulation, see Schwartzstein and Shleifer (2011).

We do not model litigation explicitly. Instead, we focus on learning over time, having judges engage in a forward looking cost-benefit analysis in interpreting precedent, and what that analysis implies for the evolution of law. For simplicity, we assume that judges discover the facts of a case on their own. This is not incompatible with the typical role of advocacy. Indeed, we could generalize the model so as to let, say, the defendant's lawyers argue that the relevant set of precedents is the one leading to dismissal of the case, while the plaintiff's lawyers argue that another set of precedents govern, leading to the opposite conclusion. Such an extension would shed light on the role of advocacy and case selection, but it would not affect our conclusions on the evolution of law. When the arguments of one party are convincing, the judge would rule in his favor based on closeness with the set of precedents the party has presented; when a definite conclusion is unclear, the judge would invest additional resources to discover which sets of precedents govern.

To our knowledge, no one has formally examined situations where argument strength turns on distance. Yet, in judge-made law, that distance is critical. Our model demonstrates a link between "closeness" and the efficient deployment of resources. As resources become more scarce, a prior case can be further away from the case at hand and still persuade the decision maker that he should decide the case without additional investigation.

Using a dynamic programming model, we build a theory of judge-made law from the ground up. We assume a single court that lives forever, consisting of judges with identical policy preferences. The judges are initially uncertain about the consequences of legal rules and have scarce resources to investigate cases. The theory yields both surprising and intuitive results.

In the model, reliance on precedent arises endogenously. Judges follow precedent not because deviations are punished, but to conserve resources. Each period, the judge interprets the prior cases, deciding how far to extrapolate the results from prior cases to new and different circumstances, without engaging in any fresh investigation. Cases that are in some sense "close" to the prior case law are decided summarily based solely on the precedent. Cases that are "far" the judge investigates. Each period, the judge decides what is close and what is far.

In making this interpretative choice, the judge balances two costs. First, there are error costs, i.e., the costs of ruling on a case incorrectly. Second, there are decision costs, the costs associated with a judge investigating a case in depth instead of relying on the precedent as a proxy for what to do. We show that the optimal balancing of these costs formalizes the usefulness of reasoning by analogy (a skill taught to every lawyer and judge) and, in accord with intuition, shows that reliance on precedent will be lower in areas of law where errors are more costly.

We next demonstrate that the judge-made law will, in general, converge. This convergence will be of two types. If decision costs are small relative to fixed error costs, the law will converge to the efficient outcome, or correct decision, in all cases. More interesting, if decision costs are high relative to fixed error costs, the law will converge to an inefficient set of legal rules. Thus, for example, judge-made law will incorrectly specify liability for some activities where no liability is the proper result and no liability where liability is the right outcome. This result obtains even though all judges share efficiency as the goal. Convergence here is second-best: spending resources to gather more information – i.e., hearing more cases – is not worthwhile in terms of the benefit of a more accurate legal rule.

Finally, we explore the implications of our theory for the evolution of law. In spite of our assumption of identical policy preferences by all judges, we show that with positive probability the court will fail to treat like cases alike. Cases with "identical" relevant facts may be decided differently. Discriminatory treatment – violation of what we refer to as the likeness principle – occurs as the court uses what it has learned to improve the law. Judges are often vilified for treating like cases differently, actions thought unfair and inconsistent with the rule of law. The model shows that strict adherence to the likeness principle inhibits judicial learning and the cost-justified updating of legal rules. Violation of the likeness principle is apt to occur when investigation of a case yields a surprising result, and hence teaches a great deal to the court. We also consider whether inconsistencies will remain in the limit. The model demonstrates that as the court learns more and more, inconsistencies become less and less likely and, in the limit, vanish. Taken together, the two results suggests inconsistencies should be observed more often when the judge-made law is in its infancy, shortly, say, after Congress passes a statute containing broad, enabling language.

The paper unfolds as follows: Related literature is reviewed in the remaining part of this introduction. Section 2 develops an economic model of legal reasoning. Section 3 demonstrates that precedent has value and studies the convergence properties of doctrine. Section 4 shows that a rational court will, with positive probability, violate the likeness principle. Section 5 suggests possible extensions and offers some concluding remarks. Throughout, where applicable, we discuss the empirical evidence consistent with the predictions of the model and offer testable implications. Although we believe the model covers a common decision making phenomenon, one benefit of focusing on judge-made law is that courts provide a wealth of data for possible future empirical tests of the theory.

Literature Review. The model closest to ours is Gennaioli and Shleifer (2007). Seeking an explanation for the empirical finding that a common-law legal origin correlates with various markers of development, they build a model of judge-made law. A judge, in their model, cannot overrule prior cases, she can only distinguish a case by searching for a different dimension along which to consider it. The act of distinguishing two cases has social value because it embeds new information into the law. Different judicial policy preferences then shape the evolution of law. Gennaioli and Shleifer's main result is the "Cardozo Theorem." It says that the legal evolution induced by distinguishing cases will, on average, be beneficial irrespective of the amount of bias in the judiciary. We take a different approach. Rather than starting from the premise that judges have conflicting policy preferences, we begin from an assumption of scarce judicial resources. From this alternative baseline, we get insights consistent with the institutional features of judge-made law.

Two other significant literatures relate to our work. The first is from law and economics scholars, the second from political science. Since Judge Posner's assertion that the common law is efficient (Posner, [1973] 2007), the law and economics literature has sought to explain why this might be so (Rubin, 1977; Priest, 1977; Bailey and Rubin, 1994; Hylton, 2006). Both Judge Posner's assertion about efficient common law and the models exploring it have been sharply contested (Bailey and Rubin, 1994; Hadfield, 1992; Hathaway, 2001). The literature has blossomed with many factors pointing toward and against efficiency (Zwyicki, 2003; Klerman, 2007; Parisi and Fon, 2008). This literature is distinct from what we do here. Case selection drives the law in these models, with judges playing little role. Ours is not a story about litigants selecting specific cases for trial and that selection dictating the path of the law. Rather, we ask what evolution to expect when judges are forward looking and have the ability to learn, but are resource constrained.

There are a few exceptions to the litigant selection story about the evolution of law. Cooter et al. (1977) and Hadfield (2009) develop models where the court can learn and ask whether rules will adapt to new circumstances and/or converge to efficiency. Unlike our model, the question of how to optimally deploy judicial resources over many periods is not examined. Dari-Mattiacci et al. (2010) develop a dynamic model where the litigants bring information to the courts and the courts issue decisions. The number of judicial decisions implies more "precedent"; distinct from us, judges don't interpret prior case law in their model.

Political scientists assume that judges, like legislators, make decisions to advance their preferred policy objective.[6] Learning doesn't occur and the informational value of precedent is not explored. In contrast, we show how the shifting interpretation of precedent can be seen as a method of efficiently managing resources to learn about the proper structure of legal rules.

[6] For a model where a judge makes decisions anticipating the likely position of Congress or the executive, see Eskridge and Ferejohn (1992). For a model where the judge is influenced by other judges sitting on the panel, see Spitzer and Talley (2011). For a model where judges face constraints imposed by the likely position of the higher court, see Songer et al. (1994) and McNollgast (1995). For a model where judges interact repeatedly over time, see O'Hara (1993).

These aspects of judicial behavior have not, to our knowledge, been formally studied elsewhere.

2 A Model of Legal Reasoning

In creating law case by case, judges mix the information from prior cases with the facts from new cases. If the new facts indicate that, as stated, the legal doctrine no longer serves its function, the judge can distinguish the prior case law and reach an alternative resolution.[7] To capture this process, suppose a judge wishes to regulate a set of activities, $x \in [0, 1]$. Activity x carries costs and benefits and there is a threshold $\theta \in [0, 1]$, below which an activity is socially valuable and above which it is not. The threshold point θ is initially unknown. We model it as random variable distributed according to $F(\theta)$ with positive density $f(\theta)$ on $[0, 1]$. The restriction of activities to the unit interval is with some loss of generality. At a higher level of abstraction, an activity might be represented as an element in a multi-dimensional space.[8] Our assumption requires that each multi-dimensional activity can be reduced to a single number and ranked against all other activities (i.e., the social value of activity 1 is lower than that of activity 3/4 which is lower than that of activity 1/2). It is akin to the common assumptions that guarantee the existence of a utility function in consumer theory.[9] Given this ordering, finding the set of socially valuable, or efficient, activities reduces to locating θ, the activity whose social value is zero.

As noted in the introduction, the court consists of a single, infinitely-lived judge with constant preferences.[10] The judge prefers to allow beneficial activities and deter harmful ones. When presented a case, the judge can either fully investigate or summarily examine it. "Summarily" in this model means spending less effort on the case, by, for example, resolving the case with an unpublished decision or one line order.

Full investigation costs C and is rewarded: the judge discovers whether undertaking activity x is, or is not, efficient; that is, whether it carries a positive or negative social value.[11,12] The

[7] Levi (1949, 7-19) contains the classic discussion of common law reasoning.

[8] The assumption that the region of uncertainty is bounded is not important. It will be used when we work with a uniform distribution, but could be dispensed with in the general model.

[9] Completeness, transitivity and continuity are sufficient; they guarantee that activities associated with nearby real numbers have nearby characteristics from the point of view of case law. Since judges have identical preferences, they collapse the fact space into a single dimension in the same way and rank the resulting bundles the same.

[10] The conclusion suggests an alternative interpretation of the model, consisting of two courts in a hierachy, where the upper court must decide how many cases whose outcome is uncertain to delegate to the lower court to resolve.

[11] It simplifies the analysis to assume that there is no noise or mistake in the discovery process, and that the judge correctly determines whether an activity is efficient upon its full investigation.

[12] The assumption that decision costs are constant seems a reasonable first approximation; it is not immediate whether investigation costs are lower or higher the closer is x to θ.

judge reports what he has learned in an opinion. Summary examination, by contrast, is costless, but might result in an incorrect decision.

Full investigation means studying the facts. It requires the court and the litigants to spend time and resources uncovering the relationship between the new facts and broader social policies, using all the judicial tools (oral argument, additional briefing, reading the scholarly literature, closely examining expert reports, etc.). Investigation costs are likely to be high in technical fields, like tax, environmental harms, and food and drug safety. In these areas, the parties must provide and the judge must assess complicated testimony and materials to see the likely consequences of ruling for the plaintiff or the defendant. Overall caseload also impacts decision costs. As caseloads go up, the judge's opportunity cost of investigating any single case increases. As we shall see, using caseload as a proxy for decision costs enables us to derive testable predictions from the model and link the theoretical results to the empirical literature on judicial decision making.

Two types of potential mistakes are associated with summary examination. The court might erroneously declare a socially beneficial activity impermissible or a socially harmful activity permissible. Mistakes in adjudication result in two kinds of losses. There is a variable loss $\ell\,|x-\theta|$ which depends on the distance between the activity and θ, and a fixed error loss L, which does not depend on the distance. The variable loss captures the idea that the further away the activity is from the threshold θ the higher is the social loss from a mistaken decision. The fixed error loss captures a discontinuity in the social value of activities. Judges care about perception, both from the public and from the other branches. In many cases, the public, for example, is unlikely to understand the difference between a "small" mistake and a "big" mistake and calibrate their reaction to the court decision accordingly. As a result, one might think about the fixed error cost as the discrete loss arising from the court making the wrong call.[13]

Error costs are likely to be high when other actors lack the capacity or will to correct judicial mistakes. Three examples are: constitutional cases, when the courts have the final say about what the constitution requires; cases when interest group conflict is rampant and, as a result, Congress rarely acts; courts decisions on the scope of a mandatory rule, that is, a rule that cannot be contracted around.

[13] Nothing in the model turns on including a fixed cost. Assuming that there is a fixed social loss from error stacks the deck in favor of efficient convergence. As noted above, a major claim in the law and economics literature is that judge-made law is efficient. Inefficiency is easily shown if the only social loss of incorrect decisions is the variable loss, which shrinks as activities get closer to θ. With a constant decision cost, investigation would, over time, become less and less worthwhile (the benefit of preventing mistakes would fall, while the cost remains constant). In the long run, judge-made law would surely admit errors. Interestingly, as we shall see, with a fixed error cost component convergence to inefficient rules still can occur.

2.1 Timing and The Construction of Precedent

Each period a case x is randomly selected from the interval $[0, 1]$ according to the distribution $G(x)$ with positive density $g(x)$ and brought to the attention of the court.[14] Let W_t denote the highest activity or fact pattern that the court has, as of time t, heard, fully investigated, and found acceptable, that is, of positive social value. Let R_t denote the lowest activity or fact pattern that the court has fully investigated and found unacceptable, that is, of negative social value. Prior case law, in other words, has taught that allowing activities in the interval $[0, W_t]$ is efficient, while allowing activities in $[R_t, 1]$ is inefficient. The two endpoints W_t and R_t squeeze the court's beliefs about the distribution of θ. The range of activities the court knows nothing about is (W_t, R_t).

Each period the court decides, based on prior precedent (i.e., looking at W_t and R_t), which cases to look closely at and which cases to decide summarily. Cases that are in some sense "close" to the prior precedent are decided summarily based on the precedent alone. Cases that are in some sense "far" from prior precedent are investigated. The court can easily determine the proper resolution of cases below W_t and above R_t, since the prior precedent is perfectly informative about them. The court summarily declares permissible any case below W_t and impermissible any case above R_t. On these cases, the judge expends no decision costs and makes no errors. For cases in the interval (W_t, R_t), the court determines how broadly or narrowly to construe past precedent. Reading precedent broadly means that the court decides a high percentage of new cases by extrapolating or reasoning by analogy from past cases without further inquiry (that is, lots of cases are determined to be "close" to the precedent). Given the greater extrapolation, the chance is higher that the prior cases will be off point and lead the judge to a mistaken resolution. Reading precedent narrowly is the opposite.

The interpretative decision determines two bounds, a_t and b_t. These bounds partition the interval of uncertainty (W_t, R_t) into three areas. The first area is the interval $[a_t, b_t]$. In this interval, the judge fully investigates the case. The second and third areas are (W_t, a_t) and (b_t, R_t). If a case lies in either of these intervals, the judge feels the activities are close enough to prior case law (W_t and R_t) as to be decided by application of precedent alone without spending effort. Activities in (W_t, a_t) are declared permissible, activities in (b_t, R_t) are declared impermissible. The size of these two intervals measures of how expansively the judge reads the prior precedent

If the judge investigates and learns that the activity in the case has positive value, she

[14] In specifying that the court always draws facts from the same distribution, we abstract away from the law's impact on behavior. We do this to ease the analysis and focus on judicial learning. The assumption is a reasonable first approximation, so long as parties make mistakes about the contours of the law when deciding their actions, or face a small probability of getting caught and sued.

writes an opinion and updates the precedent stock ($W_{t+1} > W_t$). Likewise, if the judge investigates and finds the activity to have negative value, she updates the precedent stock accordingly ($R_{t+1} < R_t$). If the new case is summarily examined, the court learns nothing and the precedent stock remain constant.

As an example motivating our model, consider punitive damage awards in tort cases. The Supreme Court has the power to strike down punitive awards set by lower courts as violating the due process clause of the Constitution.[15] The legal issue is how much is too much – what multiple of compensatory damages renders the award unconstitutional. Suppose awards range between 0 and 1000 times the compensatory award. Imagine the first case the Supreme Court hears involves a punitive award of 500 times the compensatory award. The Court investigates, finds this multiple too much and strikes the award down. In the language of the model, the Court learns that $\theta < \frac{500}{1000} = .5$. It writes an opinion saying so, setting $R_t = .5$.

Suppose that the next case involves a multiple of 100. The Court might interpret its prior case to also cover this case and summarily strike down the award (setting $b_t = .1$). In so doing, the Court broadly construes its own precedent. Such a move saves resources, but might be wrong. If the ideal rule θ lies in the interval $(.1, .5)$, the court will have made a mistake. Now suppose instead that the second case has a multiple of 450. Striking this award down summarily is less likely to generate a mistake. It only does so if θ happens to be in the interval $(.45, .5)$. Because of the "closeness" between the 450 case and the 500 precedential case, relying on precedent saves resources without unduly increasing error costs. In our model, in each period the judge interprets precedent to balance the two costs, while understanding that the investigation today benefits future judges because it reveals information about the optimal legal rule.[16]

2.2 The Optimization Problem

We can now write the dynamic optimization problem that the court must solve. Let δ be the discount factor and $V(W_t, R_t)$ be the court's value function at time t, with state variables W_t, R_t. The court chooses the interpretative bounds a_t, b_t subject to $W_t \leq a_t \leq b_t \leq R_t$ to

[15] See Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1 (1991).

[16] This example is drawn from the case law. See Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 23 (1991) (finding that a ratio of four to one between punitive damages and compensatory damages was not excessive); TXO Prod. Corp. v. Alliance Resource Corp., 509 U.S. 443, 462 (1993) (noting that a 10 to 1 ratio between the punitive award and the harm if the illicit plan had succeeded would not "jar one's constitutional sensibilities"); B.M.W., Inc. v. Gore, 517 U.S. 559, 582 (1996) (finding that an award of 500 times the actual harm was grossly excessive); State Farm Mutual Automobile Ins. Co. v. Campbell, 538 U.S. 408, 429 (2003) (finding that an award of 145 to 1 was an "irrational and arbitrary deprivation of the property of the defendant"). For other examples of this kind of evolution of law, see Niblett (2010).

maximize its expected discounted payoff:

$$
\begin{aligned}
V(W_t, R_t) = \max_{W_t \le a_t \le b_t \le R_t} \Big\{ & \delta E_t V(W_{t+1}, R_{t+1}) - C\left[G(b_t) - G(a_t)\right] \\
& - \int_{W_t}^{a_t} \int_{W_t}^{x_t} \frac{Lf(\theta)}{F(R_t) - F(W_t)} d\theta g(x_t) dx_t - \int_{b_t}^{R_t} \int_{x_t}^{R_t} \frac{Lf(\theta)}{F(R_t) - F(W_t)} d\theta g(x_t) dx_t \\
& - \int_{W_t}^{a_t} \int_{W_t}^{x_t} \frac{\ell\left(x_t - \theta\right) f(\theta)}{F(R_t) - F(W_t)} d\theta g(x_t) dx_t - \int_{b_t}^{R_t} \int_{x_t}^{R_t} \frac{\ell\left(\theta - x_t\right) f(\theta)}{F(R_t) - F(W_t)} d\theta g(x_t) dx_t \Big\}.
\end{aligned} \tag{1}
$$

The second term is the expected cost of having to decide a case in period t. For example, if $a_t = b_t$, then the court construes the prior precedent as deciding the law for all activities and does not incur any decision costs at time t. The greater the distance between a_t and b_t, the greater the chance a case is drawn where the court views the law as unsettled by prior precedent and is willing to expend effort.

The third and fourth term reflect the expected one-period losses due to the fixed error component L, while the fifth and sixth term reflect the expected losses due to the variable error component $\ell\left|x - \theta\right|$. Consider the third and fifth term. If the judge sets $a_t \ge W_t$, there is a chance the case drawn x_t is between W_t and a_t. For cases in this gap, the court will base its decision solely on the prior precedent and rule summarily that the activity is permissible. The expression $\frac{f(\theta)}{F(R_t) - F(W_t)}$ is the probability the court attaches to the possibility that θ is less than x_t. In that event, x_t creates a social loss; the court has made a mistake in its summary ruling. The fourth and sixth term follow from a similar analysis on the upper region of the interval of uncertainty; here precedent induces the court to rule the activity as impermissible, but, in fact, it is socially valuable.

The first term in (1) is the discounted expectation of the value of the court's objective function at the end of period t, given its interpretative choices at time t. This term captures the dynamic learning considerations described above. It can be written explicitly as the sum of three components:

$$
\begin{aligned}
E_t V(\cdot) = {} & V(W_t, R_t)\left[1 - \left(G(b_t) - G(a_t)\right)\right] \\
& + \int_{a_t}^{b_t} V(W_t, x_t) \frac{F\left(x_t\right) - F(W_t)}{F(R_t) - F(W_t)} g(x_t) dx_t + \int_{a_t}^{b_t} V(x_t, R_t) \frac{F\left(R_t\right) - F(x_t)}{F(R_t) - F(W_t)} g(x_t) dx_t.
\end{aligned} \tag{2}
$$

The first component is the current value function times the probability that no learning takes place because the randomly selected activity x is outside the interval of investigation $[a_t, b_t]$. The second component is the expected value function when the case x is brought to court, investigated upon, and determined to be above θ; in such an instance the new interpretative interval becomes $[W_t, x_t]$. The third component is the expected value function when x is discovered to be below θ.

A special version of the model is when the distributions $F(\theta)$ and $G(x)$ are both uniform. This version has the advantage of simplifying the analysis; only the size $\Delta_t = R_t - W_t$ of the interval (W_t, R_t) matters to the court when deciding the interpretative bounds. It is convenient to express the objective function of the court as a function of the proportion of cases α_t in the interval Δ_t that the court considers permissible and the proportion β_t that it considers impermissible after summary investigation:

$$\begin{aligned}\alpha_t &= \frac{a_t - W_t}{\Delta_t} \\ \beta_t &= \frac{R_t - b_t}{\Delta_t}.\end{aligned}$$

Thus, the proportion of cases in Δ_t that are investigated is $1 - \alpha_t - \beta_t$. It is immediate that the following restrictions must hold: $\alpha_t \geq 0$, $\beta_t \geq 0$, $1 - \alpha_t - \beta_t \geq 0$. We will use this uniform version of the model in the remainder of the paper. The appendix shows that the main results and insights extend to the general version.

Lemma 1 in the appendix proves that the value function $V(\Delta_t)$ for the uniform version exists, is unique, continuously differentiable and can be written as:

$$\begin{aligned}V(\Delta_t) = \max_{\alpha_t \geq 0, \beta_t \geq 0, 1-\alpha_t-\beta_t \geq 0} \Bigg\{ & -C\left(1 - \alpha_t - \beta_t\right)\Delta_t - \frac{L}{2}\left(\alpha_t^2 + \beta_t^2\right)\Delta_t \\ & - \frac{\ell}{6}\left(\alpha_t^3 + \beta_t^3\right)\Delta_t^2 + \delta V(\Delta_t)\left[1 - \left(1 - \alpha_t - \beta_t\right)\Delta_t\right] \\ & + \delta \int_{\alpha_t}^{1-\alpha_t} V(\Delta_t x_t)\Delta_t x_t dx_t + \delta \int_{\beta_t}^{1-\beta_t} V(\Delta_t x_t)\Delta_t x_t dx_t \Bigg\}.\end{aligned} \tag{3}$$

Lemma 1 also shows that $V(\Delta)$ is negative and decreasing in Δ, with $V(0) = 0$.

3 The Value of Precedent and the Convergence of Doctrine

The most narrow construction of precedent in period t sets $(a_t, b_t) = (W_t, R_t)$, or $\alpha_t = \beta_t = 0$. In so doing, the court maximizes the learning that case load provides by looking at every possible case where the resolution is uncertain; each of these cases carries a bit of information about the location of θ. If the decision cost C of looking at a case is sufficiently small relative to the error losses, L and ℓ, shouldn't the court use all the information potentially available in each period, setting $\alpha_t = \beta_t = 0$? As we shall see in this section, while this approach seems like a good idea, it never is, no matter how small C is relative to L and ℓ.

Proposition 1 shows that the constraints $\alpha_t \geq 0$ and $\beta_t \geq 0$ never bind if $\Delta_t > 0$ – that is, a positive proportion of cases is always summarily examined. It also proves that the solution is symmetric, $\alpha_t = \beta_t$. The court always relies on precedent and never investigates every case in the interval of uncertainty Δ_t. On the other hand, the constraint $1 - \alpha_t - \beta_t \geq 0$ may bind; when this happens $\alpha_t = \beta_t = \frac{1}{2}$ and no cases are fully investigated. In legal terms, the court construes the prior precedent as covering all potential future cases.

Define the precedent ratio λ_t as the proportion of cases in the interval of uncertainty Δ_t that will be ruled by precedent in period t:

$$\lambda_t = \alpha_t + \beta_t.$$

As the bounds a_t and b_t get closer for a given level of uncertainty, λ_t increases, meaning that the court decides a higher proportion of cases by reference to precedent alone.

Proposition 1. *There exists $\overline{\lambda}(\Delta_t) > 0$ such that in each period t when the interval of uncertainty is Δ_t the court chooses $\alpha_t = \beta_t = \frac{\lambda_t}{2} \geq \frac{\overline{\lambda}(\Delta_t)}{2}$, and thus selects a precedent ratio λ_t bounded away from zero. The precedent ratio λ_t is an increasing function of C_t and a decreasing function of L_t, ℓ_t and δ_t.*[17]

Proof. Define

$$V^{\alpha}(\Delta_t) = -C\left(\frac{1}{2} - \alpha_t\right)\Delta_t - \frac{L}{2}\alpha_t^2\Delta_t - \frac{\ell}{6}\alpha_t^3\Delta_t^2 + \delta V(\Delta_t)\left[\frac{1}{2} - \left(\frac{1}{2} - \alpha_t\right)\Delta_t\right] + \delta\int_{\alpha_t}^{1-\alpha_t} V(\Delta_t x_t)\Delta_t x_t dx_t.$$

$$V^{\beta}(\Delta_t) = -C\left(\frac{1}{2} - \beta_t\right)\Delta_t - \frac{L}{2}\beta_t^2\Delta_t - \frac{\ell}{6}\beta_t^3\Delta_t^2 + \delta V(\Delta_t)\left[\frac{1}{2} - \left(\frac{1}{2} - \beta_t\right)\Delta_t\right] + \delta\int_{\beta_t}^{1-\beta_t} V(\Delta_t x_t)\Delta_t x_t dx_t.$$

The court's objective function (3) can be written as

$$V(\Delta_t) = \max_{\alpha_t \geq 0,\ \beta_t \geq 0,\ 1-\alpha_t-\beta_t \geq 0} V^{\alpha}(\Delta_t) + V^{\beta}(\Delta_t). \tag{4}$$

Note that $V^{\alpha}(\Delta_t)$ does not depend on β_t and $V^{\beta}(\Delta_t)$ does not depend on α_t. Moreover,

[17] In the proof, we derive analytical solutions for the comparative statics when the parameters C, L, ℓ and δ only change at time t. (This is why the subscript t is attached to the parameters in the statement of the proposition). Matlab simulations show that the sign of each comparative statics result is the same if we vary the relevant parameter at each point in time.

$V^{\beta}(\Delta_t) = V^{\alpha}(\Delta_t)|_{\alpha_t=\beta_t}$; that is, after switching the independent variable α_t with β_t, the functions $V^{\alpha}(\Delta_t)$ and $V^{\beta}(\Delta_t)$ are identical. Finally, recall that $V(\Delta)$ is negative, decreasing in Δ and note that $\lim_{\Delta\to 0} V(\Delta_t) = \lim_{\Delta\to 0} V^{\alpha}(\Delta_t) = \lim_{\Delta\to 0} V^{\beta}(\Delta_t) = 0$. Note also that:

$$\frac{\partial V^{\alpha}}{\partial \alpha_t} = \Delta_t \left[C - L\alpha_t - \frac{\ell}{2}\alpha_t^2\Delta_t + \delta V(\Delta_t) - \delta(1-\alpha_t) V(\Delta_t(1-\alpha_t)) - \delta\alpha_t V(\Delta_t\alpha_t)\right] \quad (5)$$

$$\frac{\partial V^{\beta}}{\partial \beta_t} = \Delta_t \left[C - L\beta_t - \frac{\ell}{2}\beta_t^2\Delta_t + \delta V(\Delta_t) - \delta(1-\beta_t) V(\Delta_t(1-\beta_t)) - \delta\beta_t V(\Delta_t\beta_t)\right]. \quad (6)$$

Let $\mu_t^{\alpha}, \mu_t^{\beta}$ and ρ_t be the (non-negative) multipliers on the constraints $\alpha_t \geq 0$, $\beta_t \geq 0$, and $1 - \alpha_t - \beta_t \geq 0$, respectively. The first order conditions of program (4) with respect to α_t and β_t are:

$$\frac{\partial V^{\alpha}(\Delta_t)}{\partial \alpha_t} + \mu_t^{\alpha} - \rho_t = 0 \quad (7)$$

$$\frac{\partial V^{\beta}(\Delta_t)}{\partial \beta_t} + \mu_t^{\beta} - \rho_t = 0 \quad (8)$$

We now show that α_t and β_t (and hence λ_t) are bounded away from zero. Suppose that $\alpha_t = 0$. Then (7) and (5) imply

$$0 < \Delta_t C \leq \Delta_t C + \mu_t^{\alpha} = \rho_t.$$

By complementary slackness, the constraint $1 - \alpha_t - \beta_t \geq 0$ must bind and hence $\beta_t = 1$. As a consequence, it is $\mu^{\beta} = 0$ and, using (8) and (6):

$$\Delta_t \left[C - L - \frac{\ell}{2}\Delta_t\right] = \rho_t,$$

a contradiction since $\mu_t^{\alpha} \geq 0$. This shows that α_t cannot be zero. Indeed, since $\lim_{\alpha\to 0} \frac{\partial V^{\alpha}}{\partial \alpha_t} = \Delta_t C$, for α_t "arbitrarily close" to zero it must be $\rho_t \simeq \Delta_t C > 0$ and $\beta_t = 1 - \alpha_t$. In addition, it must be $\rho_t \simeq \lim_{\beta\to 1} \frac{\partial V^{\beta}}{\partial \beta_t} = \Delta_t \left[C - L - \frac{\ell}{2}\Delta_t\right]$. Since this is a contradiction, α_t must be bounded away from zero; that is, there must be a lower bound $\overline{\alpha}(\Delta_t) > 0$ such that $\alpha_t > \overline{\alpha}(\Delta_t)$. The argument that β_t is bounded away from zero is analogous.

We now show that it must be $\alpha_t = \beta_t$ (this also implies $\lambda_t > \overline{\lambda}(\Delta_t) = 2\overline{\alpha}(\Delta_t)$). We have already seen that the constraints $\alpha_t \geq 0$ and $\beta_t \geq 0$ are slack at the solution of the maximization problem in the right hand side of (4). Since $V^{\alpha}(\Delta_t)$ depends only on α_t and $V^{\beta}(\Delta_t)$ depends only on β_t, if the constraint $1 - \alpha_t - \beta_t \geq 0$ is also slack, then (4) can be written as

$$V(\Delta_t) = \max_{\alpha_t} V^{\alpha}(\Delta_t) + \max_{\beta_t} V^{\beta}(\Delta_t).$$

It follows from $V^{\beta}(\Delta_t) = V^{\alpha}(\Delta_t)|_{\alpha_t=\beta_t}$, that at the optimum it must be $\alpha_t = \beta_t$. It remains to consider the case in which the constraint $1 - \alpha_t - \beta_t \geq 0$ is binding. Replacing $\beta_t = 1 - \alpha_t$ in (4) and recalling that the constraints $\alpha_t \geq 0$ and $1 - \alpha_t \geq 0$ are slack, we can write:

$$V(\Delta_t) = \max_{\alpha_t} \left\{ -\frac{L}{2} \left(\alpha_t^2 + (1-\alpha_t)^2 \right) \Delta_t - \frac{\ell}{6} \left(\alpha_t^3 + (1-\alpha_t)^3 \right) \Delta_t^2 + \delta V(\Delta_t) \right\}.$$

The first order condition gives

$$-L\left(\alpha_t - (1-\alpha_t)\right)\Delta_t - \frac{\ell}{2}\left(\alpha_t^2 - (1-\alpha_t)^2\right)\Delta_t^2 = 0,$$

from which we may conclude that the solution is $\alpha_t = \beta_t = 1/2$.

We now derive the comparative statics results. Small changes in C, L, ℓ and δ have no impact on $\lambda_t = \alpha_t + \beta_t$ in the case of a boundary solution $\alpha_t = \beta_t = 1/2$. Consider an interior solution. Exploiting the symmetry of the solution (i.e., $\alpha_t = \beta_t = \lambda_t/2$) and (5), and using the subscript t to keep track of the fact that we are only changing C, L, ℓ, δ at time t, we may write the interior first order condition $\frac{\partial V^{\alpha}}{\partial \alpha_t} = 0$ as:[18]

$$\Phi(\cdot) := C_t - \frac{L_t}{2}\lambda_t - \frac{\ell_t}{8}\lambda_t^2\Delta_t + \delta_t V(\Delta_t) - \delta_t\left(1 - \frac{\lambda_t}{2}\right)V\left(\Delta_t\left(1-\frac{\lambda_t}{2}\right)\right) - \delta_t\frac{\lambda_t}{2}V\left(\Delta_t\frac{\lambda_t}{2}\right) = 0. \tag{9}$$

Let Φ_{z_t} be the partial derivative of Φ with respect to the variable $z_t \in \{C_t, L_t, \ell_t, \delta_t\}$. By totally differentiating (9), it is immediate that $\partial\lambda_t/\partial z_t = -\Phi_{z_t}/\Phi_{\lambda_t}$. Since the second order condition of the court's maximization problem requires $\Phi_{\lambda_t} < 0$, the sign of $\partial\lambda_t/\partial z_t$ is the same as the sign of Φ_{z_t}. Observe that

$$\begin{aligned}
\Phi_{C_t} &= 1 > 0 \\
\Phi_{L_t} &= -\frac{\lambda_t}{2} < 0 \\
\Phi_{\ell_t} &= -\frac{\lambda_t^2}{8}\Delta_t < 0 \\
\Phi_{\delta_t} &= V(\Delta_t) - \left(1-\frac{\lambda_t}{2}\right)V\left(\Delta_t\left(1-\frac{\lambda_t}{2}\right)\right) - \frac{\lambda_t}{2}V\left(\Delta_t\frac{\lambda_t}{2}\right) < 0,
\end{aligned}$$

[18] In equation (9), $V(\cdot)$ is the value function at time $t+1$, which does not depend on $C_t, L_t, \ell_t, \delta_t$, the values of the parameters at time t. When the parameters, and hence the value function, vary in time, Lemma 1 can be extended to show existence, uniqueness, differentiability and negative valuedness of the value function at each point in time.

where the fourth inequality follows from $V(\Delta)$ being negative and decreasing in Δ.[19,20] ■

The value of precedent, *stare decisis*, emerges endogenously in our model. Proposition 1 formalizes a well-known view on precedent. Assuming the prior judgments were correct, the court can take those rulings as given and focus on new issues. As pointed out by Judge Benjamin Cardozo "the labor of judges would be increased almost to the breaking point if every past decision could be reopened in every case, and one could not lay one's own course of brick on the secure foundation of the courses laid by others who had gone before him." (Cardozo, 1921, 249).

The reason the court relies on precedent is simple. At the margin the cost of looking at cases near the boundary of the interval of uncertainty always outweighs the expected social loss of relying on precedent instead. To see this, suppose the court considers the marginal impact of relying on precedent and setting $\alpha_t = \beta_t = \varepsilon/2$ and hence $\lambda_t = \varepsilon$, rather than $\alpha_t = \beta_t = \lambda_t = 0$, where ε is a small number. The probability that x is drawn from the set $(W_t, a_t) \cup (b_t, R_t)$ and hence summarily decided is ε, while the probability that x is in the set and erroneously decided (i.e., $x > \theta$ in (W_t, a_t) or $x < \theta$ in (b_t, R_t)) is proportional to ε^2. It follows that the marginal benefit of relying on precedent (i.e., the expected saving on the decision cost) exceeds the marginal cost (i.e., the expected losses from an error and from giving up learning), for a sufficiently small ε.

The narrowest possible construction or interpretation of prior precedent ($\lambda_t = 0$), means that the court expends effort also on cases close to the boundary points W_t and R_t where an error is extremely unlikely. To tie with our motivating example, the Supreme Court investigates and hears a case involving a multiple of 499 times the compensatory award when it has previously found a 500 multiple impermissible. That is a waste of judicial resources. The court isn't relying enough on reasoning by analogy, i.e., extrapolating costs and benefits from similar past cases.

These results shed light on a number of phenomena. First, the judge, or decision maker, asks how close the case is to previous cases and, if sufficiently close, decides according to the precedent. The model justifies this common form of analogical analysis, which is deployed in

[19] Since $V(\Delta)$ is the value function at time $t+1$ and $C_{t+1} = C_{t+2} = ... = C$, Lemma 1 applies and $V(\Delta)$ is decreasing in Δ.

[20] When considering a change in one of the parameters C, L, ℓ, δ at each point in time, we must account for the impact of the change on the future value function. For example, differentiating (9) with respect to $C = C_t = C_{t+1} = ...$ yields:

$$\Phi_C = 1 + \delta \frac{\partial V(\Delta_t)}{\partial C} - \delta \left(1 - \frac{\lambda_t}{2}\right) \frac{\partial V\left(\Delta_t \left(1 - \frac{\lambda_t}{2}\right)\right)}{\partial C} - \delta \frac{\lambda_t}{2} \frac{\partial V\left(\Delta_t \frac{\lambda_t}{2}\right)}{\partial C}.$$

Matlab simulations show that the sign of Φ_z is the same as the sign of Φ_{z_t} for all $z \in \{C, L, \ell, \delta\}$.

many institutional settings. Although such reasoning conserves resources, there are limits on how far the analogy can be pushed. Some cases are just too different and, as a result, the prior decision is not terribly informative. Then, the analogy fails and is replaced by a fresh look.

Second, legal argument often involves statements about how close the case is to the prior case law. In our model, the closer the case is, the more likely the court will be to defer to the precedent rather than take a fresh look. Given this feature, effective advocacy involves explaining to the court why the case is closer to one set of precedential materials than to a second set of precedential materials. Indeed, judges often must decide which set of precedents more aptly apply in a new case (Posner, 2006, 63, quoting Radin, 1925, 359). In our model, this corresponds to deciding whether x_t is closer to the precedent associated with W_t or the one associated with R_t. We assume the court knows where x_t lies on the interval, but we could expand the model to allow lawyers on one side to argue that x_t is close to the precedent case W_t and hence is a valuable activity which ought to be declared permissible, while lawyers on the other side would argue that x_t is closer to R_t, the other precedent, and ought to be impermissible.

Third, reliance on precedent occurs even when the court faces no penalty for failing to follow precedent. It thus provides one explanation for why the Supreme Court defers to its own prior decisions and federal circuit courts defer to sister circuits or state courts, when they face no sanction for a failure to do so.

3.1 Precedent Ratio: The Empirical Literature and Testable Implications

Prior models of judicial decision making assume that judges do not learn and have unchanging preferences over how to divide a factual space into segments: permissible and impermissible (e.g., see Kornhueser, 1992a, 1992b and Lax, 2011, 2012). These models implicitly assume the judge is indifferent between ruling by summary disposition or in a lengthy opinion; they cannot explain why or when judges will find precedent persuasive. They also cannot explain the existence or form of legal argument and why closeness between the current case and prior cases seems to matter in the law.

Our model is different. For example, by Proposition 1 an increase in C raises the court's cost to learn about the social value of a particular activity by investigating the merits of the case. Accordingly, it encourages the court to reason by analogy to a greater extent and to resolve more cases by summary disposition. This prediction is consistent with empirical evidence. Huang (2011) studies the effect of increased caseloads on the rate at which circuit courts reverse district courts. As noted, an increase in caseload can be thought of as an increase in the decision cost of a judge. After September 11, 2001, the Attorney General made it a priority to quickly clear the deportation backlog. This move caused a flood of immigration appeals directly to two circuit courts. Huang looks at reversal rates in another kind of case – "civil" cases – in these jurisdictions. He finds that "when flooded by the agency cases, the affected circuit courts began to reverse district court rulings less often – in the civil cases. In these circuits, it seems, deference increased." (Huang, 2011, 1115). Enhanced deference can be viewed as greater use of summary review by the circuit court.

On related lines, Epstein, Landes, and Posner (2011, 101) find that "[i]n the court of appeals, the frequency of dissents is negatively related to the caseload." A dissent is more likely in a published opinion where the judges, collectively, have spent effort looking at a case and disagreements remain. Stras and Pettigrew (2010) examine a single circuit's response to a rising caseload. They report that, as the caseload has increased, the Fourth Circuit "adopted certain procedural reforms to adapt to its increased caseload. The most important and controversial of these reforms is a reduction in the percentage of cases allotted oral argument time and an increase in the percentage of cases decided through unpublished opinions." (Stras and Pettigrew, 2010, 432).

Next consider some testable implications of the comparative static on error costs. As the cost of a mistake increases, the court is less likely to decide by applying precedent alone and hence less likely to decide by summary disposition. Thus, we can empirically sort cases by issue area and then ask about the chance of summary disposition, given our expectation of the likely error cost.[21] Courts, say, should be less likely to resolve constitutional law cases without a hearing or without a written decision – a statement that can be tested. In capital punishment cases, the courts should routinely grant oral argument. Contract cases, by contrast, should often be treated summarily, especially when they involve the setting of default rules.

Finally, consider a decrease in discount factor, δ. Proposition 1 predicts that judges who care less about the future should be more likely to rely on summary disposition. Like the error cost comparative statics, this result can be tested. Some state court judges are elected; others are appointed for a given time period. Holding all else constant, elected state court judges might be assumed to care less about the distant future and learning about the proper scope of the law; as a result, they ought to be more likely to resolve cases summarily.

Given the wealth of data on judicial decision making, we believe that looking at courts provides a nice place to test our more general theory of the mechanics behind up-or-down problem solving.

[21] Of course, any empirical test based on the chance of summary disposition must account for the probability of filing by the litigants. Knowing that the case is likely to be summarily resolved, the plaintiff might not file in the first place, which, in turn makes observing summary disposition less likely.

3.2 Convergence

Legal academics, policy-makers, and advocates often criticize the law articulated by courts as imperfect or wrong-headed. For example, in tort law a defendant will be found negligent if he acts without reasonable care. The negligence standard is a knife-edge inquiry. If the defendant is found negligent, he is liable for all the resulting damage. If he is found non negligent, he pays nothing. Calfee and Craswell (1984) show that, when the defendant is uncertain about the legal standard, negligence can result in too much deterrence. The courts have not fine-tuned negligence law to account for the risk of over-deterrence identified by Calfee and Craswell. On this score, negligence law is imperfect.

As we shall see, our model shows that imperfections in doctrine are inevitable when the cost of deciding cases is sufficiently high. This is true even if the court shares the underlying values of those criticizing the decisions. The next proposition demonstrates that the court will eventually stop learning and will exclusively rely on precedent if and only if $C > L/2$. In such a case, at some point, the benefits of further refinement – tweaking the doctrine to better advance society's interests – are smaller than the costs. The court, then, refuses to refine the doctrine and lets all the new cases that come along to be ruled by precedent. On the contrary, if $C \leq L/2$ the court will never stop learning until it reaches perfect knowledge of the parameter θ.

Proposition 2. (1) *If $C > L/2$ the law converges (with probability one) without the court fully learning about θ; there exists a threshold value Δ_S such that the court chooses $\lambda_t = 1$, equivalently $a_t = b_t$ ($\neq \theta$ with probability one), when $\Delta_t \leq \Delta_S$.* (2) *If $C \leq L/2$ the court eventually fully learns; $\lambda_t < 1$ whenever $\Delta_t > 0$ and* $\lim_{t\to\infty} \Delta_t = 0$, $\lim_{t\to\infty} a_t = \lim_{t\to\infty} b_t = \theta$. Furthermore, $\lim_{t\to\infty} \lambda_t = 2C/L$.

Proof. We know from Proposition 1 and its proof that the optimal policy is $\alpha_t = \beta_t = \lambda_t/2 > 0$. Learning stops, and hence the law converges without full learning when $\alpha_t = \beta_t = 1/2$ or $\lambda_t = 1$. Replacing the policy values $\alpha_t = \beta_t = 1/2$ on the right hand side of (3), we see that the value function when it is optimal to stop learning is

$$V(\Delta_t) = -\frac{L}{4}\Delta_t - \frac{\ell}{24}\Delta_t^2 + \delta V(\Delta_t) = -\frac{\left(6L\Delta_t + \ell\Delta_t^2\right)}{24\,(1-\delta)}. \tag{10}$$

From (9) we know that the first order condition at an interior solution is $\Phi\left(\lambda_t, \Delta_t\right) = 0$. Note that $\lim_{\Delta_t\to 0} \Phi\left(\lambda_t, \Delta_t\right) = C - \frac{L}{2}\lambda_t$. Since it is $\lambda_t \leq 1$, if $C > L/2$ then it cannot be $\Phi\left(\lambda_t, \Delta_t\right) = 0$ for a "sufficiently small" value of Δ_t; that is, there cannot be an interior solution for Δ_t "small." Formally, there must exist $\overline{\Delta}$ such that $\Phi\left(\lambda_t = 1, \Delta_t\right) > 0$ if $\Delta_t < \overline{\Delta}$; for $\Delta_t < \overline{\Delta}$ there is a corner solution $\lambda_t = 1$. To derive the threshold $\overline{\Delta}$ such that $\Phi\left(\lambda_t = 1, \overline{\Delta}\right) = 0$, first replace $\lambda_t = 1$ in (9):

$$C - \frac{L}{2} - \frac{\ell}{8}\overline{\Delta} + \delta V(\overline{\Delta}) - \delta V\left(\frac{\overline{\Delta}}{2}\right) = 0; \tag{11}$$

then insert the value function from (10) to obtain

$$\begin{aligned}
0 &= C - \frac{L}{2} - \frac{\ell}{8}\overline{\Delta} - \frac{\delta\left(6L\overline{\Delta} + \ell\overline{\Delta}^2\right)}{24(1-\delta)} + \frac{\delta\left(3L\overline{\Delta} + \ell\frac{\overline{\Delta}^2}{4}\right)}{24(1-\delta)} \\
&= 3(1-\delta)\left(8C - 4L - \ell\overline{\Delta}\right) - \delta\left(3L\overline{\Delta} + \frac{3}{4}\ell\overline{\Delta}^2\right) \\
&= 12(1-\delta)(2C - L) - 3(1-\delta)\ell\overline{\Delta} - \delta\left(3L\overline{\Delta} + \frac{3}{4}\ell\overline{\Delta}^2\right) \\
&= -\frac{1}{4}\delta\ell\overline{\Delta}^2 - (\delta L + (1-\delta)\ell)\overline{\Delta} + 4(1-\delta)(2C - L)
\end{aligned}$$

from which we have a unique positive solution (as long as $2C > L$)

$$\overline{\Delta} = \frac{-2(\delta L + (1-\delta)\ell) + 2\sqrt{(\delta L + (1-\delta)\ell)^2 + 4\delta\ell(1-\delta)(2C - L)}}{\delta\ell}. \tag{12}$$

Define $\Delta_S = \min\left\{\overline{\Delta}, 1\right\}$ as the threshold interval of uncertainty. If $2C > L$ and $\overline{\Delta} \leq 1$, then (12) defines the threshold interval of uncertainty. If $\overline{\Delta} > 1$, then the threshold is 1, the maximum size of the interval of uncertainty. If $\Delta_t \leq \Delta_S$ then the court sets $\lambda_t = 1$ and no learning takes places.

Now consider the case $2C \leq L$ and suppose, to the contrary, that there is a value $\Delta_S > 0$ at which learning stops. By (11) it must be

$$\begin{aligned}
V\left(\frac{\Delta_S}{2}\right) &= V(\Delta_S) + \frac{1}{\delta}\left(C - \frac{L}{2} - \frac{\ell}{8}\Delta_S\right) \\
&< V(\Delta_S)
\end{aligned}$$

which contradicts the fact, established in Lemma 1, that V is decreasing. Hence, if $2C \leq L$ it is optimal to set $\lambda_t < 1$ in all periods t. With probability $(1-\lambda_t)\Delta_t$ the court learns in period t and in such a case $\left(1 - \frac{\lambda_t}{2}\right)\Delta_t$ is an upper bound on the new interval of uncertainty Δ_{t+1}. It follows that $\lim_{t\to\infty}\Delta_t = 0$ and hence $\lim_{t\to\infty} a_t = \lim_{t\to\infty} b_t = \theta$. Learning never stops and in the limit the court discovers the true value of θ. It is immediate from (9) that, since Δ_t (and hence $V(\Delta_t)$) converges to zero as t goes to infinity, it must be $\lim_{t\to\infty}\lambda_t = 2C/L$. ■

If decision costs are greater than half the fixed cost of errors, the doctrine stabilizes with imperfections remaining in the law.[22] The court sets $a_t = b_t$ without knowing the exact location of θ. The court realizes the doctrine might not apply well in some circumstances, but correcting those imperfections is not cost-justified. As expected, if the fixed component of the error loss is zero, imperfect convergence always happens.

Loosely speaking, with imperfect convergence, the first few investigated cases and decisions will be important to the long run determination of which activities are permissible and which ones are not. Because of high decision costs, the courts will make inferences from the first cases investigated to cover lots of future cases (deciding those summarily). The first cases will anchor the discussion and play a disproportional role in the path of the law. With efficient convergence, by contrast, such anchoring vanishes. Although we do not explicitly consider the possibility of litigants selecting specific cases for trial, the insights from Proposition 2 suggest that any such selection effect will be more important when decision costs are large relative to the fixed error cost. The reason is that litigants will race to get a case on the books that anchors the law in their favor.

As noted in the introduction, some economists speculate that the efficiency of the common law provides a theoretical justification for the main finding of the law and finance literature – that a common law origin positively correlates with economic development. Our imperfect convergence result suggests that it can be inefficient for the law to be perfect across all cases and areas of law. That is to say, with scarce judicial resources, we should expect judges to promulgate and then stick with imperfect doctrines.[23] This is true even if the judges care solely about efficiency as the relevant benchmark for legal rules.

From Proposition 2, it is also immediate that the court will not learn at all, it will set $\lambda_1 = 1$, or $\alpha_1 = \beta_1 = 1/2$, as long as $\Delta_S = 1$. When the cost of examining a case relative to the fixed and variable loss from error is above a threshold, the court never investigates at all – no precedent stock is created.[24] The court summarily declares all cases below the error minimizing point 1/2 as permissible and all cases above this point as impermissible. There is no uncertainty in the application of law and no reason for litigants to bring cases. This suggests

[22] Notably, there is empirical evidence inconsistent with the convergence of judge-made law. Niblett et al. (2010) study the evolution of the economic loss rule (the rule limiting tort damage unless the loss results in personal injury or property damage). They find that "the doctrine has evolved in a way that cannot be easily described as convergence to efficiency" (Niblett et al., 2010, 354). They also find that the law has become less predictable in the last decade of their sample, as many state courts respond with their own exceptions to the rule. *Id.* While recent changes in state court preferences is an explanation consistent with our model, it is quite possible that factors that we left outside our model play an important part in explaining these findings.

[23] Note that, even when the law converges to efficiency, judge made law will still have errors along the way, simply because the precedent ratio is always positive.

[24] After simple calculation, (12) shows that $\overline{\Delta} \geq 1$, and hence $\Delta_S = 1$, is equivalent to $C \geq \frac{L}{2} + \frac{\delta}{1-\delta}\frac{\ell}{32} + \frac{1}{8}\left(\frac{\delta}{1-\delta}L + \ell\right)$, which needs not hold even if $L = 0$.

that, when decision costs are high relative to error costs, direct government regulation, via, say, a detailed statute, may be preferable to regulating by litigation and the courts. We might think of the regulator spending a fixed amount of resources locating θ ex ante. The natural question is when fixed expenditures are better than case by case expenditures. Intuitively, fixed expenditures won't be effective if the consequences of the regulation are unforeseeable and can only be discovered following, say, the materialization of an injury.

The extent of the expected inaccuracy in the converged doctrine is captured by Δ_S; the interval of uncertainty beyond which the law stops being refined.[25] The next proposition specifies the relationship between the scope of the "inaccuracy" in doctrine and the parameters of the model. The proof, which is relegated to the appendix, follows from the definition of Δ_S after straightforward computations.

Proposition 3. *Assuming imperfect convergence of doctrine, $C > L/2$, the expected inaccuracy of the ultimate legal rule, Δ_S, increases in the decision cost C and decreases in the error losses L and ℓ and the discount factor δ.*

The intuition is straightforward: an increase in the cost of examining a case, a decrease in the loss of an error and a reduction in the value attached to the future all have the effect of making learning less valuable and hence lead to greater inaccuracy in the law.

Like Proposition 1, Proposition 3 has testable implications. First, as decision costs go up, judge-made law should be more likely to involve per se rules – generalized rules that apply without consideration for the specific circumstances. One can think of per se rules as a commitment by the court to not investigate the merits of cases. It is well-known that per se rules allow for cheaper resolution of cases, but create room for errors (Easterbrook, 1984, 14-15). That trade-off is the one formally captured in our imperfect convergence result. Second, in times of weak judicial budgets, judicial mistakes should be more common. Assuming Congressional reversal of judge-made decisions involving, say, interpretation of a statute correlates with those decisions being mistaken, we should observe Congressional reversal increasing. Of course, Congress makes funding decisions for the judiciary and also decides when to step in and reverse a statutory interpretation. To do the test, one would need an exogenous shock to decision costs of the courts, induced by an institutional actor other than Congress, as in the Huang (2011) study discussed before.

Taken together, Propositions 2 and 3 suggests two ways to minimize mistakes in the creation and application of law. One way, perhaps, is to have Congress define the statute ex ante, but this will be costly, especially if the needed information is unavailable. A second way – which

[25] Because activities x are randomly drawn, the actual size of the uncertainty interval at which learning stops is a random variable. The expected size is an increasing function of Δ_S.

turns directly on the breath of judicial decisions – is to allocate the courts more resources. With more resources the court will read prior decisions more narrowly and, as a result, make fewer mistakes, both along the doctrinal path and once the doctrine converges.

4 The Likeness Principle

Having derived the endogenous use of precedent and studied the convergence of doctrine, this section asks a more fundamental question about the evolution of law. Will evolution be consistent with the rule-of-law value that like cases be treated alike or will like cases be treated differently?[26]

A doctrinal example demonstrates how judges might change the construction of precedent over time. In *New York v. Belton*,[27] a police officer pulled over a car with four men inside. Smelling drugs, the officer asked the men to get out of the car, and placed them in four separate areas on the highway. The officer proceeded to search the passenger compartment, where he found drugs in a jacket one of the men had left behind. In upholding the search, the Supreme Court held that when "a policeman has made a lawful custodial arrest of the occupant of an automobile, he may as a contemporaneous incident of that arrest, search the passenger compartment of that automobile." *Id.* at 460. In *Thorton v. United States*,[28] the defendant exited his car before interacting with the police officer. The police officer found drugs on the defendant's body, handcuffed him and placed him in the back of the patrol car. He then searched the defendant's car, where he found a firearm. The *Thorton* court upheld the search of the vehicle. In *Arizona v. Gant*,[29] the defendant got out of his car. While walking from his car, the officer arrested the defendant for driving with a suspended license, handcuffed him and placed him in the back of the patrol car. A search of the defendant's vehicle revealed a bag of cocaine and a gun. The Court found the search impermissible under *Belton*. To sum up, in cases with similar, perhaps even identical facts (*Thorton* and *Gant*), the court read the prior case law – *Belton* – as requiring different outcomes.

Inconsistent judicial decision making is well-established in the legal and political science literatures. Many empirical studies find that judges reach outcomes at significantly different rates depending on some measure of their political beliefs (Revesz, 1997; Cross and Tiller, 1998; Sunstein et al., 2006; Cox and Miles, 2008; Boyd, Epstein, and Martin, 2010). In a

[26] On the topic, the legal and philosphical literature is vast, see Fuller (1958), Hart (1958), Tamanaha (2004), McCubbins et al. (2010). Many influential scholars have stressed that like cases should be treated alike. See, for example, Rawls (1971, 208-209), Dworkin (1977, 113), Whittington (1999, 169).

[27] 453 U.S. 454 (1981).

[28] 541 U.S. 615 (2004).

[29] 556 U.S. 332 (2009).

study of unconscionability cases, Niblett (2009, 4) finds "where the facts of cases fall within the ambit of precedent, the outcomes of the case and the precedent are inconsistent in about 23% of the case-precedent pairs."

Our model shows that different policy preferences are not necessary to generate inconsistent judicial decision making. In addition, the model predicts lots of consistency too, as shown below. The end result is a mix of consistent and inconsistent judicial-making, depending in large part on whether the law at issue is in its infancy or not.[30]

4.1 Deriving The Likeness Principle

The ability to characterize precedent one way at one time and another way at another time provides the courts flexibility, which will allow them to incorporate new information into the law. As the next proposition shows, if $\lambda_t < 1$ and learning occurs at time t, then with positive probability next period the interpretation of precedent will vary in an inconsistent way. To see how this will occur, suppose the case x_{t-1} at time $t-1$ is deemed impermissible by appealing to precedent. Now suppose the case x_t arising at time t is investigated by the court and the activity is found to be socially valuable. As a result, the court learns and the interval of uncertainty changes. Suppose the new investigation bounds are such that $x_{t-1} < a_{t+1}$. This is possible and, as the next proposition will show, has positive probability of occurring. Finally, suppose the case x_{t+1} at time $t+1$ satisfies $x_{t-1} \leq x_{t+1} < a_{t+1}$. The court will judge it summarily and declare it permissible. There is inconsistency between the decision at $t-1$ and the decision at $t+1$. Since $x_{t-1} \leq x_{t+1}$, activity x_{t-1} is at least as socially valuable as activity x_{t+1}, but x_{t-1} is deemed not permissible at $t-1$ while x_{t+1} is deemed permissible at $t+1$. The inconsistency is due to the fact that at time t the court has learned and revised upwards its estimate of where the efficiency threshold θ lies.

Intuitively, inconsistency is more likely following an unexpected result of an investigation. If x_t is only marginally smaller than b_t, the court expects the likely outcome of the investigation to be that the activity has negative social value. Upon discovering that this is not so, that x_t is indeed valuable, the court learns a lot and in later periods may declare permissible activities that in the past would have been declared impermissible by summary disposition. In addition, the magnitude of the inconsistency depends on how much the court learns from the unexpected decision: the more the court learns, the greater the potential magnitude. Since the

[30] Recently, Fishman (2011) has estimated bounds on the range of inconsistency by studying asylum adjudication in the New York immigration courts. He finds that "a randomly selected pair of judges would disagree about the disposition of a randomly selected case at least one-quarter of the time, and perhaps as often as one-half the time." *Id.* at 30.

first few cases present a large opportunity for learning, we should expect greater inconsistency early in the development of judge made law. Notably, the amount of inconsistency is not necessarily small and transpiring just around the threshold, where it doesn't much matter which way the court decides the case. Instead, inconsistency turns on drawing a case close to the previous precedent boundary and having investigation reveal an outcome contrary to the court's intuition.

Inconsistent interpretation of prior cases that occurs with positive probability implies a violation of the likeness principle. Formally, let $\Pr(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1})$ be the probability of the following event occurring. The case x_{t-1} in period $t-1$ is at least as socially valuable as the case x_{t+1} in period $t+1$; both cases are judged according to precedent; x_{t-1} is deemed not permissible; x_{t+1} is deemed permissible. $\Pr(a_{t-1} > x_{t-1} \geq x_{t+1} > b_{t+1})$ is similarly defined. We say that the evolution of jurisprudence follows the *likeness principle* in period t if

$$\Pr(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1}) + \Pr(a_{t-1} > x_{t-1} \geq x_{t+1} > b_{t+1}) = 0.$$

The next proposition shows that the evolution of jurisprudence from a court that cares about errors and has limited judicial resources is not always consistent with the lofty rule-of-law value that identical cases be treated alike.

Proposition 4. *If $\lambda_t < 1$, then with positive probability the interpretation of prior case law at time $t+1$ is inconsistent with the interpretation at time t; that is, the likeness principle is violated at time t.*

Proof. First we show that $a_{t+1} > b_t$ with positive probability. An analogous argument could be made to show that $b_{t+1} < a_t$ with positive probability. Let $\gamma^* = 1 - \lambda_t = (b_t - a_t)/\Delta_t$. Since $\lambda_t < 1$, it is $\gamma^* > 0$ and $a_t < x(\gamma) = b_t - \gamma\Delta_t \leq b_t$ for all γ in the interval $I^* = [0, \gamma^*)$. Since $\theta > x(\gamma)$ with positive probability, with positive probability $W_{t+1} = x(\gamma)$ for some $\gamma \in I^*$ and

$$\begin{aligned} a_{t+1} &= W_{t+1} + \frac{\lambda_{t+1}}{2}[R_t - W_{t+1}] = x(\gamma) + \frac{\lambda_{t+1}}{2}[R_t - x(\gamma)] \\ &= b_t - \gamma\Delta_t + \frac{\lambda_{t+1}}{2}[R_t - b_t + \gamma\Delta_t] \\ &= b_t - \gamma\Delta_t + \frac{\lambda_{t+1}}{2}\left[\frac{\lambda_t}{2}\Delta_t + \gamma\Delta_t\right]. \end{aligned} \tag{13}$$

It follows that $a_{t+1} > b_t$ as long as

$$\gamma < \frac{\lambda_t \lambda_{t+1}}{2(2 - \lambda_{t+1})} = \gamma^{**}.$$

Since, by Proposition 1, λ_t and λ_{t+1} are bounded away from zero (i.e., $\lambda_t \geq \overline{\lambda}(\Delta_t) > 0$ and $\lambda_{t+1} \geq \overline{\lambda}(R_t - x(\gamma)) > 0$), it is a positive probability event that in period t the selected case is $x(\gamma)$ with $\gamma \in [0, \min\{\gamma^*, \gamma^{**}\})$ and $\theta > x(\gamma)$. In such an event $a_{t+1} > b_t$. This shows that $\Pr(a_{t+1} > b_t) > 0$. Now observe that with positive probability $x_{t-1} \in (b_{t-1}, R_{t-1})$ and no learning takes place, so that $W_t = W_{t-1}$, $a_t = a_{t-1}$, $b_t = b_{t-1}$, $R_t = R_{t-1}$ and $\Delta_t = \Delta_{t-1}$. It follows that with positive probability $a_{t+1} > b_{t-1}$.

We conclude the proof by showing that $\Pr(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1}) > 0$. First note that with positive probability at time $t-1$ an activity x_{t-1} is selected in the interval $(b_{t-1}, b_{t-1} + \mu\Delta_t)$, where $\mu > 0$ is a constant to be defined later. The activity x_{t-1} is judged as being not permissible and the court does not learn. Since $\lambda_t < 1$, then with positive probability at t the selected activity is some $x(\gamma) = b_{t-1} - \gamma\Delta_t$, with $\gamma > 0$, which is investigated and viewed as efficient. Then, as shown in (13), the court will set $a_{t+1} = b_{t-1} - \gamma\Delta_t + \frac{\lambda_{t+1}}{2}\left[\frac{\lambda_t}{2}\Delta_t + \gamma\Delta_t\right]$ and with positive probability $a_{t+1} > b_{t-1}$. Since $x_{t-1} < b_{t-1} + \mu\Delta_t$, a sufficient condition for the event $x_{t-1} \leq x_{t+1} \leq a_{t+1}$ is

$$b_{t-1} + \mu\Delta_t < b_{t-1} - \gamma\Delta_t + \frac{\lambda_{t+1}}{2}\left[\frac{\lambda_t}{2}\Delta_t + \gamma\Delta_t\right],$$

or, equivalently,

$$\mu + \gamma < \frac{\lambda_{t+1}}{2}\left[\frac{\lambda_t}{2} + \gamma\right]. \tag{14}$$

Since λ_t and λ_{t+1} are bounded away from zero, it is possible to choose a set of values $\gamma > 0$ and $\mu > 0$ having positive probability measure and such that (14) holds. This concludes the proof that the likeness principle is violated at t when $\lambda_t < 1$. ■

Development economists and legal scholars recommend the adoption of rule-of-law values across countries. Surprisingly, and contrary to the conventional wisdom, Proposition 4 suggests that treating like cases alike, one of the central ingredients of the rule of law, is not always socially optimal in the short run and should not always be expected from a rational, benevolent court system. The benefits of non-discrimination must be traded off against learning. The court could avoid treating like cases differently by consistently construing precedent in the narrowest way. But, as explained in Proposition 1, this interpretative approach taxes judicial resources, without enough of an offsetting benefit from what can be learned from the case.

Our model does not take into account some costs of inconsistent decision making. Predictability allows parties to engage in long-run plans, without fear that an activity deemed permissible today will be found impermissible tomorrow. The ability to plan, then, facilitates "law-specific" investment by providing for settled expectations about the legal system. For example, inconsistent changes in the relevant tort standard may induce a firm not to invest on the production line of a newly designed product. We do not therefore question the general value of a reliable legal system. What our model shows is that full predictability comes with a price tag – costly judicial investigation of cases close to the prior case law.

Having established that inconsistencies can occur in the short run, the next question is whether they will persist throughout time. Will the law oscillate with, say, activities deemed permissible and then impermissible and then permissible again as time goes by?[31] First, it is obvious that if learning eventually stops, so too will inconsistency. Inconsistency requires adjustments of the beliefs about the location of θ. If the court stops investigating, it will never update its beliefs.

What if learning never stops? Recall that to generate inconsistent decisions at time t three events must occur. First, the case at time $t-1$ must be in the portion of the interval of uncertainty Δ_{t-1} where the judge makes a summary decision. Second, the case at time t must be in the portion of the interval of uncertainty $\Delta_t = \Delta_{t-1}$ where the judge investigates the case. Third, the case in period $t+1$ must be summarily decided inconsistently with the case at time $t-1$. As time goes by and the court learns, the interval of uncertainty converges to zero. It follows that the probability of inconsistent decisions vanishes and hence the likeness principle holds in the limit. Formally, we say that the *limit likeness principle* holds if:

$$\lim_{t\to\infty} \left[\Pr\left(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1}\right) + \Pr\left(a_{t-1} > x_{t-1} \geq x_{t+1} > b_{t+1}\right)\right] = 0$$

As we have informally argued, the following proposition holds.

Proposition 5. *The optimal evolution of doctrine satisfies the limit likeness principle.*

Proof. If $C > L/2$, then by Proposition 2 $\lim_{t\to\infty} \lambda_t = 1$ and hence $\lim_{t\to\infty} a_t = \lim_{t\to} b_t$. If $C \leq L/2$, then by Proposition 2 $\lim_{t\to\infty} a_t = \lim_{t\to} b_t = \theta$. Hence, in both cases

$$\lim_{t\to\infty} \Pr\left(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1}\right) = \lim_{t\to\infty} \Pr\left(a_{t-1} > x_{t-1} \geq x_{t+1} > b_{t+1}\right) = 0.$$

■

With the use of a simple continuity argument, observe that if the limit likeness principle holds then, given any finite number of periods n and any arbitrarily small probability mass ε,

[31] In our model, standards, in effect, harden into rules. We do not speak to oscillation, where standards turn into rules and back into standards. Such movement might be incorporated in the model by assuming that the location of θ changes with some probability each period. On this type of oscillation more generally and the convergence of rules and standards, see Baker and Kim (2011), Schauer (2003), Johnston (1995), and Rose (1989).

the probability that the interpretation of precedent at time t is inconsistent with the interpretation of precedent in any of the previous n periods becomes less than ε as t grows large. The probability of an inconsistent interpretation at t with any of the previous n periods goes to zero in the limit.

4.2 Testable Implications of Inconsistent Judicial Decision Making

Combined, Proposition 4 and 5 lead to several testable implications. First, as noted, one would expect to observe more inconsistency when judge-made law is in its infancy – shortly, say, after Congress passes a statute containing a broad delegation of authority to the courts. This prediction can be run in a horse race against the preference-based account of judicial decision making, the dominant account. If inconsistency is solely the result of different judicial preferences, the length of time between the passage of the enabling statute and the amount of inconsistent decision making should not be significant.[32]

Second, under our model the same judge can be inconsistent over time. To the extent that one finds judges switching their own positions in the application of the legal standard, this is evidence in favor of our model and against a preference-based model of judicial decision making (unless, of course, one claims that a judge's preferences change over time, a claim which would make the preference-based model untestable).

Third, our model suggests that inconsistent dispositions should be more likely following a surprise decision. To test this, one might take an opinion that caught the legal commentators off guard – an unexpected resolution of a legal issue. Then, one might find a set of cases previously decided summarily, by, say, unpublished decision or the denial of certiorari (a form of summary disposition), and ask whether courts in subsequent cases with roughly similar facts granted oral argument or resolved the case with a written opinion.[33]

Fourth, our model predicts a higher probability of inconsistency between two unpublished decisions or a published decision and an unpublished decision than between two published decisions. Again, if judicial preferences were the sole driver of inconsistency, the form of resolution should not make a difference to the amount of inconsistency.

Finally, the inconsistency result can be linked back to the motivating example involving the

[32] Niblett (2009) provides some suggestive evidence of the sort we envision. In the unconscionability cases, he finds that "the level of inconsistency in the system, on average, continues to increase until about the 100th case. After this, the level of inconsisteny falls – quite dramatically." *Id.* at 7. Niblett finds also that conflicting judicial politics is a good predictor of inconsistency. *Id.* at 30. As explained above, unlike our model and as a matter of theory, judicial preferences alone cannot explain why the amount of inconsistency varies over time.

[33] A certiorari is a formal request to a court (in this example the Supreme Court) challenging a legal decision of an administrative tribunal, judicial office or organization.

constitutional limit on punitive damages. In 1991 in *Pacific Mut. Life Ins. Co. v. Haslip*,[34] the Court found constitutional an award with a ratio of four to one between punitive damages and compensatory damages. In 1993, the Court in *TXO Prod. Corp. v. Alliance Resource Corp*[35] found that an award of 10 to 1 passed constitutional muster. Suppose that, during the period 1991 and 1993, the court failed to grant certiorari and allowed a circuit court decision to stand striking down an award of 6 to 1. That, in our model, would be evidence of inconsistency. Likewise, in 1996 in *B.M.W., Inc. v. Gore*,[36] the Court found an award of 500 to 1 grossly excessive. In 2003, the court found an award of 145 to 1 grossly excessive in *State Farm Mutual Automobile Ins. Co. v. Campbell*.[37] To test for inconsistency, one might search for cases between 1996 and 2003, in which the Court summarily resolved a case – i.e., denied certiorari – where a circuit court had upheld an award with a multiple between 145 and 499.

5 Conclusion

We believe the results derived here are applicable to any situation where decision makers learn from experience and make yes or no decisions on a case by case basis. Take a front-line employee deciding how to deal with grievances by his subordinates. He must decide which grievances should be sent up the chain of command and which should not. Suppose he refers the first grievance up the chain. That decision establishes a precedent. The next grievance he encounters requires reflection on how close that grievance is to the previous one. If the two are close, the front-line employee saves resources by simply following precedent, rather than investigating the pros and cons of sending that specific grievance to his superior. But there is risk of mistake. Perhaps the second grievance is one that he should handle. The same results follow in this situation as in the model – the endogenous following of precedent, inconsistent decisions, the failure to always treat like cases alike and the making of rules that sticks, despite the decision maker realizing the rules work improperly in some circumstances.

Turning back to judge-made law, a few remarks are in order. In contrast to what is done here, the non-formal legal literature on this topic often speak of judges "writing" broadly or narrowly (Sunstein, 1999, 10). In reality, of course, both the writing judge and reading judge play a role in the creation of judge-made law. The "writing" judge might use expansive language in the opinion to signal something to future judges or lower courts about what he learned from his investigation and about how the precedent could be applied in future cases. Here, such

[34] 499 U.S. 1 (1991).
[35] 509 U.S. 443 (1993).
[36] 517 U.S. 559 (1996).
[37] 538 U.S. 408 (2003).

signaling is unnecessary because the prior judges never make mistakes upon investigation and opinions perfectly convey the investigation results.

The model might be extended to include judges with different preferences and opinions containing mistakes in transmission. The interesting insights from this extension would come from the distortions to the interpretation of precedent that the current judge may introduce in order to counteract the unwanted effects on the law of a future judge with different policy preferences. What's surprising is that our model has descriptive power, while maintaining the strong assumption that judges share the same normative commitments. For instance, optimal deviations from the likeness principle arise even when nothing about judicial preferences or the underlying environment has changed.

In addition, the model is framed in terms of a single infinitely lived judge. Instead, one might interpret the model in terms of a judicial hierarchy, where the upper court has high decision costs. To economize on these costs, the upper court hears some cases in the interval of uncertainty. It then delegates other cases close to its prior decisions to the lower court, which has low decision costs. The interpretative choice, then, becomes how many cases to delegate. This interpretation of the model makes some headway in explaining both the degree of delegation and the reason lower courts defer to higher court precedents (deference occurs because the higher court decisions result from investigation and, as a result, are correct). A delegation-style model of this sort bears resemblance to recent work by Ellison and Holden (2010).

One final worthwhile extension would be to endogenize the decision of which cases are brought before the court. For example, it could be that one party, say the party interested in the activity being declared permissible, is in a stronger position to bring cases to court. The court will then tend to see a biased sample of cases and it is natural to conjecture that the court will want to skew its reliance on precedent against the stronger party, in order to facilitate learning.

We conclude by stressing that our model reflects what judges claim to be doing: (1) looking at facts; (2) surveying prior precedent for guidance about what to do; and (3) trying to reach the best result. Notably, many of the features we observe in judge-made law flow as a natural consequences of judges doing what they say they are doing.

Appendix

In this appendix, first we prove existence and uniqueness of the value function of the court's maximization problem, then we prove Proposition 3 and finally we show that Proposition 1 and 2, proven in the main body for the uniform model, extends to the general model.

Lemma 1. *In the uniform model, the value function $V(\Delta)$ is continuously differentiable, negative valued, decreasing in* Δ, with $V(0) = 0$ *and uniquely defined by:*

$$V(\Delta_t) = \max_{\alpha_t \geq 0,\ \beta_t \geq 0,\ 1-\alpha_t-\beta_t \geq 0} \left\{ -C(1-\alpha_t-\beta_t)\Delta_t - \frac{L}{2}\left(\alpha_t^2+\beta_t^2\right)\Delta_t \right.$$
$$- \frac{\ell}{6}\left(\alpha_t^3+\beta_t^3\right)\Delta_t^2 + \delta V(\Delta_t)\left[1-(1-\alpha_t-\beta_t)\Delta_t\right] \tag{15}$$
$$\left. +\delta \int_{\alpha_t}^{1-\alpha_t} V(\Delta_t x_t)\Delta_t x_t dx_t + \delta \int_{\beta_t}^{1-\beta_t} V(\Delta_t x_t)\Delta_t x_t dx_t \right\}.$$

Proof. Recalling that $\alpha_t\Delta_t = a_t - W_t$, $\beta_t\Delta_t = R_t - b_t$, and using (1) and (2), we can write the court's objective function for the uniform version as:

$$V(\Delta_t) = \max_{\alpha_t \geq 0,\ \beta_t \geq 0,\ 1-\alpha_t-\beta_t \geq 0} \left\{ -C(1-\alpha_t-\beta_t)\Delta_t - \frac{L}{2}\left(\alpha_t^2+\beta_t^2\right)\Delta_t \right.$$
$$- \ell \int_{W_t}^{a_t} \frac{(x_t-W_t)^2}{2\Delta_t}dx_t - \ell \int_{b_t}^{R_t} \frac{(R_t-x_t)^2}{2\Delta_t}dx_t + \delta V(\Delta_t)\left[1-(1-\alpha_t-\beta_t)\Delta_t\right]$$
$$\left. +\delta \int_{W_t+\alpha_t\Delta_t}^{R_t-\beta_t\Delta_t} V(y_t-W_t)\frac{y_t-W_t}{\Delta_t}dy_t + \delta \int_{W_t+\alpha_t\Delta_t}^{R_t-\beta_t\Delta_t} V(R_t-y_t)\frac{R_t-y_t}{\Delta_t}dy_t \right\}.$$

Changing the variable of integration from y_t to $x_t = (y_t - W_t)/\Delta_t$ in the third integral and to $x_t = (R_t - y_t)/\Delta_t$ in the fourth integral yields

$$V(\Delta_t) = \max_{\alpha_t \geq 0,\ \beta_t \geq 0,\ 1-\alpha_t-\beta_t \geq 0} \left\{ -C(1-\alpha_t-\beta_t)\Delta_t - \frac{L}{2}\left(\alpha_t^2+\beta_t^2\right)\Delta_t \right.$$
$$- \frac{\ell}{6}\left(\alpha_t^3+\beta_t^3\right)\Delta_t^2 + \delta V(\Delta_t)\left[1-(1-\alpha_t-\beta_t)\Delta_t\right]$$
$$\left. +\delta \int_{\beta_t}^{1-\alpha_t} V(\Delta_t x_t)\Delta_t x_t dx_t + \delta \int_{\alpha_t}^{1-\beta_t} V(\Delta_t x_t)\Delta_t x_t dx_t \right\}.$$

Rearranging, we obtain (15).

We now show existence, uniqueness and differentiability of the value function. Recall that $\Delta \in [0,1]$ and note that $V(0) = 0$. Let S be the metric space of continuously differentiable, real valued functions $\omega : [0,1] \to \mathbb{R}$. Let the metric on S be $\rho\left(\omega^0,\omega^1\right) = \sup_{\Delta\in[0,1]} \left|\omega^0(\Delta) - \omega^1(\Delta)\right|$. Define the operator T, mapping the metric space S onto itself as follows:

$$T\omega(\Delta)=\max_{\alpha\geq0,\ \beta\geq0,\ 1-\alpha-\beta\geq0}\frac{\Delta}{1-\delta\left(1-\Delta\right)}\left\{-C\left(1-\alpha-\beta\right)-\frac{L}{2}\left(\alpha^{2}+\beta^{2}\right)\right.$$
$$\left.-\frac{\ell}{6}\left(\alpha^{3}+\beta^{3}\right)\Delta+\delta\omega(\Delta)\left(\alpha+\beta\right)+\delta\int_{\alpha}^{1-\alpha}\omega(\Delta x)xdx+\delta\int_{\beta}^{1-\beta}\omega(\Delta x)xdx\right\}. \tag{16}$$

Interpreting ω as an initial value-function guess, the mapping T associates to ω an updated guess $T\omega$. The value function V defined in (15) is a fixed point of the mapping T. (The parameter $\frac{\Delta}{1-\delta(1-\Delta)}$ is obtained by moving the term $V\left(\Delta\right)\delta\left(1-\Delta\right)$ from the right to the left hand side of (15) and then dividing both sides by $1-\delta\left(1-\Delta\right)$.) We now show that T is a contraction mapping and hence V exists and is unique (and continuously differentiable). We apply Blackwell's Theorem (see Blackwell, 1965, or Stokey and Lucas, 1989, p.54). We need to show that T satisfies monotonicity and discounting. (1) Take $\omega^{0}\left(\Delta\right)\leq\omega^{1}\left(\Delta\right)$ for all $\Delta\in[0,1]$. It is immediate that $T\omega^{0}\left(\Delta\right)\leq T\omega^{1}\left(\Delta\right)$ and hence monotonicity holds. (2) To see that discounting also holds, let z be a non negative constant map defined by $z\left(\Delta\right)=z$ for all $\Delta\in[0,1]$. Let the map $\omega+z$ be defined by $\left(\omega+z\right)\left(\Delta\right)=\omega(\Delta)+z$. We need to show that $T\left(\omega+z\right)\left(\Delta\right)=T\omega(\Delta)+\gamma z$ for some $\gamma\in\left(0,1\right)$. It is

$$\begin{aligned}
T\left(\omega+z\right)\left(\Delta\right)=&\max_{\alpha\geq0,\ \beta\geq0,\ 1-\alpha-\beta\geq0}\frac{\Delta}{1-\delta\left(1-\Delta\right)}\left\{-C\left(1-\alpha-\beta\right)-\frac{L}{2}\left(\alpha^{2}+\beta^{2}\right)\right.\\
&-\frac{\ell}{6}\left(\alpha^{3}+\beta^{3}\right)\Delta+\delta\left(\omega(\Delta)+z\right)\left(\alpha+\beta\right)\\
&\left.+\delta\int_{\alpha}^{1-\alpha}\left(\omega(\Delta x)+z\right)xdx+\delta\int_{\beta}^{1-\beta}\left(\omega(\Delta x)+z\right)xdx\right\}\\
=&\max_{\alpha\geq0,\ \beta\geq0,\ 1-\alpha-\beta\geq0}\frac{\Delta}{1-\delta\left(1-\Delta\right)}\left\{-C\left(1-\alpha-\beta\right)-\frac{L}{2}\left(\alpha^{2}+\beta^{2}\right)\right.\\
&-\frac{\ell}{6}\left(\alpha^{3}+\beta^{3}\right)\Delta+\delta\omega(\Delta)\left(\alpha+\beta\right)+\delta\int_{\alpha}^{1-\alpha}\omega(\Delta x)xdx\\
&\left.+\delta\int_{\beta}^{1-\beta}\omega(\Delta x)xdx+\delta z\left(\alpha+\beta\right)+\frac{\delta z}{2}\left(\left(1-\alpha\right)^{2}-\alpha^{2}+\left(1-\beta\right)^{2}-\beta^{2}\right)\right\}\\
=&T\omega(\Delta)+\frac{\Delta\delta z}{1-\delta\left(1-\Delta\right)}\left(\alpha+\beta+\frac{\left(1+\alpha^{2}-2\alpha\right)-\alpha^{2}+\left(1+\beta^{2}-2\beta\right)-\beta^{2}}{2}\right)\\
=&T\omega(\Delta)+\frac{\delta\Delta}{1-\delta\left(1-\Delta\right)}z,
\end{aligned}$$

Since $\frac{\delta\Delta}{1-\delta(1-\Delta)}=\frac{1}{1+\frac{1-\delta}{\delta\Delta}}\in(0,1)$, this proves that discounting holds and hence that T is a contraction. Its unique fixed point is the continuously differentiable value function V.

To see that V is negative and decreasing, consider the set S' of continuously differentiable functions which are negative-valued and decreasing (i.e., such that $\omega(\Delta) \leq 0$ and $\omega'(\Delta) \leq 0$). The set S' is a closed subset of S. We need to show that T maps S' onto itself. Then we can conclude that the fixed point of T, the value function V, is negative-valued and decreasing (e.g., see Corollary 1 to the Contraction Mapping Theorem in Stokey and Lucas, 1989, 52.). First, it is immediate from (16) that if $\omega(\Delta) \leq 0$, then it is also $T\omega(\Delta) \leq 0$. Hence T maps negative functions into negative functions. It only remains to show that T maps decreasing function into decreasing functions. Suppose $\omega(\Delta)$ is decreasing and differentiate (16) at the solution values α_*, β_* to obtain:

$$\begin{aligned}
\frac{\partial T\omega(\Delta)}{\partial \Delta} &= \frac{1-\delta}{[1-\delta(1-\Delta)]^2}\left\{\frac{1-\delta(1-\Delta)}{\Delta}T\omega(\Delta)\right\} \\
&\quad + \frac{\Delta}{1-\delta(1-\Delta)}\left\{-\frac{\ell}{6}\left(\alpha_*^3+\beta_*^3\right)+\delta\omega'(\Delta)(\alpha_*+\beta_*)\right. \\
&\quad \left. +\delta\int_{\alpha_*}^{1-\alpha_*}\omega'(\Delta x)x^2dx+\delta\int_{\beta_*}^{1-\beta_*}\omega'(\Delta x)x^2dx\right\} \\
&\leq 0.
\end{aligned}$$

where the first term is curly brackets is equal to the term in curly brackets in (16) and the inequality follows from all terms being negative. ■

We now provide the proof of Proposition 3.

Proposition 3. *Assuming imperfect convergence of doctrine, $C > L/2$, the expected inaccuracy of the ultimate legal rule, Δ_S, increases in the decision cost C and decreases in the error losses L and ℓ and the discount factor δ.*

Proof. Recall that Δ_S is implicitly defined by:

$$\varphi(\cdot) := -\frac{1}{4}\delta\ell\Delta_S^2 - (\delta L + (1-\delta)\ell)\Delta_S + 4(1-\delta)(2C-L) = 0$$

Since $\partial\varphi/\partial\Delta_S < 0$, the sign of the impact on Δ_S of a change in exogenous variable $z \in \{C, L, \ell, \delta\}$ is given by $\partial\varphi/\partial z$.

It immediate that $\partial\varphi/\partial C > 0$, $\partial\varphi/\partial L < 0$ and $\partial\varphi/\partial\ell < 0$; hence Δ_S is increasing in C and decreasing in L and ℓ. Moreover:

$$\begin{aligned}
\partial\varphi/\partial\delta &= -\frac{1}{4}\ell\Delta_S^2 - (L-\ell)\Delta_S - 4(2C-L) \\
&= \frac{1}{\delta}[\varphi + \ell\Delta_s - 4(2C-L)] \\
&= \frac{1}{\delta}[\ell\Delta_s - 4(2C-L)].
\end{aligned}$$

Recall from (12) that

$$\ell\Delta_S = \frac{-2\left(\delta L + (1-\delta)\,\ell\right) + 2\sqrt{\left(\delta L + (1-\delta)\,\ell\right)^2 + 4\delta\ell\,(1-\delta)\,(2C-L)}}{\delta}$$

and hence the sign of $\partial\varphi/\partial\delta$ is the same as the sign of

$$-2\left(\delta L + (1-\delta)\,\ell\right) + 2\sqrt{\left(\delta L + (1-\delta)\,\ell\right)^2 + 4\delta\ell\,(1-\delta)\,(2C-L)} - 4\delta\,(2C-L)\,,$$

which is negative since

$$\sqrt{\left(\delta L + (1-\delta)\,\ell\right)^2 + 4\delta\ell\,(1-\delta)\,(2C-L)} < \left(\delta L + (1-\delta)\,\ell\right) + 2\delta\,(2C-L)\,.$$

It follows that Δ_S is decreasing in δ. ■

We now consider the general model and provide sketches of the proofs of Lemma 1 and Propositions 1 and 2; the proofs of Propositions 4 and 5 only require minor modifications (e.g., keeping track of the fact that it is not any longer true that $\alpha_t = \beta_t = \frac{\lambda_t}{2}$) and are omitted.

Lemma $1^{\mathbf{G}}$. *The value function $V(W_t, R_t)$ is continuously differentiable and negative valued,* with $V(0) = 0$.

Proof Sketch. Let S be the metric space of continuously differentiable, real valued functions $\omega : D \to \mathbb{R}$, where $D = \left\{(x,y) \in [0,1]^2 : x \leq y\right\}$. Define the operator T, mapping the metric space S onto itself as follows:

$$\begin{aligned}
T\omega(W_t,R_t) = \frac{1}{[F(R_t)-F(W_t)]} \max_{W_t \leq a_t \leq b_t \leq R_t} \Big\{ & \delta\omega(W_t,R_t)\left[1-(G(b_t)-G(a_t))\right][F(R_t)-F(W_t)] \\
& + \delta \int_{a_t}^{b_t} \omega(W_t,x_t)\left[F(x_t)-F(W_t)\right] g(x_t)dx_t + \delta \int_{a_t}^{b_t} \omega(x_t,R_t)\left[F(R_t)-F(x_t)\right] g(x_t)dx_t \\
& - C\left[G(b_t)-G(a_t)\right]\left[F(R_t)-F(W_t)\right] \qquad (17) \\
& - \int_{W_t}^{a_t}\int_{W_t}^{x_t} Lf(\theta)\,d\theta g(x_t)dx_t - \int_{b_t}^{R_t}\int_{x_t}^{R_t} Lf(\theta)d\theta g(x_t)dx_t \\
& - \int_{W_t}^{a_t}\int_{W_t}^{x_t} \ell\,(x_t-\theta)\,f(\theta)d\theta g(x_t)dx_t - \int_{b_t}^{R_t}\int_{x_t}^{R_t} \ell\,(\theta-x_t)\,f(\theta)d\theta g(x_t)dx_t \Big\}.
\end{aligned}$$

The value function V defined in (1) is a fixed point of the mapping T. To apply Blackwell (1965) Theorem (i.e., show existence of V), we must show that T satisfies monotonicity and discounting. Monotonicity is immediate, since $\omega^0(W_t,R_t) \leq \omega^1(W_t,R_t)$ for all $(W_t,R_t) \in D$ implies $T\omega^0(W_t,R_t) \leq T\omega^1(W_t,R_t)$. To see that discounting also holds, let z be a non negative constant map defined by $z(W_t, R_t) = z$ for all $(W_t, R_t) \in D$. Let the map $\omega + z$ be defined by $(\omega + z)(W_t, R_t) = \omega(W_t, R_t) + z$. We must show that $T(\omega + z)(W_t, R_t) = T\omega(W_t, R_t) + \gamma z$ for some $\gamma \in (0,1)$. The equality holds for $\gamma = \delta$, since

$$
\begin{aligned}
& T(\omega+z)(W_t,R_t) \\
& = \frac{1}{[F(R_t)-F(W_t)]} \max_{W_t \le a_t \le b_t \le R_t} \Big\{ \delta\left[\omega(W_t,R_t)+z\right]\left[1-(G(b_t)-G(a_t))\right]\left[F(R_t)-F(W_t)\right] \\
& + \delta \int_{a_t}^{b_t} \left\{ \left[\omega(W_t,x_t)+z\right]\left[F(x_t)-F(W_t)\right] + \left[\omega(x_t,R_t)+z\right]\left[F(R_t)-F(x_t)\right] \right\} g(x_t)dx_t \\
& - C\left[G(b_t)-G(a_t)\right]\left[F(R_t)-F(W_t)\right] - \int_{W_t}^{a_t}\int_{W_t}^{x_t} Lf(\theta)d\theta g(x_t)dx_t - \int_{b_t}^{R_t}\int_{x_t}^{R_t} Lf(\theta)d\theta g(x_t)dx_t \\
& - \int_{W_t}^{a_t}\int_{W_t}^{x_t} \ell\left(x_t-\theta\right)f(\theta)d\theta g(x_t)dx_t - \int_{b_t}^{R_t}\int_{x_t}^{R_t} \ell\left(\theta-x_t\right)f(\theta)d\theta g(x_t)dx_t \Big\} \\
& = T(\omega+z)(W_t,R_t) + \delta z.
\end{aligned}
$$

Now consider the set S' of continuously differentiable functions which are negative-valued. It is immediate to conclude that the value function V is negative-valued, since T maps S' onto itself: $\omega(W_t, R_t) \le 0$ implies $T\omega(W_t, R_t) \le 0$. ■

The following first order conditions are obtained by differentiating (1) with respect to a_t and b_t, with multiplier η_t associated to the constraint $a_t \le b_t$, and multipliers μ_t^a, μ_t^b associated to constraints $W_t \le a_t$ and $b_t \le R_t$:

$$
\begin{aligned}
0 \;=\; & C - L\frac{F(a_t)-F(W_t)}{F(R_t)-F(W_t)} - \ell\int_{W_t}^{a_t}\frac{(a_t-\theta)f(\theta)}{F(R_t)-F(W_t)}d\theta + \delta V(W_t,R_t) \\
& -\delta V(a_t,R_t)\frac{F(R_t)-F(a_t)}{F(R_t)-F(W_t)} - \delta V(W_t,a_t)\frac{F(a_t)-F(W_t)}{F(R_t)-F(W_t)} - \frac{(\eta_t-\mu_t^a)}{g(a_t)}
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
0 \;=\; & -C + L\frac{F(R_t)-F(b_t)}{F(R_t)-F(W_t)} + \ell\int_{b_t}^{R_t}\frac{(\theta-b_t)f(\theta)}{F(R_t)-F(W_t)}d\theta - \delta V(W_t,R_t) \\
& +\delta V(b_t,R_t)\frac{F(R_t)-F(b_t)}{F(R_t)-F(W_t)} + \delta V(W_t,b_t)\frac{F(b_t)-F(W_t)}{F(R_t)-F(W_t)} + \frac{\left(\eta_t-\mu_t^b\right)}{g(b_t)}.
\end{aligned}
\tag{19}
$$

Proposition $1^{\mathbf{G}}$. *In each period t the court chooses $a_t > W_t$ and $b_t < R_t$.*

Proof Sketch. Suppose, to the contrary, that $a_t = W_t < R_t$. Equation (18) becomes $C = \frac{(\eta_t - \mu_t^a)}{g(a_t)}$, which can only be satisfied if $b_t = a_t$. Then (19) becomes $C - L - \ell\int_{W_t}^{R_t}\frac{(\theta - W_t)f(\theta)}{F(R_t)-F(W_t)}d\theta = \frac{(\eta_t - \mu_t^b)}{g(b_t)}$, which can only be satisfied if $b_t = R_t$. This contradicts $b_t = a_t = W_t$. Hence it must be $a_t > W_t$ whenever $W_t < R_t$.

Similarly, suppose $b_t = R_t > W_t$. Then (19) becomes $C = \frac{(\eta_t - \mu_t^b)}{g(a_t)}$, which requires $a_t = b_t$, while (18) becomes $C - L - \ell \int_{W_t}^{R_t} \frac{(a_t - \theta) f(\theta)}{F(R_t) - F(W_t)} d\theta = \frac{(\eta_t - \mu_t^a)}{g(a_t)}$, which requires $a_t = W_t$, a contradiction. Hence it must be $b_t < R_t$ whenever $R_t > W_t$. ■

Proposition 2^G. (1) *If* $C > L/2$ *the law converges (with probability one) without the court fully learning about* θ. (2) *If* $C \leq L/2$ *the court eventually fully learns,* $\lim_{t\to\infty} a_t = \lim_{t\to\infty} b_t = \theta$.

Proof Sketch. For full learning to take place in the limit (i.e., $a_t \to \theta$, $b_t \to \theta$) it must be $a_t \neq b_t$ whenever $W_t \neq R_t$. On the other hand, learning stops if $a_t = b_t$. By Proposition 1^G, it is $\mu_t^a = \mu_t^b = 0$, and hence $V(W_t, R_t) < V(a_t, R_t)$ and $V(W_t, R_t) < V(W_t, a_t)$ whenever $W_t \neq R_t$. Furthermore, the last five terms on the right hand side of (18) and (19) are bounded and converge to zero as $R_t - W_t$ converges to zero. Thus, (18) and (19) imply that there exists a positive constant M such that

$$\begin{aligned} M\,(R_t - W_t) &> C - L\frac{F(a_t) - F(W_t)}{F(R_t) - F(W_t)} > 0 \\ M\,(R_t - W_t) &> C - L\frac{F(R_t) - F(b_t)}{F(R_t) - F(W_t)} > 0, \end{aligned}$$

as long as $R_t \neq W_t$. Adding up, we obtain

$$2M\,(R_t - W_t) > 2C - L\left(1 - \frac{F(b_t) - F(a_t)}{F(R_t) - F(W_t)}\right) > 0. \tag{20}$$

Suppose $C > L/2$ and, contrary to the proposition, learning never stops ($a_t \neq b_t$ for all t). Then it must be $\lim_{t\to\infty} R_t - W_t = 0$, and (20) implies that $2C - L = \lim_{t\to\infty} L\frac{F(a_t) - F(b_t)}{F(R_t) - F(W_t)} \leq 0$, a contradiction. This shows that if $C > L/2$ the law converges without the court fully learning.

Now suppose $C \leq L/2$. Note that (20) cannot be satisfied when $a_t = b_t$, because it cannot be $2C - L > 0$. Hence in such a case learning never stops. ■

References

Baker, Scott, and Pauline Kim. 2011. "A Dynamic Model of Legal Doctrine." Manuscript. School of Law, Washington University in St. Louis.

Bailey, Martin, and Paul Rubin. 1994. "A Positive Theory of Legal Change." *International Review of Law and Economics* 14 (December): 467-77.

Blackwell, David. 1965. "Discounted Dynamic Programming." *Annals of Mathematical Statistics* 36 (February): 226-235.

Boyd, Christina L., Lee Epstein, and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54 (April): 389–411.

Calfee, John E. and Richard Craswell. 1984. "Some Effects of Uncertainty on Compliance with Legal Standards." *Virginia Law Review*, 70 (June): 965-1003.

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

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

Cox, Adam B. and Thomas J. Miles. 2008. "Judging the Voting Rights Act." *Columbia Law Review* 108 (January): 1–54.

Cross, Frank B., and Emerson H. Tiller. 1998. "Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals." *Yale Law Journal* 107 (May): 2155–2176.

Dari-Mattiacci, Giuseppe, Bruno Deffainis, and Bruno Lovat. 2011 "The Dynamics of the Legal System." *Journal of Economic Behavior and Organization* 79 (March advance access): 95-107.

Dworkin, Ronald. 1977. *Taking Rights Seriously.* Cambridge: Harvard University Press.

Easterbrook, Frank H. 1984. "The Limits of Antitrust." *Texas Law Review* 63 (August): 1-40.

Ellison, Glen and Richard Holden. 2010. "A Theory of Rule Development." Manuscript. Department of Economics, MIT.

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 (Spring): 101-137.

Eskridge, William N. and John Ferejohn. 1992. "The Article I Section 7 Game." *Georgetown Law Journal* 80 (February): 523-564.

Fishman, Joshua. 2011. "Inconsistency, Indeterminacy, and Error in Adjudication." Manuscript. Law School, University of Virginia.

Fuller, Lon. L. 1958. "Positivism and Fidelity to Law – A Reply to Professor Hart." *Harvard Law Review* 71 (February): 630-672.

Gennaioli, Nicola, and Andrei Shleifer. 2007. "The Evolution of the Common Law," *Journal of Political Economy* 115 (February): 43-68.

Hadfield, Gillian K. 1992. "Bias in the Evolution of Legal Rules." *Georgetown Law Journal* 80 (February): 583-616.

Hadfield, Gillian K. 2011. "The Dynamic Quality of Law: The Role of Judicial Incentives and Legal Human Capital." *Journal of Economic, Behavior, and Organization* 79 (March advance access): 80-94.

Hart, H.L.A. 1958. "Positivism and Fidelity to the Separation of Law and Morals," *Harvard Law Review* 71 (February): 593-629.

Hathaway, Oona A. 2001. "Path Dependence in the Law: The Course and Pattern of Legal Change in a Common Law System." *Iowa Law Review* 86 (January): 601-65.

Huang, Bert I. 2011. "Lightened Scrutiny." *Harvard Law Review* 124 (March): 1109-1152.

Hylton, Keith. 2006. "Information, Litigation, and the Common Law Evolution." *American Law and Economics Review* 8 (Spring): 33-61.

Johnston, Jason Scott. 1991. "Uncertainty, Chaos and the Torts Process: An Economic Analysis of Legal Form." *Cornell Law Review* 76 (January): 341-400.

Klerman, Daniel. 2007. "Jurisdictional Competition and the Evolution of the Common Law," *University of Chicago Law Review* 74 (Fall): 1179-1226.

Kornhauser, Lewis A. 1992a. "Modeling Collegial Courts I: Path-Dependence." *International Review of Law and Economics* 12 (June):169-85.

Kornhauser, Lewis A. 1992b. "Modeling Collegial Courts II: Legal Doctrine." *Journal of Law, Economics & Organization* 8 (October): 441-70.

La Porta, Rafael, Florencio Lopez-de-Silanes, Cristian Pop-Eleches, and Andrei Schleifer. 2004. "Judicial Checks and Balances." *Journal of Political Economy* 112 (April): 445-470.

La Porta, Rafael, Florencio Lopez–de-Silanes, Andrei Shleifer, and Robert Vishny. 1998. "Law and Finance." *Journal of Political Economy* 106 (December): 1113-1155.

Lax, Jeffrey A. 2011. "The New Judicial Politics of Legal Doctrine." *Annual Review of Political Science* 14 (February Advance Access): 131-157.

Lax, Jeffrey A. 2012. "Political Constraints on Legal Doctrine: How Hierarchy Shapes the Law." Manuscript. Department of Political Science, Columbia University.

Levi, Edward H. 1949. *An Introduction to Legal Reasoning.* Chicago: University of Chicago Press.

McCubbins, Matthew D., Daniel B. Rodriguez, and Barry R. Weingast."The Rule of Law Unplugged." Manuscript. Law School, University of Texas.

McNollgast. 1995. "Politics and the Courts: A Positive Theory of Judicial Doctrine and the Rule of Law." *Southern California Law Review* 68 (September): 1631-1683.

Niblett, Anthony. 2009. “Tracking Inconsistent Judicial Behavior.” Manuscript. Law School, University of Chicago.

Niblett, Anthony. 2010. “Case-by-Case Adjudication and the Path of Law.” Manuscript. Law School, University of Chicago.

Niblett, Anthony, Richard A. Posner, and Andrei Shleifer. 2010. “The Evolution of a Legal Rule.” *Journal of Legal Studies* 39 (June): 325-358.

O’Hara, Erin. 1993. “Social Constraint or Implicit Collusion? Toward a Game Theoretic Analysis of Stare Decisis.” *Seton Hall Law Review* 24: 736-778.

Parisi, Francesco and Fon, Vincy. 2008. *The Economics of Lawmaking*. New York, NY: Oxford University Press.

Posner, Richard A. [1973] 2007. *Economic Analysis of Law*. 7th ed. Boston: Little Brown Press.

Posner, Richard A. 2006. “Reasoning by Analogy.” *Cornell Law Review* 91 (March): 761-774.

Priest, George L. 1977. “The Common Law Process and the Selection of Efficient Rules.” *Journal of Legal Studies* 6 (January): 65-82.

Radin, Max. 1925. “The Theory of Judicial Decision: Or How Judges Think.” *American Bar Association Journal* 11 (June): 357-362.

Rawls, John. 1971. *A Theory of Justice*. Cambridge Mass: Harvard University Press, revised edition.

Revesz, Richard L. 1997. “Environmental Regulation, Ideology, and the D.C. Circuit.” *Virginia Law Review* 83 (November): 1717–1772.

Rose, Carol M. 1988. “Crystals and Mud in Property Law.” *Stanford Law Review* 40 (February): 577-610.

Rubin, Paul H. 1977. “Why is the Common Law Efficient?” *Journal of Legal Studies* 6 (January): 51-63.

Schauer, Frederick. 2003. “The Convergence of Rules and Standards.” *New Zealand Law Review* 2003: 303-328.

Schwartzstein, Joshua, Andrei Schleifer. 2011. “An Activity-Generating Theory of Regulation.” Manuscript. Department of Economics, Dartmouth College.

Shavell, Steven. 2004. *Foundations of Economic Analysis of Law*. Cambridge: Harvard University Press.

Stephenson, Matthew C. 2009. “Legal Realism for Economists.” *Journal of Economic Perspectives* 23 (Spring): 191-211.

Songer, Donald R., Jeffrey A. Segal and Charles M. Cameron. 1994. "Testing a Principal Agent Model of Supreme Court-Circuit Court Interactions." *American Journal of Political Science* 38 (August): 673-696.

Spitzer, Matthew, and Eric Talley. 2011. "Left, Right, and Center: Strategic Information Acquisition and Diversity in Judicial Panels." *Journal of Law, Economics, and Organization* (September Advance Access): 1-43.

Stokey, Nancy, and Robert E. Lucas Jr. 1989. *Recursive Methods in Economic Dynamics.* Cambridge: Harvard University Press.

Stras, David R., and Shaun M. Pettigrew. 2010. "The Rising Caseload in the Fourth Circuit: A Statistical and Institutional Analysis." *Southern Carolina Law Review* 61 (Spring): 421-444.

Sunstein, Cass R. 1999. *One Case at a Time.* Cambridge: Harvard University Press.

Sunstein, Cass R. 2001. "Cost-Benefit Default Principles." *Michigan Law Review* 99 (June): 1651-1723.

Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki. 2006. *Are Judges Political? An Empirical Analysis of the Federal Judiciary.* Washington, DC: Brookings Institution Press.

Tamanaha, Brian Z. 2004. *On the Rule of Law: History, Politics, Theory.* Cambridge: Cambridge University Press.

Whittington, Keith E. 1999. *Constitutional Interpretation: Textual Meaning, Original Intent, and Judicial Review.* Lawrence: University of Kansas Press.

Zywicki, Todd. 2003. "The Rise and Fall of Efficiency in the Common Law: A Supply-Side Explanation." *Northwestern Law Review* 97 (Summer): 1551-1633.

Cases Cited

Arizona v. Gant, 556 U.S. 332 (2009).

B.M.W., Inc. v Gore, 517 U.S. 559 (1996).

New York v. Belton, 453 U.S. 454 (1981).

Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1 (1991).

State Farm Mutual Automobile Ins. Co. v. Campbell, 538 U.S. 408 (2003).

Thorton v. United States, 541 U.S. 615 (2004).

TXO Prod. Corp. v. Alliance Resource Corp., 509 U.S. 443 (1993).

THE UNIVERSITY OF
MELBOURNE

Department of Economics

Working Paper Series

A Theory of Rational Jurisprudence

Scott Baker and Claudio Mezzetti

May 2012

Research Paper Number 1144

ISSN: 0819-2642
ISBN: 978 0 7340 4494 5

Department of Economics
The University of Melbourne
Parkville VIC 3010
www.economics.unimelb.edu.au

A Theory of Rational Jurisprudence*

Scott Baker † and Claudio Mezzetti ‡

† Washington University in St. Louis
‡ University of Melbourne

May 7, 2012

Abstract

We examine a dynamic model of up-or-down problem solving. A decision maker can either spend resources investigating a new problem before deciding what to do, or decide based on similarity with precedent problems. Over time, a decision making framework, or jurisprudence, develops. We focus on the model's application to judge-made law. We show that judges summarily apply precedent in some cases. The law may converge to efficient or inefficient rules. With positive probability, identical cases are treated differently. As the court learns over time, inconsistencies become less likely. We discuss the existing empirical evidence and the model's testable implications.

Keywords: Law and Economics, Incompleteness of Law, Judge-Made Law, Evolution of Legal Rules.

JEL Classification Numbers: K10, K40.

1 Introduction

We develop a dynamic model of up-or-down problem solving. In it, a decision maker must devise a method for dealing with a series of problems, project proposals, or cases, with an up-or-down, yes-or-no, answer. As matters arise, the decision maker has a choice: He can spend time and resources investigating the merits in depth and, based on the results, decide what to

*We would like to thank the associate editor and two anonymous referees; their insightful comments have led to a substantial improvement of the paper. We also thank participants at workshops at Tilburg University, the University of Chicago, the University of Illinois, Northwestern University, Boston University, USC, the University of Amsterdam, Washington University, AETW 2012 at UNSW, the 2010 American Law and Economics Association Meetings, and the 2009 Triangle Workshop on Law and Economics.

do. Alternatively, the decision maker can decide that a case or project is "close enough" to prior cases or projects that it can be decided similarly without additional thought. Reasoning in this way saves on decision costs but creates the chance of error. Over time the decision maker develops a decision making framework, a jurisprudence. The paper focuses on the properties of this jurisprudence: Will a practice of following precedent emerge? Will "like" cases or projects be treated "alike"? Under what conditions will the case law converge to efficiency – where a party wins or a project is approved if and only if the benefits outweigh the costs?

The problem solving we model is something many organizations must do. CEOs and venture capitalists must decide which projects to move ahead and which projects to forgo; at the same time, they must decide how to spend limited resources deciding which projects go in which bin. In handling customer complaints, firms must decide which complaints should be investigated and which ones should be declined or approved based solely on similarity to past complaints.

Although we believe there are many applications, our primary motivation is the study of judge-made law, a subject rarely studied by economists.[1] In common law countries, judicial decisions set the rules for property, govern the interpretation of contracts, and control negative externalities through the allocation of liability. Further, because ex-ante specification of statutes is costly, legislatures often use statutory language which includes broad phrases or terms. In antitrust, for example, Section 2 of the Sherman Act imposes liability on "every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize."[2] In intellectual property, a patent will not issue if the difference between the prior art and the claimed invention would have been "obvious" at the time of the invention.[3] Courts, then, fill in the law on a case by case basis.

In the United States, constitutional law is judge-made. As a result, judicial decisions set the ground rules for the political process and define the state's reach into private lives. Even in areas of law where agency regulation dominates (say, environmental law and occupational health and safety) judges play a significant oversight role. For example, judges review an agency's cost-benefit analysis and can, in some scenarios, decide whether the agency must do cost-benefit analysis at all (Sunstein, 2001).

Case by case adjudication differs from ex ante legislative or agency rule making in fundamental ways. Unlike other law-making bodies, judges don't announce policies. They are

[1] For a recent paper outlining several reasons why economists need to pay closer attention to judicial behavior, see Stephenson (2009). An enhanced focus on judicial decision-making and common law evolution is also critical in light of the empirical findings of the legal origins literature (see La Porta et al., 1998, and La Porta et al., 2004).

[2] 15 U.S.C. § 2.

[3] 35 U.S.C. § 103.

reactive. Generally speaking, judges don't respond unless an injury materializes.[4] The injury leads to a complaint or case, with each case providing grounds for revising the judge's belief about the benefits and costs of a particular judge-made law. In deciding cases, judges have to consider that careful examination of a case is costly. They need to select which cases are important to look at closely and which cases are not. One lesson from our model is that this triage process – an important act in judging – has consequences for the ultimate policy and the consistency of judicial decision making along the way.

Just because case by case decision making is different from ex-ante direct regulation of behavior doesn't make it important. But there is an interaction between the effectiveness of judicial decision making and the legislature's choice to delegate – via a vague statute – to the courts. Detailed statutes are expensive to draft and inflexible. If judicial decision making is likely to be poor or mistaken, the legislature should be more willing to spend resources setting the contours of the statute in the first place, and vice versa.

Judicial decision making is likely to be poor when the court must spend lots of resources figuring out the correct answer to the resolution of a case.[5] This might happen in complex areas of the law, such as tax. In addition, when judicial caseload is high, courts will react to additional delegation (passage of another statute with vague language) by reading prior cases "broadly." The judge will extrapolate a great deal from prior cases to new cases with different facts; new cases where the outcome of the prior case is not terribly informative. As a result, the court will make more errors, simply as a consequence of having to adjudicate all cases on its docket.

On the other hand, case by case adjudication is well suited to uncover unintended side-effects. By observing materialized injuries, judges learn the extent of the injury before setting the doctrine. Judges can thus incorporate this information in the doctrine, adjusting the rules over time as more information comes available.

To be concrete, zoning makes sure everyone knows what they can and cannot do with their property ex ante and saves on litigation costs. Regulation via nuisance claims, by contrast, allows for ex-post considerations and, thus, for better tailoring of the regulation to situation specific circumstances. Which is better or should we have a little of both? In short, the choice of the degree of ex-post versus ex-ante regulation turns on how good judicial decision making is likely to be. The answer to that question requires an understanding of how judges make decisions with scarce resources, a decision making exercise which forms the basis of our inquiry.

[4] For an informal discussion of the factors making legal intervention before the harm relatively more efficient than legal intervention after the harm, see Shavell (2004, 572-578).

[5] For a model relating judicial errors and the desirability of ex ante regulation, see Schwartzstein and Shleifer (2011).

We do not model litigation explicitly. Instead, we focus on learning over time, having judges engage in a forward looking cost-benefit analysis in interpreting precedent, and what that analysis implies for the evolution of law. For simplicity, we assume that judges discover the facts of a case on their own. This is not incompatible with the typical role of advocacy. Indeed, we could generalize the model so as to let, say, the defendant's lawyers argue that the relevant set of precedents is the one leading to dismissal of the case, while the plaintiff's lawyers argue that another set of precedents govern, leading to the opposite conclusion. Such an extension would shed light on the role of advocacy and case selection, but it would not affect our conclusions on the evolution of law. When the arguments of one party are convincing, the judge would rule in his favor based on closeness with the set of precedents the party has presented; when a definite conclusion is unclear, the judge would invest additional resources to discover which sets of precedents govern.

To our knowledge, no one has formally examined situations where argument strength turns on distance. Yet, in judge-made law, that distance is critical. Our model demonstrates a link between "closeness" and the efficient deployment of resources. As resources become more scarce, a prior case can be further away from the case at hand and still persuade the decision maker that he should decide the case without additional investigation.

Using a dynamic programming model, we build a theory of judge-made law from the ground up. We assume a single court that lives forever, consisting of judges with identical policy preferences. The judges are initially uncertain about the consequences of legal rules and have scarce resources to investigate cases. The theory yields both surprising and intuitive results.

In the model, reliance on precedent arises endogenously. Judges follow precedent not because deviations are punished, but to conserve resources. Each period, the judge interprets the prior cases, deciding how far to extrapolate the results from prior cases to new and different circumstances, without engaging in any fresh investigation. Cases that are in some sense "close" to the prior case law are decided summarily based solely on the precedent. Cases that are "far" the judge investigates. Each period, the judge decides what is close and what is far.

In making this interpretative choice, the judge balances two costs. First, there are error costs, i.e., the costs of ruling on a case incorrectly. Second, there are decision costs, the costs associated with a judge investigating a case in depth instead of relying on the precedent as a proxy for what to do. We show that the optimal balancing of these costs formalizes the usefulness of reasoning by analogy (a skill taught to every lawyer and judge) and, in accord with intuition, shows that reliance on precedent will be lower in areas of law where errors are more costly.

We next demonstrate that the judge-made law will, in general, converge. This convergence will be of two types. If decision costs are small relative to fixed error costs, the law will converge to the efficient outcome, or correct decision, in all cases. More interesting, if decision costs are high relative to fixed error costs, the law will converge to an inefficient set of legal rules. Thus, for example, judge-made law will incorrectly specify liability for some activities where no liability is the proper result and no liability where liability is the right outcome. This result obtains even though all judges share efficiency as the goal. Convergence here is second-best: spending resources to gather more information – i.e., hearing more cases – is not worthwhile in terms of the benefit of a more accurate legal rule.

Finally, we explore the implications of our theory for the evolution of law. In spite of our assumption of identical policy preferences by all judges, we show that with positive probability the court will fail to treat like cases alike. Cases with "identical" relevant facts may be decided differently. Discriminatory treatment – violation of what we refer to as the likeness principle – occurs as the court uses what it has learned to improve the law. Judges are often vilified for treating like cases differently, actions thought unfair and inconsistent with the rule of law. The model shows that strict adherence to the likeness principle inhibits judicial learning and the cost-justified updating of legal rules. Violation of the likeness principle is apt to occur when investigation of a case yields a surprising result, and hence teaches a great deal to the court. We also consider whether inconsistencies will remain in the limit. The model demonstrates that as the court learns more and more, inconsistencies become less and less likely and, in the limit, vanish. Taken together, the two results suggests inconsistencies should be observed more often when the judge-made law is in its infancy, shortly, say, after Congress passes a statute containing broad, enabling language.

The paper unfolds as follows: Related literature is reviewed in the remaining part of this introduction. Section 2 develops an economic model of legal reasoning. Section 3 demonstrates that precedent has value and studies the convergence properties of doctrine. Section 4 shows that a rational court will, with positive probability, violate the likeness principle. Section 5 suggests possible extensions and offers some concluding remarks. Throughout, where applicable, we discuss the empirical evidence consistent with the predictions of the model and offer testable implications. Although we believe the model covers a common decision making phenomenon, one benefit of focusing on judge-made law is that courts provide a wealth of data for possible future empirical tests of the theory.

Literature Review. The model closest to ours is Gennaioli and Shleifer (2007). Seeking an explanation for the empirical finding that a common-law legal origin correlates with various markers of development, they build a model of judge-made law. A judge, in their model, cannot overrule prior cases, she can only distinguish a case by searching for a different dimension along which to consider it. The act of distinguishing two cases has social value because it embeds new information into the law. Different judicial policy preferences then shape the evolution of law. Gennaioli and Shleifer's main result is the "Cardozo Theorem." It says that the legal evolution induced by distinguishing cases will, on average, be beneficial irrespective of the amount of bias in the judiciary. We take a different approach. Rather than starting from the premise that judges have conflicting policy preferences, we begin from an assumption of scarce judicial resources. From this alternative baseline, we get insights consistent with the institutional features of judge-made law.

Two other significant literatures relate to our work. The first is from law and economics scholars, the second from political science. Since Judge Posner's assertion that the common law is efficient (Posner, [1973] 2007), the law and economics literature has sought to explain why this might be so (Rubin, 1977; Priest, 1977; Bailey and Rubin, 1994; Hylton, 2006). Both Judge Posner's assertion about efficient common law and the models exploring it have been sharply contested (Bailey and Rubin, 1994; Hadfield, 1992; Hathaway, 2001). The literature has blossomed with many factors pointing toward and against efficiency (Zwyicki, 2003; Klerman, 2007; Parisi and Fon, 2008). This literature is distinct from what we do here. Case selection drives the law in these models, with judges playing little role. Ours is not a story about litigants selecting specific cases for trial and that selection dictating the path of the law. Rather, we ask what evolution to expect when judges are forward looking and have the ability to learn, but are resource constrained.

There are a few exceptions to the litigant selection story about the evolution of law. Cooter et al. (1977) and Hadfield (2009) develop models where the court can learn and ask whether rules will adapt to new circumstances and/or converge to efficiency. Unlike our model, the question of how to optimally deploy judicial resources over many periods is not examined. Dari-Mattiacci et al. (2010) develop a dynamic model where the litigants bring information to the courts and the courts issue decisions. The number of judicial decisions implies more "precedent"; distinct from us, judges don't interpret prior case law in their model.

Political scientists assume that judges, like legislators, make decisions to advance their preferred policy objective.[6] Learning doesn't occur and the informational value of precedent is not explored. In contrast, we show how the shifting interpretation of precedent can be seen as a method of efficiently managing resources to learn about the proper structure of legal rules.

[6] For a model where a judge makes decisions anticipating the likely position of Congress or the executive, see Eskridge and Ferejohn (1992). For a model where the judge is influenced by other judges sitting on the panel, see Spitzer and Talley (2011). For a model where judges face constraints imposed by the likely position of the higher court, see Songer et al. (1994) and McNollgast (1995). For a model where judges interact repeatedly over time, see O'Hara (1993).

These aspects of judicial behavior have not, to our knowledge, been formally studied elsewhere.

2 A Model of Legal Reasoning

In creating law case by case, judges mix the information from prior cases with the facts from new cases. If the new facts indicate that, as stated, the legal doctrine no longer serves its function, the judge can distinguish the prior case law and reach an alternative resolution.[7] To capture this process, suppose a judge wishes to regulate a set of activities, $x \in [0, 1]$. Activity x carries costs and benefits and there is a threshold $\theta \in [0, 1]$, below which an activity is socially valuable and above which it is not. The threshold point θ is initially unknown. We model it as random variable distributed according to $F(\theta)$ with positive density $f(\theta)$ on $[0, 1]$. The restriction of activities to the unit interval is with some loss of generality. At a higher level of abstraction, an activity might be represented as an element in a multi-dimensional space.[8] Our assumption requires that each multi-dimensional activity can be reduced to a single number and ranked against all other activities (i.e., the social value of activity 1 is lower than that of activity 3/4 which is lower than that of activity 1/2). It is akin to the common assumptions that guarantee the existence of a utility function in consumer theory.[9] Given this ordering, finding the set of socially valuable, or efficient, activities reduces to locating θ, the activity whose social value is zero.

As noted in the introduction, the court consists of a single, infinitely-lived judge with constant preferences.[10] The judge prefers to allow beneficial activities and deter harmful ones. When presented a case, the judge can either fully investigate or summarily examine it. “Summarily” in this model means spending less effort on the case, by, for example, resolving the case with an unpublished decision or one line order.

Full investigation costs C and is rewarded: the judge discovers whether undertaking activity x is, or is not, efficient; that is, whether it carries a positive or negative social value.[11,12] The

[7] Levi (1949, 7-19) contains the classic discussion of common law reasoning.

[8] The assumption that the region of uncertainty is bounded is not important. It will be used when we work with a uniform distribution, but could be dispensed with in the general model.

[9] Completeness, transitivity and continuity are sufficient; they guarantee that activities associated with nearby real numbers have nearby characteristics from the point of view of case law. Since judges have identical preferences, they collapse the fact space into a single dimension in the same way and rank the resulting bundles the same.

[10] The conclusion suggests an alternative interpretation of the model, consisting of two courts in a hierachy, where the upper court must decide how many cases whose outcome is uncertain to delegate to the lower court to resolve.

[11] It simplifies the analysis to assume that there is no noise or mistake in the discovery process, and that the judge correctly determines whether an activity is efficient upon its full investigation.

[12] The assumption that decision costs are constant seems a reasonable first approximation; it is not immediate whether investigation costs are lower or higher the closer is x to θ.

judge reports what he has learned in an opinion. Summary examination, by contrast, is costless, but might result in an incorrect decision.

Full investigation means studying the facts. It requires the court and the litigants to spend time and resources uncovering the relationship between the new facts and broader social policies, using all the judicial tools (oral argument, additional briefing, reading the scholarly literature, closely examining expert reports, etc.). Investigation costs are likely to be high in technical fields, like tax, environmental harms, and food and drug safety. In these areas, the parties must provide and the judge must assess complicated testimony and materials to see the likely consequences of ruling for the plaintiff or the defendant. Overall caseload also impacts decision costs. As caseloads go up, the judge's opportunity cost of investigating any single case increases. As we shall see, using caseload as a proxy for decision costs enables us to derive testable predictions from the model and link the theoretical results to the empirical literature on judicial decision making.

Two types of potential mistakes are associated with summary examination. The court might erroneously declare a socially beneficial activity impermissible or a socially harmful activity permissible. Mistakes in adjudication result in two kinds of losses. There is a variable loss $\ell\,|x-\theta|$ which depends on the distance between the activity and θ, and a fixed error loss L, which does not depend on the distance. The variable loss captures the idea that the further away the activity is from the threshold θ the higher is the social loss from a mistaken decision. The fixed error loss captures a discontinuity in the social value of activities. Judges care about perception, both from the public and from the other branches. In many cases, the public, for example, is unlikely to understand the difference between a "small" mistake and a "big" mistake and calibrate their reaction to the court decision accordingly. As a result, one might think about the fixed error cost as the discrete loss arising from the court making the wrong call.[13]

Error costs are likely to be high when other actors lack the capacity or will to correct judicial mistakes. Three examples are: constitutional cases, when the courts have the final say about what the constitution requires; cases when interest group conflict is rampant and, as a result, Congress rarely acts; courts decisions on the scope of a mandatory rule, that is, a rule that cannot be contracted around.

[13] Nothing in the model turns on including a fixed cost. Assuming that there is a fixed social loss from error stacks the deck in favor of efficient convergence. As noted above, a major claim in the law and economics literature is that judge-made law is efficient. Inefficiency is easily shown if the only social loss of incorrect decisions is the variable loss, which shrinks as activities get closer to θ. With a constant decision cost, investigation would, over time, become less and less worthwhile (the benefit of preventing mistakes would fall, while the cost remains constant). In the long run, judge-made law would surely admit errors. Interestingly, as we shall see, with a fixed error cost component convergence to inefficient rules still can occur.

2.1 Timing and The Construction of Precedent

Each period a case x is randomly selected from the interval $[0, 1]$ according to the distribution $G(x)$ with positive density $g(x)$ and brought to the attention of the court.[14] Let W_t denote the highest activity or fact pattern that the court has, as of time t, heard, fully investigated, and found acceptable, that is, of positive social value. Let R_t denote the lowest activity or fact pattern that the court has fully investigated and found unacceptable, that is, of negative social value. Prior case law, in other words, has taught that allowing activities in the interval $[0, W_t]$ is efficient, while allowing activities in $[R_t, 1]$ is inefficient. The two endpoints W_t and R_t squeeze the court's beliefs about the distribution of θ. The range of activities the court knows nothing about is (W_t, R_t).

Each period the court decides, based on prior precedent (i.e., looking at W_t and R_t), which cases to look closely at and which cases to decide summarily. Cases that are in some sense "close" to the prior precedent are decided summarily based on the precedent alone. Cases that are in some sense "far" from prior precedent are investigated. The court can easily determine the proper resolution of cases below W_t and above R_t, since the prior precedent is perfectly informative about them. The court summarily declares permissible any case below W_t and impermissible any case above R_t. On these cases, the judge expends no decision costs and makes no errors. For cases in the interval (W_t, R_t), the court determines how broadly or narrowly to construe past precedent. Reading precedent broadly means that the court decides a high percentage of new cases by extrapolating or reasoning by analogy from past cases without further inquiry (that is, lots of cases are determined to be "close" to the precedent). Given the greater extrapolation, the chance is higher that the prior cases will be off point and lead the judge to a mistaken resolution. Reading precedent narrowly is the opposite.

The interpretative decision determines two bounds, a_t and b_t. These bounds partition the interval of uncertainty (W_t, R_t) into three areas. The first area is the interval $[a_t, b_t]$. In this interval, the judge fully investigates the case. The second and third areas are (W_t, a_t) and (b_t, R_t). If a case lies in either of these intervals, the judge feels the activities are close enough to prior case law (W_t and R_t) as to be decided by application of precedent alone without spending effort. Activities in (W_t, a_t) are declared permissible, activities in (b_t, R_t) are declared impermissible. The size of these two intervals measures of how expansively the judge reads the prior precedent

If the judge investigates and learns that the activity in the case has positive value, she

[14] In specifying that the court always draws facts from the same distribution, we abstract away from the law's impact on behavior. We do this to ease the analysis and focus on judicial learning. The assumption is a reasonable first approximation, so long as parties make mistakes about the contours of the law when deciding their actions, or face a small probability of getting caught and sued.

writes an opinion and updates the precedent stock ($W_{t+1} > W_t$). Likewise, if the judge investigates and finds the activity to have negative value, she updates the precedent stock accordingly ($R_{t+1} < R_t$). If the new case is summarily examined, the court learns nothing and the precedent stock remain constant.

As an example motivating our model, consider punitive damage awards in tort cases. The Supreme Court has the power to strike down punitive awards set by lower courts as violating the due process clause of the Constitution.[15] The legal issue is how much is too much – what multiple of compensatory damages renders the award unconstitutional. Suppose awards range between 0 and 1000 times the compensatory award. Imagine the first case the Supreme Court hears involves a punitive award of 500 times the compensatory award. The Court investigates, finds this multiple too much and strikes the award down. In the language of the model, the Court learns that $\theta < \frac{500}{1000} = .5$. It writes an opinion saying so, setting $R_t = .5$.

Suppose that the next case involves a multiple of 100. The Court might interpret its prior case to also cover this case and summarily strike down the award (setting $b_t = .1$). In so doing, the Court broadly construes its own precedent. Such a move saves resources, but might be wrong. If the ideal rule θ lies in the interval $(.1, .5)$, the court will have made a mistake. Now suppose instead that the second case has a multiple of 450. Striking this award down summarily is less likely to generate a mistake. It only does so if θ happens to be in the interval $(.45, .5)$. Because of the "closeness" between the 450 case and the 500 precedential case, relying on precedent saves resources without unduly increasing error costs. In our model, in each period the judge interprets precedent to balance the two costs, while understanding that the investigation today benefits future judges because it reveals information about the optimal legal rule.[16]

2.2 The Optimization Problem

We can now write the dynamic optimization problem that the court must solve. Let δ be the discount factor and $V(W_t, R_t)$ be the court's value function at time t, with state variables W_t, R_t. The court chooses the interpretative bounds a_t, b_t subject to $W_t \leq a_t \leq b_t \leq R_t$ to

[15] See Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1 (1991).

[16] This example is drawn from the case law. See Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 23 (1991) (finding that a ratio of four to one between punitive damages and compensatory damages was not excessive); TXO Prod. Corp. v. Alliance Resource Corp., 509 U.S. 443, 462 (1993) (noting that a 10 to 1 ratio between the punitive award and the harm if the illicit plan had succeeded would not "jar one's constitutional sensibilities"); B.M.W., Inc. v. Gore, 517 U.S. 559, 582 (1996) (finding that an award of 500 times the actual harm was grossly excessive); State Farm Mutual Automobile Ins. Co. v. Campbell, 538 U.S. 408, 429 (2003) (finding that an award of 145 to 1 was an "irrational and arbitrary deprivation of the property of the defendant"). For other examples of this kind of evolution of law, see Niblett (2010).

maximize its expected discounted payoff:

$$\begin{aligned} V(W_t, R_t) = \max_{W_t \le a_t \le b_t \le R_t} \Big\{ & \delta E_t V(W_{t+1}, R_{t+1}) - C\left[G(b_t) - G(a_t)\right] \\ & - \int_{W_t}^{a_t} \int_{W_t}^{x_t} \frac{Lf(\theta)}{F(R_t) - F(W_t)} d\theta g(x_t) dx_t - \int_{b_t}^{R_t} \int_{x_t}^{R_t} \frac{Lf(\theta)}{F(R_t) - F(W_t)} d\theta g(x_t) dx_t \\ & - \int_{W_t}^{a_t} \int_{W_t}^{x_t} \frac{\ell\left(x_t - \theta\right) f(\theta)}{F(R_t) - F(W_t)} d\theta g(x_t) dx_t - \int_{b_t}^{R_t} \int_{x_t}^{R_t} \frac{\ell\left(\theta - x_t\right) f(\theta)}{F(R_t) - F(W_t)} d\theta g(x_t) dx_t \Big\}. \end{aligned} \tag{1}$$

The second term is the expected cost of having to decide a case in period t. For example, if $a_t = b_t$, then the court construes the prior precedent as deciding the law for all activities and does not incur any decision costs at time t. The greater the distance between a_t and b_t, the greater the chance a case is drawn where the court views the law as unsettled by prior precedent and is willing to expend effort.

The third and fourth term reflect the expected one-period losses due to the fixed error component L, while the fifth and sixth term reflect the expected losses due to the variable error component $\ell\left|x - \theta\right|$. Consider the third and fifth term. If the judge sets $a_t \geq W_t$, there is a chance the case drawn x_t is between W_t and a_t. For cases in this gap, the court will base its decision solely on the prior precedent and rule summarily that the activity is permissible. The expression $\frac{f(\theta)}{F(R_t) - F(W_t)}$ is the probability the court attaches to the possibility that θ is less than x_t. In that event, x_t creates a social loss; the court has made a mistake in its summary ruling. The fourth and sixth term follow from a similar analysis on the upper region of the interval of uncertainty; here precedent induces the court to rule the activity as impermissible, but, in fact, it is socially valuable.

The first term in (1) is the discounted expectation of the value of the court's objective function at the end of period t, given its interpretative choices at time t. This term captures the dynamic learning considerations described above. It can be written explicitly as the sum of three components:

$$\begin{aligned} E_t V(\cdot) = {} & V(W_t, R_t)\left[1 - (G(b_t) - G(a_t))\right] \\ & + \int_{a_t}^{b_t} V(W_t, x_t) \frac{F\left(x_t\right) - F(W_t)}{F(R_t) - F(W_t)} g(x_t) dx_t + \int_{a_t}^{b_t} V(x_t, R_t) \frac{F\left(R_t\right) - F(x_t)}{F(R_t) - F(W_t)} g(x_t) dx_t. \end{aligned} \tag{2}$$

The first component is the current value function times the probability that no learning takes place because the randomly selected activity x is outside the interval of investigation $[a_t, b_t]$. The second component is the expected value function when the case x is brought to court, investigated upon, and determined to be above θ; in such an instance the new interpretative interval becomes $[W_t, x_t]$. The third component is the expected value function when x is discovered to be below θ.

A special version of the model is when the distributions $F(\theta)$ and $G(x)$ are both uniform. This version has the advantage of simplifying the analysis; only the size $\Delta_t = R_t - W_t$ of the interval (W_t, R_t) matters to the court when deciding the interpretative bounds. It is convenient to express the objective function of the court as a function of the proportion of cases α_t in the interval Δ_t that the court considers permissible and the proportion β_t that it considers impermissible after summary investigation:

$$\begin{aligned} \alpha_t &= \frac{a_t - W_t}{\Delta_t} \\ \beta_t &= \frac{R_t - b_t}{\Delta_t}. \end{aligned}$$

Thus, the proportion of cases in Δ_t that are investigated is $1 - \alpha_t - \beta_t$. It is immediate that the following restrictions must hold: $\alpha_t \geq 0$, $\beta_t \geq 0$, $1 - \alpha_t - \beta_t \geq 0$. We will use this uniform version of the model in the remainder of the paper. The appendix shows that the main results and insights extend to the general version.

Lemma 1 in the appendix proves that the value function $V(\Delta_t)$ for the uniform version exists, is unique, continuously differentiable and can be written as:

$$\begin{aligned} V(\Delta_t) = \max_{\alpha_t \geq 0, \beta_t \geq 0, 1-\alpha_t-\beta_t \geq 0} \Bigg\{ & -C\left(1-\alpha_t-\beta_t\right)\Delta_t - \frac{L}{2}\left(\alpha_t^2+\beta_t^2\right)\Delta_t \\ & - \frac{\ell}{6}\left(\alpha_t^3+\beta_t^3\right)\Delta_t^2 + \delta V(\Delta_t)\left[1-\left(1-\alpha_t-\beta_t\right)\Delta_t\right] \\ & + \delta \int_{\alpha_t}^{1-\alpha_t} V(\Delta_t x_t)\Delta_t x_t dx_t + \delta \int_{\beta_t}^{1-\beta_t} V(\Delta_t x_t)\Delta_t x_t dx_t \Bigg\}. \end{aligned} \tag{3}$$

Lemma 1 also shows that $V(\Delta)$ is negative and decreasing in Δ, with $V(0) = 0$.

3 The Value of Precedent and the Convergence of Doctrine

The most narrow construction of precedent in period t sets $(a_t, b_t) = (W_t, R_t)$, or $\alpha_t = \beta_t = 0$. In so doing, the court maximizes the learning that case load provides by looking at every possible case where the resolution is uncertain; each of these cases carries a bit of information about the location of θ. If the decision cost C of looking at a case is sufficiently small relative to the error losses, L and ℓ, shouldn't the court use all the information potentially available in each period, setting $\alpha_t = \beta_t = 0$? As we shall see in this section, while this approach seems like a good idea, it never is, no matter how small C is relative to L and ℓ.

Proposition 1 shows that the constraints $\alpha_t \geq 0$ and $\beta_t \geq 0$ never bind if $\Delta_t > 0$ – that is, a positive proportion of cases is always summarily examined. It also proves that the solution is symmetric, $\alpha_t = \beta_t$. The court always relies on precedent and never investigates every case in the interval of uncertainty Δ_t. On the other hand, the constraint $1 - \alpha_t - \beta_t \geq 0$ may bind; when this happens $\alpha_t = \beta_t = \frac{1}{2}$ and no cases are fully investigated. In legal terms, the court construes the prior precedent as covering all potential future cases.

Define the precedent ratio λ_t as the proportion of cases in the interval of uncertainty Δ_t that will be ruled by precedent in period t:

$$\lambda_t = \alpha_t + \beta_t.$$

As the bounds a_t and b_t get closer for a given level of uncertainty, λ_t increases, meaning that the court decides a higher proportion of cases by reference to precedent alone.

Proposition 1. *There exists $\overline{\lambda}(\Delta_t) > 0$ such that in each period t when the interval of uncertainty is Δ_t the court chooses $\alpha_t = \beta_t = \frac{\lambda_t}{2} \geq \frac{\overline{\lambda}(\Delta_t)}{2}$, and thus selects a precedent ratio λ_t bounded away from zero. The precedent ratio λ_t is an increasing function of C_t and a decreasing function of L_t, ℓ_t and δ_t.*[17]

Proof. Define

$$\begin{aligned} V^\alpha(\Delta_t) = &-C\left(\frac{1}{2} - \alpha_t\right)\Delta_t - \frac{L}{2}\alpha_t^2\Delta_t - \frac{\ell}{6}\alpha_t^3\Delta_t^2 \\ &+ \delta V(\Delta_t)\left[\frac{1}{2} - \left(\frac{1}{2} - \alpha_t\right)\Delta_t\right] + \delta\int_{\alpha_t}^{1-\alpha_t} V(\Delta_t x_t)\Delta_t x_t dx_t. \end{aligned}$$

$$\begin{aligned} V^\beta(\Delta_t) = &-C\left(\frac{1}{2} - \beta_t\right)\Delta_t - \frac{L}{2}\beta_t^2\Delta_t - \frac{\ell}{6}\beta_t^3\Delta_t^2 \\ &+ \delta V(\Delta_t)\left[\frac{1}{2} - \left(\frac{1}{2} - \beta_t\right)\Delta_t\right] + \delta\int_{\beta_t}^{1-\beta_t} V(\Delta_t x_t)\Delta_t x_t dx_t. \end{aligned}$$

The court's objective function (3) can be written as

$$V(\Delta_t) = \max_{\alpha_t \geq 0,\ \beta_t \geq 0,\ 1-\alpha_t-\beta_t \geq 0} V^\alpha(\Delta_t) + V^\beta(\Delta_t). \tag{4}$$

Note that $V^\alpha(\Delta_t)$ does not depend on β_t and $V^\beta(\Delta_t)$ does not depend on α_t. Moreover,

[17] In the proof, we derive analytical solutions for the comparative statics when the parameters C, L, ℓ and δ only change at time t. (This is why the subscript t is attached to the parameters in the statement of the proposition). Matlab simulations show that the sign of each comparative statics result is the same if we vary the relevant parameter at each point in time.

$V^{\beta}(\Delta_t) = V^{\alpha}(\Delta_t)|_{\alpha_t=\beta_t}$; that is, after switching the independent variable α_t with β_t, the functions $V^{\alpha}(\Delta_t)$ and $V^{\beta}(\Delta_t)$ are identical. Finally, recall that $V(\Delta)$ is negative, decreasing in Δ and note that $\lim_{\Delta\to 0} V(\Delta_t) = \lim_{\Delta\to 0} V^{\alpha}(\Delta_t) = \lim_{\Delta\to 0} V^{\beta}(\Delta_t) = 0$. Note also that:

$$\frac{\partial V^{\alpha}}{\partial \alpha_t} = \Delta_t \left[C - L\alpha_t - \frac{\ell}{2}\alpha_t^2 \Delta_t + \delta V(\Delta_t) - \delta(1-\alpha_t) V(\Delta_t(1-\alpha_t)) - \delta\alpha_t V(\Delta_t \alpha_t) \right] \quad (5)$$

$$\frac{\partial V^{\beta}}{\partial \beta_t} = \Delta_t \left[C - L\beta_t - \frac{\ell}{2}\beta_t^2 \Delta_t + \delta V(\Delta_t) - \delta(1-\beta_t) V(\Delta_t(1-\beta_t)) - \delta\beta_t V(\Delta_t \beta_t) \right]. \quad (6)$$

Let $\mu_t^{\alpha}, \mu_t^{\beta}$ and ρ_t be the (non-negative) multipliers on the constraints $\alpha_t \geq 0$, $\beta_t \geq 0$, and $1 - \alpha_t - \beta_t \geq 0$, respectively. The first order conditions of program (4) with respect to α_t and β_t are:

$$\frac{\partial V^{\alpha}(\Delta_t)}{\partial \alpha_t} + \mu_t^{\alpha} - \rho_t = 0 \quad (7)$$

$$\frac{\partial V^{\beta}(\Delta_t)}{\partial \beta_t} + \mu_t^{\beta} - \rho_t = 0 \quad (8)$$

We now show that α_t and β_t (and hence λ_t) are bounded away from zero. Suppose that $\alpha_t = 0$. Then (7) and (5) imply

$$0 < \Delta_t C \leq \Delta_t C + \mu_t^{\alpha} = \rho_t.$$

By complementary slackness, the constraint $1 - \alpha_t - \beta_t \geq 0$ must bind and hence $\beta_t = 1$. As a consequence, it is $\mu^{\beta} = 0$ and, using (8) and (6):

$$\Delta_t \left[C - L - \frac{\ell}{2}\Delta_t \right] = \rho_t,$$

a contradiction since $\mu_t^{\alpha} \geq 0$. This shows that α_t cannot be zero. Indeed, since $\lim_{\alpha\to 0} \frac{\partial V^{\alpha}}{\partial \alpha_t} = \Delta_t C$, for α_t "arbitrarily close" to zero it must be $\rho_t \simeq \Delta_t C > 0$ and $\beta_t = 1 - \alpha_t$. In addition, it must be $\rho_t \simeq \lim_{\beta\to 1} \frac{\partial V^{\beta}}{\partial \beta_t} = \Delta_t \left[C - L - \frac{\ell}{2}\Delta_t \right]$. Since this is a contradiction, α_t must be bounded away from zero; that is, there must be a lower bound $\overline{\alpha}(\Delta_t) > 0$ such that $\alpha_t > \overline{\alpha}(\Delta_t)$. The argument that β_t is bounded away from zero is analogous.

We now show that it must be $\alpha_t = \beta_t$ (this also implies $\lambda_t > \overline{\lambda}(\Delta_t) = 2\overline{\alpha}(\Delta_t)$). We have already seen that the constraints $\alpha_t \geq 0$ and $\beta_t \geq 0$ are slack at the solution of the maximization problem in the right hand side of (4). Since $V^{\alpha}(\Delta_t)$ depends only on α_t and $V^{\beta}(\Delta_t)$ depends only on β_t, if the constraint $1 - \alpha_t - \beta_t \geq 0$ is also slack, then (4) can be written as

$$V(\Delta_t) = \max_{\alpha_t} V^{\alpha}(\Delta_t) + \max_{\beta_t} V^{\beta}(\Delta_t).$$

It follows from $V^{\beta}(\Delta_t) = V^{\alpha}(\Delta_t)|_{\alpha_t=\beta_t}$, that at the optimum it must be $\alpha_t = \beta_t$. It remains to consider the case in which the constraint $1 - \alpha_t - \beta_t \geq 0$ is binding. Replacing $\beta_t = 1 - \alpha_t$ in (4) and recalling that the constraints $\alpha_t \geq 0$ and $1 - \alpha_t \geq 0$ are slack, we can write:

$$V(\Delta_t) = \max_{\alpha_t} \left\{ -\frac{L}{2} \left(\alpha_t^2 + (1-\alpha_t)^2 \right) \Delta_t - \frac{\ell}{6} \left(\alpha_t^3 + (1-\alpha_t)^3 \right) \Delta_t^2 + \delta V(\Delta_t) \right\}.$$

The first order condition gives

$$-L\left(\alpha_t - (1-\alpha_t)\right)\Delta_t - \frac{\ell}{2}\left(\alpha_t^2 - (1-\alpha_t)^2\right)\Delta_t^2 = 0,$$

from which we may conclude that the solution is $\alpha_t = \beta_t = 1/2$.

We now derive the comparative statics results. Small changes in C, L, ℓ and δ have no impact on $\lambda_t = \alpha_t + \beta_t$ in the case of a boundary solution $\alpha_t = \beta_t = 1/2$. Consider an interior solution. Exploiting the symmetry of the solution (i.e., $\alpha_t = \beta_t = \lambda_t/2$) and (5), and using the subscript t to keep track of the fact that we are only changing C, L, ℓ, δ at time t, we may write the interior first order condition $\frac{\partial V^{\alpha}}{\partial \alpha_t} = 0$ as:[18]

$$\Phi(\cdot) := C_t - \frac{L_t}{2}\lambda_t - \frac{\ell_t}{8}\lambda_t^2\Delta_t + \delta_t V(\Delta_t) - \delta_t\left(1 - \frac{\lambda_t}{2}\right)V\left(\Delta_t\left(1-\frac{\lambda_t}{2}\right)\right) - \delta_t\frac{\lambda_t}{2}V\left(\Delta_t\frac{\lambda_t}{2}\right) = 0. \tag{9}$$

Let Φ_{z_t} be the partial derivative of Φ with respect to the variable $z_t \in \{C_t, L_t, \ell_t, \delta_t\}$. By totally differentiating (9), it is immediate that $\partial\lambda_t/\partial z_t = -\Phi_{z_t}/\Phi_{\lambda_t}$. Since the second order condition of the court's maximization problem requires $\Phi_{\lambda_t} < 0$, the sign of $\partial\lambda_t/\partial z_t$ is the same as the sign of Φ_{z_t}. Observe that

$$\begin{aligned}
\Phi_{C_t} &= 1 > 0 \\
\Phi_{L_t} &= -\frac{\lambda_t}{2} < 0 \\
\Phi_{\ell_t} &= -\frac{\lambda_t^2}{8}\Delta_t < 0 \\
\Phi_{\delta_t} &= V(\Delta_t) - \left(1-\frac{\lambda_t}{2}\right)V\left(\Delta_t\left(1-\frac{\lambda_t}{2}\right)\right) - \frac{\lambda_t}{2}V\left(\Delta_t\frac{\lambda_t}{2}\right) < 0,
\end{aligned}$$

[18] In equation (9), $V(\cdot)$ is the value function at time $t+1$, which does not depend on $C_t, L_t, \ell_t, \delta_t$, the values of the parameters at time t. When the parameters, and hence the value function, vary in time, Lemma 1 can be extended to show existence, uniqueness, differentiability and negative valuedness of the value function at each point in time.

where the fourth inequality follows from $V(\Delta)$ being negative and decreasing in Δ.[19,20] ■

The value of precedent, *stare decisis*, emerges endogenously in our model. Proposition 1 formalizes a well-known view on precedent. Assuming the prior judgments were correct, the court can take those rulings as given and focus on new issues. As pointed out by Judge Benjamin Cardozo "the labor of judges would be increased almost to the breaking point if every past decision could be reopened in every case, and one could not lay one's own course of brick on the secure foundation of the courses laid by others who had gone before him." (Cardozo, 1921, 249).

The reason the court relies on precedent is simple. At the margin the cost of looking at cases near the boundary of the interval of uncertainty always outweighs the expected social loss of relying on precedent instead. To see this, suppose the court considers the marginal impact of relying on precedent and setting $\alpha_t = \beta_t = \varepsilon/2$ and hence $\lambda_t = \varepsilon$, rather than $\alpha_t = \beta_t = \lambda_t = 0$, where ε is a small number. The probability that x is drawn from the set $(W_t, a_t) \cup (b_t, R_t)$ and hence summarily decided is ε, while the probability that x is in the set and erroneously decided (i.e., $x > \theta$ in (W_t, a_t) or $x < \theta$ in (b_t, R_t)) is proportional to ε^2. It follows that the marginal benefit of relying on precedent (i.e., the expected saving on the decision cost) exceeds the marginal cost (i.e., the expected losses from an error and from giving up learning), for a sufficiently small ε.

The narrowest possible construction or interpretation of prior precedent ($\lambda_t = 0$), means that the court expends effort also on cases close to the boundary points W_t and R_t where an error is extremely unlikely. To tie with our motivating example, the Supreme Court investigates and hears a case involving a multiple of 499 times the compensatory award when it has previously found a 500 multiple impermissible. That is a waste of judicial resources. The court isn't relying enough on reasoning by analogy, i.e., extrapolating costs and benefits from similar past cases.

These results shed light on a number of phenomena. First, the judge, or decision maker, asks how close the case is to previous cases and, if sufficiently close, decides according to the precedent. The model justifies this common form of analogical analysis, which is deployed in

[19] Since $V(\Delta)$ is the value function at time $t+1$ and $C_{t+1} = C_{t+2} = ... = C$, Lemma 1 applies and $V(\Delta)$ is decreasing in Δ.

[20] When considering a change in one of the parameters C, L, ℓ, δ at each point in time, we must account for the impact of the change on the future value function. For example, differentiating (9) with respect to $C = C_t = C_{t+1} = ...$ yields:

$$\Phi_C = 1 + \delta \frac{\partial V(\Delta_t)}{\partial C} - \delta \left(1 - \frac{\lambda_t}{2}\right) \frac{\partial V\left(\Delta_t \left(1 - \frac{\lambda_t}{2}\right)\right)}{\partial C} - \delta \frac{\lambda_t}{2} \frac{\partial V\left(\Delta_t \frac{\lambda_t}{2}\right)}{\partial C}.$$

Matlab simulations show that the sign of Φ_z is the same as the sign of Φ_{z_t} for all $z \in \{C, L, \ell, \delta\}$.

many institutional settings. Although such reasoning conserves resources, there are limits on how far the analogy can be pushed. Some cases are just too different and, as a result, the prior decision is not terribly informative. Then, the analogy fails and is replaced by a fresh look.

Second, legal argument often involves statements about how close the case is to the prior case law. In our model, the closer the case is, the more likely the court will be to defer to the precedent rather than take a fresh look. Given this feature, effective advocacy involves explaining to the court why the case is closer to one set of precedential materials than to a second set of precedential materials. Indeed, judges often must decide which set of precedents more aptly apply in a new case (Posner, 2006, 63, quoting Radin, 1925, 359). In our model, this corresponds to deciding whether x_t is closer to the precedent associated with W_t or the one associated with R_t. We assume the court knows where x_t lies on the interval, but we could expand the model to allow lawyers on one side to argue that x_t is close to the precedent case W_t and hence is a valuable activity which ought to be declared permissible, while lawyers on the other side would argue that x_t is closer to R_t, the other precedent, and ought to be impermissible.

Third, reliance on precedent occurs even when the court faces no penalty for failing to follow precedent. It thus provides one explanation for why the Supreme Court defers to its own prior decisions and federal circuit courts defer to sister circuits or state courts, when they face no sanction for a failure to do so.

3.1 Precedent Ratio: The Empirical Literature and Testable Implications

Prior models of judicial decision making assume that judges do not learn and have unchanging preferences over how to divide a factual space into segments: permissible and impermissible (e.g., see Kornhueser, 1992a, 1992b and Lax, 2011, 2012). These models implicitly assume the judge is indifferent between ruling by summary disposition or in a lengthy opinion; they cannot explain why or when judges will find precedent persuasive. They also cannot explain the existence or form of legal argument and why closeness between the current case and prior cases seems to matter in the law.

Our model is different. For example, by Proposition 1 an increase in C raises the court's cost to learn about the social value of a particular activity by investigating the merits of the case. Accordingly, it encourages the court to reason by analogy to a greater extent and to resolve more cases by summary disposition. This prediction is consistent with empirical evidence. Huang (2011) studies the effect of increased caseloads on the rate at which circuit courts reverse district courts. As noted, an increase in caseload can be thought of as an increase in the decision cost of a judge. After September 11, 2001, the Attorney General made it a priority to quickly clear the deportation backlog. This move caused a flood of immigration appeals directly to two circuit courts. Huang looks at reversal rates in another kind of case – "civil" cases – in these jurisdictions. He finds that "when flooded by the agency cases, the affected circuit courts began to reverse district court rulings less often – in the civil cases. In these circuits, it seems, deference increased." (Huang, 2011, 1115). Enhanced deference can be viewed as greater use of summary review by the circuit court.

On related lines, Epstein, Landes, and Posner (2011, 101) find that "[i]n the court of appeals, the frequency of dissents is negatively related to the caseload." A dissent is more likely in a published opinion where the judges, collectively, have spent effort looking at a case and disagreements remain. Stras and Pettigrew (2010) examine a single circuit's response to a rising caseload. They report that, as the caseload has increased, the Fourth Circuit "adopted certain procedural reforms to adapt to its increased caseload. The most important and controversial of these reforms is a reduction in the percentage of cases allotted oral argument time and an increase in the percentage of cases decided through unpublished opinions." (Stras and Pettigrew, 2010, 432).

Next consider some testable implications of the comparative static on error costs. As the cost of a mistake increases, the court is less likely to decide by applying precedent alone and hence less likely to decide by summary disposition. Thus, we can empirically sort cases by issue area and then ask about the chance of summary disposition, given our expectation of the likely error cost.[21] Courts, say, should be less likely to resolve constitutional law cases without a hearing or without a written decision – a statement that can be tested. In capital punishment cases, the courts should routinely grant oral argument. Contract cases, by contrast, should often be treated summarily, especially when they involve the setting of default rules.

Finally, consider a decrease in discount factor, δ. Proposition 1 predicts that judges who care less about the future should be more likely to rely on summary disposition. Like the error cost comparative statics, this result can be tested. Some state court judges are elected; others are appointed for a given time period. Holding all else constant, elected state court judges might be assumed to care less about the distant future and learning about the proper scope of the law; as a result, they ought to be more likely to resolve cases summarily.

Given the wealth of data on judicial decision making, we believe that looking at courts provides a nice place to test our more general theory of the mechanics behind up-or-down problem solving.

[21] Of course, any empirical test based on the chance of summary disposition must account for the probability of filing by the litigants. Knowing that the case is likely to be summarily resolved, the plaintiff might not file in the first place, which, in turn makes observing summary disposition less likely.

3.2 Convergence

Legal academics, policy-makers, and advocates often criticize the law articulated by courts as imperfect or wrong-headed. For example, in tort law a defendant will be found negligent if he acts without reasonable care. The negligence standard is a knife-edge inquiry. If the defendant is found negligent, he is liable for all the resulting damage. If he is found non negligent, he pays nothing. Calfee and Craswell (1984) show that, when the defendant is uncertain about the legal standard, negligence can result in too much deterrence. The courts have not fine-tuned negligence law to account for the risk of over-deterrence identified by Calfee and Craswell. On this score, negligence law is imperfect.

As we shall see, our model shows that imperfections in doctrine are inevitable when the cost of deciding cases is sufficiently high. This is true even if the court shares the underlying values of those criticizing the decisions. The next proposition demonstrates that the court will eventually stop learning and will exclusively rely on precedent if and only if $C > L/2$. In such a case, at some point, the benefits of further refinement – tweaking the doctrine to better advance society's interests – are smaller than the costs. The court, then, refuses to refine the doctrine and lets all the new cases that come along to be ruled by precedent. On the contrary, if $C \leq L/2$ the court will never stop learning until it reaches perfect knowledge of the parameter θ.

Proposition 2. (1) *If $C > L/2$ the law converges (with probability one) without the court fully learning about θ; there exists a threshold value Δ_S such that the court chooses $\lambda_t = 1$, equivalently $a_t = b_t$ ($\neq \theta$ with probability one), when $\Delta_t \leq \Delta_S$.* (2) *If $C \leq L/2$ the court eventually fully learns; $\lambda_t < 1$ whenever $\Delta_t > 0$ and* $\lim_{t\to\infty} \Delta_t = 0$, $\lim_{t\to\infty} a_t = \lim_{t\to\infty} b_t = \theta$. Furthermore, $\lim_{t\to\infty} \lambda_t = 2C/L$.

Proof. We know from Proposition 1 and its proof that the optimal policy is $\alpha_t = \beta_t = \lambda_t/2 > 0$. Learning stops, and hence the law converges without full learning when $\alpha_t = \beta_t = 1/2$ or $\lambda_t = 1$. Replacing the policy values $\alpha_t = \beta_t = 1/2$ on the right hand side of (3), we see that the value function when it is optimal to stop learning is

$$V(\Delta_t) = -\frac{L}{4}\Delta_t - \frac{\ell}{24}\Delta_t^2 + \delta V(\Delta_t) = -\frac{\left(6L\Delta_t + \ell\Delta_t^2\right)}{24\,(1-\delta)}. \tag{10}$$

From (9) we know that the first order condition at an interior solution is $\Phi\left(\lambda_t, \Delta_t\right) = 0$. Note that $\lim_{\Delta_t \to 0} \Phi\left(\lambda_t, \Delta_t\right) = C - \frac{L}{2}\lambda_t$. Since it is $\lambda_t \leq 1$, if $C > L/2$ then it cannot be $\Phi\left(\lambda_t, \Delta_t\right) = 0$ for a "sufficiently small" value of Δ_t; that is, there cannot be an interior solution for Δ_t "small." Formally, there must exist $\overline{\Delta}$ such that $\Phi\left(\lambda_t = 1, \Delta_t\right) > 0$ if $\Delta_t < \overline{\Delta}$; for $\Delta_t < \overline{\Delta}$ there is a corner solution $\lambda_t = 1$. To derive the threshold $\overline{\Delta}$ such that $\Phi\left(\lambda_t = 1, \overline{\Delta}\right) = 0$, first replace $\lambda_t = 1$ in (9):

$$C - \frac{L}{2} - \frac{\ell}{8}\overline{\Delta} + \delta V(\overline{\Delta}) - \delta V\left(\frac{\overline{\Delta}}{2}\right) = 0; \tag{11}$$

then insert the value function from (10) to obtain

$$\begin{aligned} 0 &= C - \frac{L}{2} - \frac{\ell}{8}\overline{\Delta} - \frac{\delta\left(6L\overline{\Delta} + \ell\overline{\Delta}^2\right)}{24(1-\delta)} + \frac{\delta\left(3L\overline{\Delta} + \ell\frac{\overline{\Delta}^2}{4}\right)}{24(1-\delta)} \\ &= 3(1-\delta)\left(8C - 4L - \ell\overline{\Delta}\right) - \delta\left(3L\overline{\Delta} + \frac{3}{4}\ell\overline{\Delta}^2\right) \\ &= 12(1-\delta)(2C - L) - 3(1-\delta)\ell\overline{\Delta} - \delta\left(3L\overline{\Delta} + \frac{3}{4}\ell\overline{\Delta}^2\right) \\ &= -\frac{1}{4}\delta\ell\overline{\Delta}^2 - (\delta L + (1-\delta)\ell)\overline{\Delta} + 4(1-\delta)(2C - L) \end{aligned}$$

from which we have a unique positive solution (as long as $2C > L$)

$$\overline{\Delta} = \frac{-2(\delta L + (1-\delta)\ell) + 2\sqrt{(\delta L + (1-\delta)\ell)^2 + 4\delta\ell(1-\delta)(2C - L)}}{\delta\ell}. \tag{12}$$

Define $\Delta_S = \min\left\{\overline{\Delta}, 1\right\}$ as the threshold interval of uncertainty. If $2C > L$ and $\overline{\Delta} \leq 1$, then (12) defines the threshold interval of uncertainty. If $\overline{\Delta} > 1$, then the threshold is 1, the maximum size of the interval of uncertainty. If $\Delta_t \leq \Delta_S$ then the court sets $\lambda_t = 1$ and no learning takes places.

Now consider the case $2C \leq L$ and suppose, to the contrary, that there is a value $\Delta_S > 0$ at which learning stops. By (11) it must be

$$\begin{aligned} V\left(\frac{\Delta_S}{2}\right) &= V(\Delta_S) + \frac{1}{\delta}\left(C - \frac{L}{2} - \frac{\ell}{8}\Delta_S\right) \\ &< V(\Delta_S) \end{aligned}$$

which contradicts the fact, established in Lemma 1, that V is decreasing. Hence, if $2C \leq L$ it is optimal to set $\lambda_t < 1$ in all periods t. With probability $(1-\lambda_t)\Delta_t$ the court learns in period t and in such a case $\left(1 - \frac{\lambda_t}{2}\right)\Delta_t$ is an upper bound on the new interval of uncertainty Δ_{t+1}. It follows that $\lim_{t\to\infty}\Delta_t = 0$ and hence $\lim_{t\to\infty} a_t = \lim_{t\to\infty} b_t = \theta$. Learning never stops and in the limit the court discovers the true value of θ. It is immediate from (9) that, since Δ_t (and hence $V(\Delta_t)$) converges to zero as t goes to infinity, it must be $\lim_{t\to\infty}\lambda_t = 2C/L$. ■

If decision costs are greater than half the fixed cost of errors, the doctrine stabilizes with imperfections remaining in the law.[22] The court sets $a_t = b_t$ without knowing the exact location of θ. The court realizes the doctrine might not apply well in some circumstances, but correcting those imperfections is not cost-justified. As expected, if the fixed component of the error loss is zero, imperfect convergence always happens.

Loosely speaking, with imperfect convergence, the first few investigated cases and decisions will be important to the long run determination of which activities are permissible and which ones are not. Because of high decision costs, the courts will make inferences from the first cases investigated to cover lots of future cases (deciding those summarily). The first cases will anchor the discussion and play a disproportional role in the path of the law. With efficient convergence, by contrast, such anchoring vanishes. Although we do not explicitly consider the possibility of litigants selecting specific cases for trial, the insights from Proposition 2 suggest that any such selection effect will be more important when decision costs are large relative to the fixed error cost. The reason is that litigants will race to get a case on the books that anchors the law in their favor.

As noted in the introduction, some economists speculate that the efficiency of the common law provides a theoretical justification for the main finding of the law and finance literature – that a common law origin positively correlates with economic development. Our imperfect convergence result suggests that it can be inefficient for the law to be perfect across all cases and areas of law. That is to say, with scarce judicial resources, we should expect judges to promulgate and then stick with imperfect doctrines.[23] This is true even if the judges care solely about efficiency as the relevant benchmark for legal rules.

From Proposition 2, it is also immediate that the court will not learn at all, it will set $\lambda_1 = 1$, or $\alpha_1 = \beta_1 = 1/2$, as long as $\Delta_S = 1$. When the cost of examining a case relative to the fixed and variable loss from error is above a threshold, the court never investigates at all – no precedent stock is created.[24] The court summarily declares all cases below the error minimizing point 1/2 as permissible and all cases above this point as impermissible. There is no uncertainty in the application of law and no reason for litigants to bring cases. This suggests

[22] Notably, there is empirical evidence inconsistent with the convergence of judge-made law. Niblett et al. (2010) study the evolution of the economic loss rule (the rule limiting tort damage unless the loss results in personal injury or property damage). They find that "the doctrine has evolved in a way that cannot be easily described as convergence to efficiency" (Niblett et al., 2010, 354). They also find that the law has become less predictable in the last decade of their sample, as many state courts respond with their own exceptions to the rule. *Id.* While recent changes in state court preferences is an explanation consistent with our model, it is quite possible that factors that we left outside our model play an important part in explaining these findings.

[23] Note that, even when the law converges to efficiency, judge made law will still have errors along the way, simply because the precedent ratio is always positive.

[24] After simple calculation, (12) shows that $\overline{\Delta} \geq 1$, and hence $\Delta_S = 1$, is equivalent to $C \geq \frac{L}{2} + \frac{\delta}{1-\delta}\frac{\ell}{32} + \frac{1}{8}\left(\frac{\delta}{1-\delta}L + \ell\right)$, which needs not hold even if $L = 0$.

that, when decision costs are high relative to error costs, direct government regulation, via, say, a detailed statute, may be preferable to regulating by litigation and the courts. We might think of the regulator spending a fixed amount of resources locating θ ex ante. The natural question is when fixed expenditures are better than case by case expenditures. Intuitively, fixed expenditures won't be effective if the consequences of the regulation are unforeseeable and can only be discovered following, say, the materialization of an injury.

The extent of the expected inaccuracy in the converged doctrine is captured by Δ_S; the interval of uncertainty beyond which the law stops being refined.[25] The next proposition specifies the relationship between the scope of the "inaccuracy" in doctrine and the parameters of the model. The proof, which is relegated to the appendix, follows from the definition of Δ_S after straightforward computations.

Proposition 3. *Assuming imperfect convergence of doctrine, $C > L/2$, the expected inaccuracy of the ultimate legal rule, Δ_S, increases in the decision cost C and decreases in the error losses L and ℓ and the discount factor δ.*

The intuition is straightforward: an increase in the cost of examining a case, a decrease in the loss of an error and a reduction in the value attached to the future all have the effect of making learning less valuable and hence lead to greater inaccuracy in the law.

Like Proposition 1, Proposition 3 has testable implications. First, as decision costs go up, judge-made law should be more likely to involve per se rules – generalized rules that apply without consideration for the specific circumstances. One can think of per se rules as a commitment by the court to not investigate the merits of cases. It is well-known that per se rules allow for cheaper resolution of cases, but create room for errors (Easterbrook, 1984, 14-15). That trade-off is the one formally captured in our imperfect convergence result. Second, in times of weak judicial budgets, judicial mistakes should be more common. Assuming Congressional reversal of judge-made decisions involving, say, interpretation of a statute correlates with those decisions being mistaken, we should observe Congressional reversal increasing. Of course, Congress makes funding decisions for the judiciary and also decides when to step in and reverse a statutory interpretation. To do the test, one would need an exogenous shock to decision costs of the courts, induced by an institutional actor other than Congress, as in the Huang (2011) study discussed before.

Taken together, Propositions 2 and 3 suggests two ways to minimize mistakes in the creation and application of law. One way, perhaps, is to have Congress define the statute ex ante, but this will be costly, especially if the needed information is unavailable. A second way – which

[25] Because activities x are randomly drawn, the actual size of the uncertainty interval at which learning stops is a random variable. The expected size is an increasing function of Δ_S.

turns directly on the breath of judicial decisions – is to allocate the courts more resources. With more resources the court will read prior decisions more narrowly and, as a result, make fewer mistakes, both along the doctrinal path and once the doctrine converges.

4 The Likeness Principle

Having derived the endogenous use of precedent and studied the convergence of doctrine, this section asks a more fundamental question about the evolution of law. Will evolution be consistent with the rule-of-law value that like cases be treated alike or will like cases be treated differently?[26]

A doctrinal example demonstrates how judges might change the construction of precedent over time. In *New York v. Belton*,[27] a police officer pulled over a car with four men inside. Smelling drugs, the officer asked the men to get out of the car, and placed them in four separate areas on the highway. The officer proceeded to search the passenger compartment, where he found drugs in a jacket one of the men had left behind. In upholding the search, the Supreme Court held that when "a policeman has made a lawful custodial arrest of the occupant of an automobile, he may as a contemporaneous incident of that arrest, search the passenger compartment of that automobile." *Id.* at 460. In *Thorton v. United States*,[28] the defendant exited his car before interacting with the police officer. The police officer found drugs on the defendant's body, handcuffed him and placed him in the back of the patrol car. He then searched the defendant's car, where he found a firearm. The *Thorton* court upheld the search of the vehicle. In *Arizona v. Gant*,[29] the defendant got out of his car. While walking from his car, the officer arrested the defendant for driving with a suspended license, handcuffed him and placed him in the back of the patrol car. A search of the defendant's vehicle revealed a bag of cocaine and a gun. The Court found the search impermissible under *Belton*. To sum up, in cases with similar, perhaps even identical facts (*Thorton* and *Gant*), the court read the prior case law – *Belton* – as requiring different outcomes.

Inconsistent judicial decision making is well-established in the legal and political science literatures. Many empirical studies find that judges reach outcomes at significantly different rates depending on some measure of their political beliefs (Revesz, 1997; Cross and Tiller, 1998; Sunstein et al., 2006; Cox and Miles, 2008; Boyd, Epstein, and Martin, 2010). In a

[26] On the topic, the legal and philosphical literature is vast, see Fuller (1958), Hart (1958), Tamanaha (2004), McCubbins et al. (2010). Many influential scholars have stressed that like cases should be treated alike. See, for example, Rawls (1971, 208-209), Dworkin (1977, 113), Whittington (1999, 169).

[27] 453 U.S. 454 (1981).

[28] 541 U.S. 615 (2004).

[29] 556 U.S. 332 (2009).

study of unconscionability cases, Niblett (2009, 4) finds "where the facts of cases fall within the ambit of precedent, the outcomes of the case and the precedent are inconsistent in about 23% of the case-precedent pairs."

Our model shows that different policy preferences are not necessary to generate inconsistent judicial decision making. In addition, the model predicts lots of consistency too, as shown below. The end result is a mix of consistent and inconsistent judicial-making, depending in large part on whether the law at issue is in its infancy or not.[30]

4.1 Deriving The Likeness Principle

The ability to characterize precedent one way at one time and another way at another time provides the courts flexibility, which will allow them to incorporate new information into the law. As the next proposition shows, if $\lambda_t < 1$ and learning occurs at time t, then with positive probability next period the interpretation of precedent will vary in an inconsistent way. To see how this will occur, suppose the case x_{t-1} at time $t-1$ is deemed impermissible by appealing to precedent. Now suppose the case x_t arising at time t is investigated by the court and the activity is found to be socially valuable. As a result, the court learns and the interval of uncertainty changes. Suppose the new investigation bounds are such that $x_{t-1} < a_{t+1}$. This is possible and, as the next proposition will show, has positive probability of occurring. Finally, suppose the case x_{t+1} at time $t+1$ satisfies $x_{t-1} \leq x_{t+1} < a_{t+1}$. The court will judge it summarily and declare it permissible. There is inconsistency between the decision at $t-1$ and the decision at $t+1$. Since $x_{t-1} \leq x_{t+1}$, activity x_{t-1} is at least as socially valuable as activity x_{t+1}, but x_{t-1} is deemed not permissible at $t-1$ while x_{t+1} is deemed permissible at $t+1$. The inconsistency is due to the fact that at time t the court has learned and revised upwards its estimate of where the efficiency threshold θ lies.

Intuitively, inconsistency is more likely following an unexpected result of an investigation. If x_t is only marginally smaller than b_t, the court expects the likely outcome of the investigation to be that the activity has negative social value. Upon discovering that this is not so, that x_t is indeed valuable, the court learns a lot and in later periods may declare permissible activities that in the past would have been declared impermissible by summary disposition. In addition, the magnitude of the inconsistency depends on how much the court learns from the unexpected decision: the more the court learns, the greater the potential magnitude. Since the

[30] Recently, Fishman (2011) has estimated bounds on the range of inconsistency by studying asylum adjudication in the New York immigration courts. He finds that "a randomly selected pair of judges would disagree about the disposition of a randomly selected case at least one-quarter of the time, and perhaps as often as one-half the time." *Id.* at 30.

first few cases present a large opportunity for learning, we should expect greater inconsistency early in the development of judge made law. Notably, the amount of inconsistency is not necessarily small and transpiring just around the threshold, where it doesn't much matter which way the court decides the case. Instead, inconsistency turns on drawing a case close to the previous precedent boundary and having investigation reveal an outcome contrary to the court's intuition.

Inconsistent interpretation of prior cases that occurs with positive probability implies a violation of the likeness principle. Formally, let $\Pr(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1})$ be the probability of the following event occurring. The case x_{t-1} in period $t-1$ is at least as socially valuable as the case x_{t+1} in period $t+1$; both cases are judged according to precedent; x_{t-1} is deemed not permissible; x_{t+1} is deemed permissible. $\Pr(a_{t-1} > x_{t-1} \geq x_{t+1} > b_{t+1})$ is similarly defined. We say that the evolution of jurisprudence follows the *likeness principle* in period t if

$$\Pr(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1}) + \Pr(a_{t-1} > x_{t-1} \geq x_{t+1} > b_{t+1}) = 0.$$

The next proposition shows that the evolution of jurisprudence from a court that cares about errors and has limited judicial resources is not always consistent with the lofty rule-of-law value that identical cases be treated alike.

Proposition 4. *If $\lambda_t < 1$, then with positive probability the interpretation of prior case law at time $t+1$ is inconsistent with the interpretation at time t; that is, the likeness principle is violated at time t.*

Proof. First we show that $a_{t+1} > b_t$ with positive probability. An analogous argument could be made to show that $b_{t+1} < a_t$ with positive probability. Let $\gamma^* = 1 - \lambda_t = (b_t - a_t)/\Delta_t$. Since $\lambda_t < 1$, it is $\gamma^* > 0$ and $a_t < x(\gamma) = b_t - \gamma\Delta_t \leq b_t$ for all γ in the interval $I^* = [0, \gamma^*)$. Since $\theta > x(\gamma)$ with positive probability, with positive probability $W_{t+1} = x(\gamma)$ for some $\gamma \in I^*$ and

$$\begin{aligned} a_{t+1} &= W_{t+1} + \frac{\lambda_{t+1}}{2}[R_t - W_{t+1}] = x(\gamma) + \frac{\lambda_{t+1}}{2}[R_t - x(\gamma)] \\ &= b_t - \gamma\Delta_t + \frac{\lambda_{t+1}}{2}[R_t - b_t + \gamma\Delta_t] \\ &= b_t - \gamma\Delta_t + \frac{\lambda_{t+1}}{2}\left[\frac{\lambda_t}{2}\Delta_t + \gamma\Delta_t\right]. \end{aligned} \tag{13}$$

It follows that $a_{t+1} > b_t$ as long as

$$\gamma < \frac{\lambda_t\lambda_{t+1}}{2(2-\lambda_{t+1})} = \gamma^{**}.$$

Since, by Proposition 1, λ_t and λ_{t+1} are bounded away from zero (i.e., $\lambda_t \geq \overline{\lambda}(\Delta_t) > 0$ and $\lambda_{t+1} \geq \overline{\lambda}(R_t - x(\gamma)) > 0$), it is a positive probability event that in period t the selected case is $x(\gamma)$ with $\gamma \in [0, \min\{\gamma^*, \gamma^{**}\})$ and $\theta > x(\gamma)$. In such an event $a_{t+1} > b_t$. This shows that $\Pr(a_{t+1} > b_t) > 0$. Now observe that with positive probability $x_{t-1} \in (b_{t-1}, R_{t-1})$ and no learning takes place, so that $W_t = W_{t-1}$, $a_t = a_{t-1}$, $b_t = b_{t-1}$, $R_t = R_{t-1}$ and $\Delta_t = \Delta_{t-1}$. It follows that with positive probability $a_{t+1} > b_{t-1}$.

We conclude the proof by showing that $\Pr(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1}) > 0$. First note that with positive probability at time $t-1$ an activity x_{t-1} is selected in the interval $(b_{t-1}, b_{t-1} + \mu\Delta_t)$, where $\mu > 0$ is a constant to be defined later. The activity x_{t-1} is judged as being not permissible and the court does not learn. Since $\lambda_t < 1$, then with positive probability at t the selected activity is some $x(\gamma) = b_{t-1} - \gamma\Delta_t$, with $\gamma > 0$, which is investigated and viewed as efficient. Then, as shown in (13), the court will set $a_{t+1} = b_{t-1} - \gamma\Delta_t + \frac{\lambda_{t+1}}{2}\left[\frac{\lambda_t}{2}\Delta_t + \gamma\Delta_t\right]$ and with positive probability $a_{t+1} > b_{t-1}$. Since $x_{t-1} < b_{t-1} + \mu\Delta_t$, a sufficient condition for the event $x_{t-1} \leq x_{t+1} \leq a_{t+1}$ is

$$b_{t-1} + \mu\Delta_t < b_{t-1} - \gamma\Delta_t + \frac{\lambda_{t+1}}{2}\left[\frac{\lambda_t}{2}\Delta_t + \gamma\Delta_t\right],$$

or, equivalently,

$$\mu + \gamma < \frac{\lambda_{t+1}}{2}\left[\frac{\lambda_t}{2} + \gamma\right]. \tag{14}$$

Since λ_t and λ_{t+1} are bounded away from zero, it is possible to choose a set of values $\gamma > 0$ and $\mu > 0$ having positive probability measure and such that (14) holds. This concludes the proof that the likeness principle is violated at t when $\lambda_t < 1$. ■

Development economists and legal scholars recommend the adoption of rule-of-law values across countries. Surprisingly, and contrary to the conventional wisdom, Proposition 4 suggests that treating like cases alike, one of the central ingredients of the rule of law, is not always socially optimal in the short run and should not always be expected from a rational, benevolent court system. The benefits of non-discrimination must be traded off against learning. The court could avoid treating like cases differently by consistently construing precedent in the narrowest way. But, as explained in Proposition 1, this interpretative approach taxes judicial resources, without enough of an offsetting benefit from what can be learned from the case.

Our model does not take into account some costs of inconsistent decision making. Predictability allows parties to engage in long-run plans, without fear that an activity deemed permissible today will be found impermissible tomorrow. The ability to plan, then, facilitates "law-specific" investment by providing for settled expectations about the legal system. For example, inconsistent changes in the relevant tort standard may induce a firm not to invest on the production line of a newly designed product. We do not therefore question the general value of a reliable legal system. What our model shows is that full predictability comes with a price tag – costly judicial investigation of cases close to the prior case law.

Having established that inconsistencies can occur in the short run, the next question is whether they will persist throughout time. Will the law oscillate with, say, activities deemed permissible and then impermissible and then permissible again as time goes by?[31] First, it is obvious that if learning eventually stops, so too will inconsistency. Inconsistency requires adjustments of the beliefs about the location of θ. If the court stops investigating, it will never update its beliefs.

What if learning never stops? Recall that to generate inconsistent decisions at time t three events must occur. First, the case at time $t-1$ must be in the portion of the interval of uncertainty Δ_{t-1} where the judge makes a summary decision. Second, the case at time t must be in the portion of the interval of uncertainty $\Delta_t = \Delta_{t-1}$ where the judge investigates the case. Third, the case in period $t+1$ must be summarily decided inconsistently with the case at time $t-1$. As time goes by and the court learns, the interval of uncertainty converges to zero. It follows that the probability of inconsistent decisions vanishes and hence the likeness principle holds in the limit. Formally, we say that the *limit likeness principle* holds if:

$$\lim_{t\to\infty} \left[\Pr\left(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1}\right) + \Pr\left(a_{t-1} > x_{t-1} \geq x_{t+1} > b_{t+1}\right)\right] = 0$$

As we have informally argued, the following proposition holds.

Proposition 5. *The optimal evolution of doctrine satisfies the limit likeness principle.*

Proof. If $C > L/2$, then by Proposition 2 $\lim_{t\to\infty} \lambda_t = 1$ and hence $\lim_{t\to\infty} a_t = \lim_{t\to} b_t$. If $C \leq L/2$, then by Proposition 2 $\lim_{t\to\infty} a_t = \lim_{t\to} b_t = \theta$. Hence, in both cases

$$\lim_{t\to\infty} \Pr\left(b_{t-1} < x_{t-1} \leq x_{t+1} < a_{t+1}\right) = \lim_{t\to\infty} \Pr\left(a_{t-1} > x_{t-1} \geq x_{t+1} > b_{t+1}\right) = 0.$$

■

With the use of a simple continuity argument, observe that if the limit likeness principle holds then, given any finite number of periods n and any arbitrarily small probability mass ε,

[31] In our model, standards, in effect, harden into rules. We do not speak to oscillation, where standards turn into rules and back into standards. Such movement might be incorporated in the model by assuming that the location of θ changes with some probability each period. On this type of oscillation more generally and the convergence of rules and standards, see Baker and Kim (2011), Schauer (2003), Johnston (1995), and Rose (1989).

the probability that the interpretation of precedent at time t is inconsistent with the interpretation of precedent in any of the previous n periods becomes less than ε as t grows large. The probability of an inconsistent interpretation at t with any of the previous n periods goes to zero in the limit.

4.2 Testable Implications of Inconsistent Judicial Decision Making

Combined, Proposition 4 and 5 lead to several testable implications. First, as noted, one would expect to observe more inconsistency when judge-made law is in its infancy – shortly, say, after Congress passes a statute containing a broad delegation of authority to the courts. This prediction can be run in a horse race against the preference-based account of judicial decision making, the dominant account. If inconsistency is solely the result of different judicial preferences, the length of time between the passage of the enabling statute and the amount of inconsistent decision making should not be significant.[32]

Second, under our model the same judge can be inconsistent over time. To the extent that one finds judges switching their own positions in the application of the legal standard, this is evidence in favor of our model and against a preference-based model of judicial decision making (unless, of course, one claims that a judge's preferences change over time, a claim which would make the preference-based model untestable).

Third, our model suggests that inconsistent dispositions should be more likely following a surprise decision. To test this, one might take an opinion that caught the legal commentators off guard – an unexpected resolution of a legal issue. Then, one might find a set of cases previously decided summarily, by, say, unpublished decision or the denial of certiorari (a form of summary disposition), and ask whether courts in subsequent cases with roughly similar facts granted oral argument or resolved the case with a written opinion.[33]

Fourth, our model predicts a higher probability of inconsistency between two unpublished decisions or a published decision and an unpublished decision than between two published decisions. Again, if judicial preferences were the sole driver of inconsistency, the form of resolution should not make a difference to the amount of inconsistency.

Finally, the inconsistency result can be linked back to the motivating example involving the

[32] Niblett (2009) provides some suggestive evidence of the sort we envision. In the unconscionability cases, he finds that "the level of inconsistency in the system, on average, continues to increase until about the 100th case. After this, the level of inconsisteny falls – quite dramatically." *Id.* at 7. Niblett finds also that conflicting judicial politics is a good predictor of inconsistency. *Id.* at 30. As explained above, unlike our model and as a matter of theory, judicial preferences alone cannot explain why the amount of inconsistency varies over time.

[33] A certiorari is a formal request to a court (in this example the Supreme Court) challenging a legal decision of an administrative tribunal, judicial office or organization.

constitutional limit on punitive damages. In 1991 in *Pacific Mut. Life Ins. Co. v. Haslip*,[34] the Court found constitutional an award with a ratio of four to one between punitive damages and compensatory damages. In 1993, the Court in *TXO Prod. Corp. v. Alliance Resource Corp*[35] found that an award of 10 to 1 passed constitutional muster. Suppose that, during the period 1991 and 1993, the court failed to grant certiorari and allowed a circuit court decision to stand striking down an award of 6 to 1. That, in our model, would be evidence of inconsistency. Likewise, in 1996 in *B.M.W., Inc. v. Gore*,[36] the Court found an award of 500 to 1 grossly excessive. In 2003, the court found an award of 145 to 1 grossly excessive in *State Farm Mutual Automobile Ins. Co. v. Campbell*.[37] To test for inconsistency, one might search for cases between 1996 and 2003, in which the Court summarily resolved a case – i.e., denied certiorari – where a circuit court had upheld an award with a multiple between 145 and 499.

5 Conclusion

We believe the results derived here are applicable to any situation where decision makers learn from experience and make yes or no decisions on a case by case basis. Take a front-line employee deciding how to deal with grievances by his subordinates. He must decide which grievances should be sent up the chain of command and which should not. Suppose he refers the first grievance up the chain. That decision establishes a precedent. The next grievance he encounters requires reflection on how close that grievance is to the previous one. If the two are close, the front-line employee saves resources by simply following precedent, rather than investigating the pros and cons of sending that specific grievance to his superior. But there is risk of mistake. Perhaps the second grievance is one that he should handle. The same results follow in this situation as in the model – the endogenous following of precedent, inconsistent decisions, the failure to always treat like cases alike and the making of rules that sticks, despite the decision maker realizing the rules work improperly in some circumstances.

Turning back to judge-made law, a few remarks are in order. In contrast to what is done here, the non-formal legal literature on this topic often speak of judges "writing" broadly or narrowly (Sunstein, 1999, 10). In reality, of course, both the writing judge and reading judge play a role in the creation of judge-made law. The "writing" judge might use expansive language in the opinion to signal something to future judges or lower courts about what he learned from his investigation and about how the precedent could be applied in future cases. Here, such

[34] 499 U.S. 1 (1991).
[35] 509 U.S. 443 (1993).
[36] 517 U.S. 559 (1996).
[37] 538 U.S. 408 (2003).

signaling is unnecessary because the prior judges never make mistakes upon investigation and opinions perfectly convey the investigation results.

The model might be extended to include judges with different preferences and opinions containing mistakes in transmission. The interesting insights from this extension would come from the distortions to the interpretation of precedent that the current judge may introduce in order to counteract the unwanted effects on the law of a future judge with different policy preferences. What's surprising is that our model has descriptive power, while maintaining the strong assumption that judges share the same normative commitments. For instance, optimal deviations from the likeness principle arise even when nothing about judicial preferences or the underlying environment has changed.

In addition, the model is framed in terms of a single infinitely lived judge. Instead, one might interpret the model in terms of a judicial hierarchy, where the upper court has high decision costs. To economize on these costs, the upper court hears some cases in the interval of uncertainty. It then delegates other cases close to its prior decisions to the lower court, which has low decision costs. The interpretative choice, then, becomes how many cases to delegate. This interpretation of the model makes some headway in explaining both the degree of delegation and the reason lower courts defer to higher court precedents (deference occurs because the higher court decisions result from investigation and, as a result, are correct). A delegation-style model of this sort bears resemblance to recent work by Ellison and Holden (2010).

One final worthwhile extension would be to endogenize the decision of which cases are brought before the court. For example, it could be that one party, say the party interested in the activity being declared permissible, is in a stronger position to bring cases to court. The court will then tend to see a biased sample of cases and it is natural to conjecture that the court will want to skew its reliance on precedent against the stronger party, in order to facilitate learning.

We conclude by stressing that our model reflects what judges claim to be doing: (1) looking at facts; (2) surveying prior precedent for guidance about what to do; and (3) trying to reach the best result. Notably, many of the features we observe in judge-made law flow as a natural consequences of judges doing what they say they are doing.

Appendix

In this appendix, first we prove existence and uniqueness of the value function of the court's maximization problem, then we prove Proposition 3 and finally we show that Proposition 1 and 2, proven in the main body for the uniform model, extends to the general model.

Lemma 1. *In the uniform model, the value function $V(\Delta)$ is continuously differentiable, negative valued, decreasing in* Δ, with $V(0)=0$ *and uniquely defined by:*

$$
\begin{aligned}
V(\Delta_t) = \max_{\alpha_t \geq 0,\ \beta_t \geq 0,\ 1-\alpha_t-\beta_t \geq 0} \Bigg\{ & -C\left(1-\alpha_t-\beta_t\right)\Delta_t - \frac{L}{2}\left(\alpha_t^2+\beta_t^2\right)\Delta_t \\
& -\frac{\ell}{6}\left(\alpha_t^3+\beta_t^3\right)\Delta_t^2 + \delta V(\Delta_t)\left[1-\left(1-\alpha_t-\beta_t\right)\Delta_t\right] \\
& +\delta \int_{\alpha_t}^{1-\alpha_t} V(\Delta_t x_t)\Delta_t x_t dx_t + \delta \int_{\beta_t}^{1-\beta_t} V(\Delta_t x_t)\Delta_t x_t dx_t \Bigg\}.
\end{aligned}
\tag{15}
$$

Proof. Recalling that $\alpha_t\Delta_t = a_t - W_t$, $\beta_t\Delta_t = R_t - b_t$, and using (1) and (2), we can write the court's objective function for the uniform version as:

$$
\begin{aligned}
V(\Delta_t) = \max_{\alpha_t \geq 0,\ \beta_t \geq 0,\ 1-\alpha_t-\beta_t \geq 0} \Bigg\{ & -C\left(1-\alpha_t-\beta_t\right)\Delta_t - \frac{L}{2}\left(\alpha_t^2+\beta_t^2\right)\Delta_t \\
& -\ell \int_{W_t}^{a_t} \frac{\left(x_t-W_t\right)^2}{2\Delta_t} dx_t - \ell \int_{b_t}^{R_t} \frac{\left(R_t-x_t\right)^2}{2\Delta_t} dx_t + \delta V(\Delta_t)\left[1-\left(1-\alpha_t-\beta_t\right)\Delta_t\right] \\
& +\delta \int_{W_t+\alpha_t\Delta_t}^{R_t-\beta_t\Delta_t} V(y_t-W_t)\frac{y_t-W_t}{\Delta_t} dy_t + \delta \int_{W_t+\alpha_t\Delta_t}^{R_t-\beta_t\Delta_t} V(R_t-y_t)\frac{R_t-y_t}{\Delta_t} dy_t \Bigg\}.
\end{aligned}
$$

Changing the variable of integration from y_t to $x_t = \left(y_t - W_t\right)/\Delta_t$ in the third integral and to $x_t = \left(R_t - y_t\right)/\Delta_t$ in the fourth integral yields

$$
\begin{aligned}
V(\Delta_t) = \max_{\alpha_t \geq 0,\ \beta_t \geq 0,\ 1-\alpha_t-\beta_t \geq 0} \Bigg\{ & -C\left(1-\alpha_t-\beta_t\right)\Delta_t - \frac{L}{2}\left(\alpha_t^2+\beta_t^2\right)\Delta_t \\
& -\frac{\ell}{6}\left(\alpha_t^3+\beta_t^3\right)\Delta_t^2 + \delta V(\Delta_t)\left[1-\left(1-\alpha_t-\beta_t\right)\Delta_t\right] \\
& +\delta \int_{\beta_t}^{1-\alpha_t} V(\Delta_t x_t)\Delta_t x_t dx_t + \delta \int_{\alpha_t}^{1-\beta_t} V(\Delta_t x_t)\Delta_t x_t dx_t \Bigg\}.
\end{aligned}
$$

Rearranging, we obtain (15).

We now show existence, uniqueness and differentiability of the value function. Recall that $\Delta \in [0,1]$ and note that $V(0) = 0$. Let S be the metric space of continuously differentiable, real valued functions $\omega : [0,1] \to \mathbb{R}$. Let the metric on S be $\rho\left(\omega^0, \omega^1\right) = \sup_{\Delta\in[0,1]} \left|\omega^0(\Delta) - \omega^1(\Delta)\right|$. Define the operator T, mapping the metric space S onto itself as follows:

$$T\omega(\Delta)=\max_{\alpha\geq 0,\ \beta\geq 0,\ 1-\alpha-\beta\geq 0}\frac{\Delta}{1-\delta\left(1-\Delta\right)}\left\{-C\left(1-\alpha-\beta\right)-\frac{L}{2}\left(\alpha^2+\beta^2\right)\right.$$
$$\left.-\frac{\ell}{6}\left(\alpha^3+\beta^3\right)\Delta+\delta\omega(\Delta)\left(\alpha+\beta\right)+\delta\int_{\alpha}^{1-\alpha}\omega(\Delta x)xdx+\delta\int_{\beta}^{1-\beta}\omega(\Delta x)xdx\right\}. \tag{16}$$

Interpreting ω as an initial value-function guess, the mapping T associates to ω an updated guess $T\omega$. The value function V defined in (15) is a fixed point of the mapping T. (The parameter $\frac{\Delta}{1-\delta(1-\Delta)}$ is obtained by moving the term $V\left(\Delta\right)\delta\left(1-\Delta\right)$ from the right to the left hand side of (15) and then dividing both sides by $1-\delta\left(1-\Delta\right)$.) We now show that T is a contraction mapping and hence V exists and is unique (and continuously differentiable). We apply Blackwell's Theorem (see Blackwell, 1965, or Stokey and Lucas, 1989, p.54). We need to show that T satisfies monotonicity and discounting. (1) Take $\omega^0\left(\Delta\right)\leq\omega^1\left(\Delta\right)$ for all $\Delta\in[0,1]$. It is immediate that $T\omega^0\left(\Delta\right)\leq T\omega^1\left(\Delta\right)$ and hence monotonicity holds. (2) To see that discounting also holds, let z be a non negative constant map defined by $z\left(\Delta\right)=z$ for all $\Delta\in[0,1]$. Let the map $\omega+z$ be defined by $\left(\omega+z\right)\left(\Delta\right)=\omega(\Delta)+z$. We need to show that $T\left(\omega+z\right)\left(\Delta\right)=T\omega(\Delta)+\gamma z$ for some $\gamma\in(0,1)$. It is

$$\begin{aligned}
T\left(\omega+z\right)\left(\Delta\right)=&\max_{\alpha\geq 0,\ \beta\geq 0,\ 1-\alpha-\beta\geq 0}\frac{\Delta}{1-\delta\left(1-\Delta\right)}\left\{-C\left(1-\alpha-\beta\right)-\frac{L}{2}\left(\alpha^2+\beta^2\right)\right.\\
&-\frac{\ell}{6}\left(\alpha^3+\beta^3\right)\Delta+\delta\left(\omega(\Delta)+z\right)\left(\alpha+\beta\right)\\
&\left.+\delta\int_{\alpha}^{1-\alpha}\left(\omega(\Delta x)+z\right)xdx+\delta\int_{\beta}^{1-\beta}\left(\omega(\Delta x)+z\right)xdx\right\}\\
=&\max_{\alpha\geq 0,\ \beta\geq 0,\ 1-\alpha-\beta\geq 0}\frac{\Delta}{1-\delta\left(1-\Delta\right)}\left\{-C\left(1-\alpha-\beta\right)-\frac{L}{2}\left(\alpha^2+\beta^2\right)\right.\\
&-\frac{\ell}{6}\left(\alpha^3+\beta^3\right)\Delta+\delta\omega(\Delta)\left(\alpha+\beta\right)+\delta\int_{\alpha}^{1-\alpha}\omega(\Delta x)xdx\\
&\left.+\delta\int_{\beta}^{1-\beta}\omega(\Delta x)xdx+\delta z\left(\alpha+\beta\right)+\frac{\delta z}{2}\left(\left(1-\alpha\right)^2-\alpha^2+\left(1-\beta\right)^2-\beta^2\right)\right\}\\
=&T\omega(\Delta)+\frac{\Delta\delta z}{1-\delta\left(1-\Delta\right)}\left(\alpha+\beta+\frac{\left(1+\alpha^2-2\alpha\right)-\alpha^2+\left(1+\beta^2-2\beta\right)-\beta^2}{2}\right)\\
=&T\omega(\Delta)+\frac{\delta\Delta}{1-\delta\left(1-\Delta\right)}z,
\end{aligned}$$

Since $\frac{\delta\Delta}{1-\delta(1-\Delta)}=\frac{1}{1+\frac{1-\delta}{\delta\Delta}}\in(0,1)$, this proves that discounting holds and hence that T is a contraction. Its unique fixed point is the continuously differentiable value function V.

To see that V is negative and decreasing, consider the set S' of continuously differentiable functions which are negative-valued and decreasing (i.e., such that $\omega(\Delta) \leq 0$ and $\omega'(\Delta) \leq 0$). The set S' is a closed subset of S. We need to show that T maps S' onto itself. Then we can conclude that the fixed point of T, the value function V, is negative-valued and decreasing (e.g., see Corollary 1 to the Contraction Mapping Theorem in Stokey and Lucas, 1989, 52.). First, it is immediate from (16) that if $\omega(\Delta) \leq 0$, then it is also $T\omega(\Delta) \leq 0$. Hence T maps negative functions into negative functions. It only remains to show that T maps decreasing function into decreasing functions. Suppose $\omega(\Delta)$ is decreasing and differentiate (16) at the solution values α_*, β_* to obtain:

$$
\begin{aligned}
\frac{\partial T\omega(\Delta)}{\partial\Delta} &= \frac{1-\delta}{[1-\delta(1-\Delta)]^2}\left\{\frac{1-\delta(1-\Delta)}{\Delta}T\omega(\Delta)\right\} \\
&+\frac{\Delta}{1-\delta(1-\Delta)}\left\{-\frac{\ell}{6}\left(\alpha_*^3+\beta_*^3\right)+\delta\omega'(\Delta)\left(\alpha_*+\beta_*\right)\right. \\
&\left.+\delta\int_{\alpha_*}^{1-\alpha_*}\omega'(\Delta x)x^2dx+\delta\int_{\beta_*}^{1-\beta_*}\omega'(\Delta x)x^2dx\right\} \\
&\leq 0.
\end{aligned}
$$

where the first term is curly brackets is equal to the term in curly brackets in (16) and the inequality follows from all terms being negative. ■

We now provide the proof of Proposition 3.

Proposition 3. *Assuming imperfect convergence of doctrine, $C > L/2$, the expected inaccuracy of the ultimate legal rule, Δ_S, increases in the decision cost C and decreases in the error losses L and ℓ and the discount factor δ.*

Proof. Recall that Δ_S is implicitly defined by:

$$\varphi(\cdot) := -\frac{1}{4}\delta\ell\Delta_S^2 - (\delta L + (1-\delta)\ell)\Delta_S + 4(1-\delta)(2C-L) = 0$$

Since $\partial\varphi/\partial\Delta_S < 0$, the sign of the impact on Δ_S of a change in exogenous variable $z \in \{C, L, \ell, \delta\}$ is given by $\partial\varphi/\partial z$.

It immediate that $\partial\varphi/\partial C > 0$, $\partial\varphi/\partial L < 0$ and $\partial\varphi/\partial\ell < 0$; hence Δ_S is increasing in C and decreasing in L and ℓ. Moreover:

$$
\begin{aligned}
\partial\varphi/\partial\delta &= -\frac{1}{4}\ell\Delta_S^2 - (L-\ell)\Delta_S - 4(2C-L) \\
&= \frac{1}{\delta}\left[\varphi + \ell\Delta_s - 4(2C-L)\right] \\
&= \frac{1}{\delta}\left[\ell\Delta_s - 4(2C-L)\right].
\end{aligned}
$$

Recall from (12) that

$$\ell\Delta_S = \frac{-2\left(\delta L + (1-\delta)\,\ell\right) + 2\sqrt{\left(\delta L + (1-\delta)\,\ell\right)^2 + 4\delta\ell\,(1-\delta)\,(2C-L)}}{\delta}$$

and hence the sign of $\partial\varphi/\partial\delta$ is the same as the sign of

$$-2\left(\delta L + (1-\delta)\,\ell\right) + 2\sqrt{\left(\delta L + (1-\delta)\,\ell\right)^2 + 4\delta\ell\,(1-\delta)\,(2C-L)} - 4\delta\,(2C-L)\,,$$

which is negative since

$$\sqrt{\left(\delta L + (1-\delta)\,\ell\right)^2 + 4\delta\ell\,(1-\delta)\,(2C-L)} < \left(\delta L + (1-\delta)\,\ell\right) + 2\delta\,(2C-L)\,.$$

It follows that Δ_S is decreasing in δ. ■

We now consider the general model and provide sketches of the proofs of Lemma 1 and Propositions 1 and 2; the proofs of Propositions 4 and 5 only require minor modifications (e.g., keeping track of the fact that it is not any longer true that $\alpha_t = \beta_t = \frac{\lambda_t}{2}$) and are omitted.

Lemma $1^{\mathbf{G}}$. *The value function $V\,(W_t, R_t)$ is continuously differentiable and negative valued,* with $V\,(0) = 0$.

Proof Sketch. Let S be the metric space of continuously differentiable, real valued functions $\omega : D \to \mathbb{R}$, where $D = \left\{(x,y) \in [0,1]^2 : x \leq y\right\}$. Define the operator T, mapping the metric space S onto itself as follows:

$$\begin{aligned}
T\omega(W_t,R_t) = \frac{1}{[F(R_t)-F(W_t)]} \max_{W_t \leq a_t \leq b_t \leq R_t} \Big\{ & \delta\omega(W_t,R_t)\,[1-(G(b_t)-G(a_t))]\,[F(R_t)-F(W_t)] \\
& + \delta \int_{a_t}^{b_t} \omega(W_t,x_t)\,[F(x_t)-F(W_t)]\,g(x_t)dx_t + \delta \int_{a_t}^{b_t} \omega(x_t,R_t)\,[F(R_t)-F(x_t)]\,g(x_t)dx_t \\
& - C\,[G(b_t)-G(a_t)]\,[F(R_t)-F(W_t)] \qquad (17) \\
& - \int_{W_t}^{a_t}\int_{W_t}^{x_t} Lf\,(\theta)\,d\theta g(x_t)dx_t - \int_{b_t}^{R_t}\int_{x_t}^{R_t} Lf(\theta)d\theta g(x_t)dx_t \\
& - \int_{W_t}^{a_t}\int_{W_t}^{x_t} \ell\,(x_t-\theta)\,f(\theta)d\theta g(x_t)dx_t - \int_{b_t}^{R_t}\int_{x_t}^{R_t} \ell\,(\theta-x_t)\,f(\theta)d\theta g(x_t)dx_t \Big\}.
\end{aligned}$$

The value function V defined in (1) is a fixed point of the mapping T. To apply Blackwell (1965) Theorem (i.e., show existence of V), we must show that T satisfies monotonicity and discounting. Monotonicity is immediate, since $\omega^0(W_t,R_t) \leq \omega^1(W_t,R_t)$ for all $(W_t,R_t) \in D$ implies $T\omega^0(W_t,R_t) \leq T\omega^1(W_t,R_t)$. To see that discounting also holds, let z be a non negative constant map defined by $z(W_t, R_t) = z$ for all $(W_t, R_t) \in D$. Let the map $\omega + z$ be defined by $(\omega + z)(W_t, R_t) = \omega(W_t, R_t) + z$. We must show that $T(\omega + z)(W_t, R_t) = T\omega(W_t, R_t) + \gamma z$ for some $\gamma \in (0, 1)$. The equality holds for $\gamma = \delta$, since

$$
\begin{aligned}
& T(\omega + z)(W_t, R_t) \\
&= \frac{1}{[F(R_t) - F(W_t)]} \max_{W_t \le a_t \le b_t \le R_t} \Big\{ \delta [\omega(W_t, R_t) + z] [1 - (G(b_t) - G(a_t))] [F(R_t) - F(W_t)] \\
&+ \delta \int_{a_t}^{b_t} \{ [\omega(W_t, x_t) + z] [F(x_t) - F(W_t)] + [\omega(x_t, R_t) + z] [F(R_t) - F(x_t)] \} g(x_t) dx_t \\
&- C [G(b_t) - G(a_t)] [F(R_t) - F(W_t)] - \int_{W_t}^{a_t} \int_{W_t}^{x_t} L f(\theta) d\theta g(x_t) dx_t - \int_{b_t}^{R_t} \int_{x_t}^{R_t} L f(\theta) d\theta g(x_t) dx_t \\
&- \int_{W_t}^{a_t} \int_{W_t}^{x_t} \ell (x_t - \theta) f(\theta) d\theta g(x_t) dx_t - \int_{b_t}^{R_t} \int_{x_t}^{R_t} \ell (\theta - x_t) f(\theta) d\theta g(x_t) dx_t \Big\} \\
&= T(\omega + z)(W_t, R_t) + \delta z.
\end{aligned}
$$

Now consider the set S' of continuously differentiable functions which are negative-valued. It is immediate to conclude that the value function V is negative-valued, since T maps S' onto itself: $\omega(W_t, R_t) \le 0$ implies $T\omega(W_t, R_t) \le 0$. ■

The following first order conditions are obtained by differentiating (1) with respect to a_t and b_t, with multiplier η_t associated to the constraint $a_t \le b_t$, and multipliers μ_t^a, μ_t^b associated to constraints $W_t \le a_t$ and $b_t \le R_t$:

$$
\begin{aligned}
0 \;=\;& C - L \frac{F(a_t) - F(W_t)}{F(R_t) - F(W_t)} - \ell \int_{W_t}^{a_t} \frac{(a_t - \theta) f(\theta)}{F(R_t) - F(W_t)} d\theta + \delta V(W_t, R_t) \\
& - \delta V(a_t, R_t) \frac{F(R_t) - F(a_t)}{F(R_t) - F(W_t)} - \delta V(W_t, a_t) \frac{F(a_t) - F(W_t)}{F(R_t) - F(W_t)} - \frac{(\eta_t - \mu_t^a)}{g(a_t)}
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
0 \;=\;& -C + L \frac{F(R_t) - F(b_t)}{F(R_t) - F(W_t)} + \ell \int_{b_t}^{R_t} \frac{(\theta - b_t) f(\theta)}{F(R_t) - F(W_t)} d\theta - \delta V(W_t, R_t) \\
& + \delta V(b_t, R_t) \frac{F(R_t) - F(b_t)}{F(R_t) - F(W_t)} + \delta V(W_t, b_t) \frac{F(b_t) - F(W_t)}{F(R_t) - F(W_t)} + \frac{(\eta_t - \mu_t^b)}{g(b_t)}.
\end{aligned} \tag{19}
$$

Proposition 1$^{\mathbf{G}}$. *In each period t the court chooses $a_t > W_t$ and $b_t < R_t$.*

Proof Sketch. Suppose, to the contrary, that $a_t = W_t < R_t$. Equation (18) becomes $C = \frac{(\eta_t - \mu_t^a)}{g(a_t)}$, which can only be satisfied if $b_t = a_t$. Then (19) becomes $C - L - \ell \int_{W_t}^{R_t} \frac{(\theta - W_t) f(\theta)}{F(R_t) - F(W_t)} d\theta = \frac{(\eta_t - \mu_t^b)}{g(b_t)}$, which can only be satisfied if $b_t = R_t$. This contradicts $b_t = a_t = W_t$. Hence it must be $a_t > W_t$ whenever $W_t < R_t$.

Similarly, suppose $b_t = R_t > W_t$. Then (19) becomes $C = \frac{(\eta_t - \mu_t^b)}{g(a_t)}$, which requires $a_t = b_t$, while (18) becomes $C - L - \ell \int_{W_t}^{R_t} \frac{(a_t-\theta)f(\theta)}{F(R_t)-F(W_t)} d\theta = \frac{(\eta_t - \mu_t^a)}{g(a_t)}$, which requires $a_t = W_t$, a contradiction. Hence it must be $b_t < R_t$ whenever $R_t > W_t$. ■

Proposition 2$^{\mathbf{G}}$. (1) *If* $C > L/2$ *the law converges (with probability one) without the court fully learning about* θ. (2) *If* $C \leq L/2$ *the court eventually fully learns,* $\lim_{t\to\infty} a_t = \lim_{t\to\infty} b_t = \theta$.

Proof Sketch. For full learning to take place in the limit (i.e., $a_t \to \theta$, $b_t \to \theta$) it must be $a_t \neq b_t$ whenever $W_t \neq R_t$. On the other hand, learning stops if $a_t = b_t$. By Proposition 1^G, it is $\mu_t^a = \mu_t^b = 0$, and hence $V(W_t, R_t) < V(a_t, R_t)$ and $V(W_t, R_t) < V(W_t, a_t)$ whenever $W_t \neq R_t$. Furthermore, the last five terms on the right hand side of (18) and (19) are bounded and converge to zero as $R_t - W_t$ converges to zero. Thus, (18) and (19) imply that there exists a positive constant M such that

$$\begin{aligned} M(R_t - W_t) &> C - L\frac{F(a_t) - F(W_t)}{F(R_t) - F(W_t)} > 0 \\ M(R_t - W_t) &> C - L\frac{F(R_t) - F(b_t)}{F(R_t) - F(W_t)} > 0, \end{aligned}$$

as long as $R_t \neq W_t$. Adding up, we obtain

$$2M(R_t - W_t) > 2C - L\left(1 - \frac{F(b_t) - F(a_t)}{F(R_t) - F(W_t)}\right) > 0. \tag{20}$$

Suppose $C > L/2$ and, contrary to the proposition, learning never stops ($a_t \neq b_t$ for all t). Then it must be $\lim_{t\to\infty} R_t - W_t = 0$, and (20) implies that $2C - L = \lim_{t\to\infty} L\frac{F(a_t)-F(b_t)}{F(R_t)-F(W_t)} \leq 0$, a contradiction. This shows that if $C > L/2$ the law converges without the court fully learning.

Now suppose $C \leq L/2$. Note that (20) cannot be satisfied when $a_t = b_t$, because it cannot be $2C - L > 0$. Hence in such a case learning never stops. ■

References

Baker, Scott, and Pauline Kim. 2011. "A Dynamic Model of Legal Doctrine." Manuscript. School of Law, Washington University in St. Louis.

Bailey, Martin, and Paul Rubin. 1994. "A Positive Theory of Legal Change." *International Review of Law and Economics* 14 (December): 467-77.

Blackwell, David. 1965. "Discounted Dynamic Programming." *Annals of Mathematical Statistics* 36 (February): 226-235.

Boyd, Christina L., Lee Epstein, and Andrew D. Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54 (April): 389–411.

Calfee, John E. and Richard Craswell. 1984. "Some Effects of Uncertainty on Compliance with Legal Standards." *Virginia Law Review*, 70 (June): 965-1003.

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

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

Cox, Adam B. and Thomas J. Miles. 2008. "Judging the Voting Rights Act." *Columbia Law Review* 108 (January): 1–54.

Cross, Frank B., and Emerson H. Tiller. 1998. "Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals." *Yale Law Journal* 107 (May): 2155–2176.

Dari-Mattiacci, Giuseppe, Bruno Deffainis, and Bruno Lovat. 2011 "The Dynamics of the Legal System." *Journal of Economic Behavior and Organization* 79 (March advance access): 95-107.

Dworkin, Ronald. 1977. *Taking Rights Seriously.* Cambridge: Harvard University Press.

Easterbrook, Frank H. 1984. "The Limits of Antitrust." *Texas Law Review* 63 (August): 1-40.

Ellison, Glen and Richard Holden. 2010. "A Theory of Rule Development." Manuscript. Department of Economics, MIT.

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 (Spring): 101-137.

Eskridge, William N. and John Ferejohn. 1992. "The Article I Section 7 Game." *Georgetown Law Journal* 80 (February): 523-564.

Fishman, Joshua. 2011. "Inconsistency, Indeterminacy, and Error in Adjudication." Manuscript. Law School, University of Virginia.

Fuller, Lon. L. 1958. "Positivism and Fidelity to Law – A Reply to Professor Hart." *Harvard Law Review* 71 (February): 630-672.

Gennaioli, Nicola, and Andrei Shleifer. 2007. "The Evolution of the Common Law," *Journal of Political Economy* 115 (February): 43-68.

Hadfield, Gillian K. 1992. "Bias in the Evolution of Legal Rules." *Georgetown Law Journal* 80 (February): 583-616.

Hadfield, Gillian K. 2011. "The Dynamic Quality of Law: The Role of Judicial Incentives and Legal Human Capital." *Journal of Economic, Behavior, and Organization* 79 (March advance access): 80-94.

Hart, H.L.A. 1958. "Positivism and Fidelity to the Separation of Law and Morals," *Harvard Law Review* 71 (February): 593-629.

Hathaway, Oona A. 2001. "Path Dependence in the Law: The Course and Pattern of Legal Change in a Common Law System." *Iowa Law Review* 86 (January): 601-65.

Huang, Bert I. 2011. "Lightened Scrutiny." *Harvard Law Review* 124 (March): 1109-1152.

Hylton, Keith. 2006. "Information, Litigation, and the Common Law Evolution." *American Law and Economics Review* 8 (Spring): 33-61.

Johnston, Jason Scott. 1991. "Uncertainty, Chaos and the Torts Process: An Economic Analysis of Legal Form." *Cornell Law Review* 76 (January): 341-400.

Klerman, Daniel. 2007. "Jurisdictional Competition and the Evolution of the Common Law," *University of Chicago Law Review* 74 (Fall): 1179-1226.

Kornhauser, Lewis A. 1992a. "Modeling Collegial Courts I: Path-Dependence." *International Review of Law and Economics* 12 (June):169-85.

Kornhauser, Lewis A. 1992b. "Modeling Collegial Courts II: Legal Doctrine." *Journal of Law, Economics & Organization* 8 (October): 441-70.

La Porta, Rafael, Florencio Lopez-de-Silanes, Cristian Pop-Eleches, and Andrei Schleifer. 2004. "Judicial Checks and Balances." *Journal of Political Economy* 112 (April): 445-470.

La Porta, Rafael, Florencio Lopez–de-Silanes, Andrei Shleifer, and Robert Vishny. 1998. "Law and Finance." *Journal of Political Economy* 106 (December): 1113-1155.

Lax, Jeffrey A. 2011. "The New Judicial Politics of Legal Doctrine." *Annual Review of Political Science* 14 (February Advance Access): 131-157.

Lax, Jeffrey A. 2012. "Political Constraints on Legal Doctrine: How Hierarchy Shapes the Law." Manuscript. Department of Political Science, Columbia University.

Levi, Edward H. 1949. *An Introduction to Legal Reasoning.* Chicago: University of Chicago Press.

McCubbins, Matthew D., Daniel B. Rodriguez, and Barry R. Weingast."The Rule of Law Unplugged." Manuscript. Law School, University of Texas.

McNollgast. 1995. "Politics and the Courts: A Positive Theory of Judicial Doctrine and the Rule of Law." *Southern California Law Review* 68 (September): 1631-1683.

Niblett, Anthony. 2009. "Tracking Inconsistent Judicial Behavior." Manuscript. Law School, University of Chicago.

Niblett, Anthony. 2010. "Case-by-Case Adjudication and the Path of Law." Manuscript. Law School, University of Chicago.

Niblett, Anthony, Richard A. Posner, and Andrei Shleifer. 2010. "The Evolution of a Legal Rule." *Journal of Legal Studies* 39 (June): 325-358.

O'Hara, Erin. 1993. "Social Constraint or Implicit Collusion? Toward a Game Theoretic Analysis of Stare Decisis." *Seton Hall Law Review* 24: 736-778.

Parisi, Francesco and Fon, Vincy. 2008. *The Economics of Lawmaking.* New York, NY: Oxford University Press.

Posner, Richard A. [1973] 2007. *Economic Analysis of Law.* 7th ed. Boston: Little Brown Press.

Posner, Richard A. 2006. "Reasoning by Analogy." *Cornell Law Review* 91 (March): 761-774.

Priest, George L. 1977. "The Common Law Process and the Selection of Efficient Rules." *Journal of Legal Studies* 6 (January): 65-82.

Radin, Max. 1925. "The Theory of Judicial Decision: Or How Judges Think." *American Bar Association Journal* 11 (June): 357-362.

Rawls, John. 1971. *A Theory of Justice.* Cambridge Mass: Harvard University Press, revised edition.

Revesz, Richard L. 1997. "Environmental Regulation, Ideology, and the D.C. Circuit." *Virginia Law Review* 83 (November): 1717–1772.

Rose, Carol M. 1988. "Crystals and Mud in Property Law." *Stanford Law Review* 40 (February): 577-610.

Rubin, Paul H. 1977. "Why is the Common Law Efficient?" *Journal of Legal Studies* 6 (January): 51-63.

Schauer, Frederick. 2003. "The Convergence of Rules and Standards." *New Zealand Law Review* 2003: 303-328.

Schwartzstein, Joshua, Andrei Schleifer. 2011. "An Activity-Generating Theory of Regulation." Manuscript. Department of Economics, Dartmouth College.

Shavell, Steven. 2004. *Foundations of Economic Analysis of Law.* Cambridge: Harvard University Press.

Stephenson, Matthew C. 2009. "Legal Realism for Economists." *Journal of Economic Perspectives* 23 (Spring): 191-211.

Songer, Donald R., Jeffrey A. Segal and Charles M. Cameron. 1994. "Testing a Principal Agent Model of Supreme Court-Circuit Court Interactions." *American Journal of Political Science* 38 (August): 673-696.

Spitzer, Matthew, and Eric Talley. 2011. "Left, Right, and Center: Strategic Information Acquisition and Diversity in Judicial Panels." *Journal of Law, Economics, and Organization* (September Advance Access): 1-43.

Stokey, Nancy, and Robert E. Lucas Jr. 1989. *Recursive Methods in Economic Dynamics.* Cambridge: Harvard University Press.

Stras, David R., and Shaun M. Pettigrew. 2010. "The Rising Caseload in the Fourth Circuit: A Statistical and Institutional Analysis." *Southern Carolina Law Review* 61 (Spring): 421-444.

Sunstein, Cass R. 1999. *One Case at a Time.* Cambridge: Harvard University Press.

Sunstein, Cass R. 2001. "Cost-Benefit Default Principles." *Michigan Law Review* 99 (June): 1651-1723.

Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki. 2006. *Are Judges Political? An Empirical Analysis of the Federal Judiciary.* Washington, DC: Brookings Institution Press.

Tamanaha, Brian Z. 2004. *On the Rule of Law: History, Politics, Theory.* Cambridge: Cambridge University Press.

Whittington, Keith E. 1999. *Constitutional Interpretation: Textual Meaning, Original Intent, and Judicial Review.* Lawrence: University of Kansas Press.

Zywicki, Todd. 2003. "The Rise and Fall of Efficiency in the Common Law: A Supply-Side Explanation." *Northwestern Law Review* 97 (Summer): 1551-1633.

Cases Cited

Arizona v. Gant, 556 U.S. 332 (2009).

B.M.W., Inc. v Gore, 517 U.S. 559 (1996).

New York v. Belton, 453 U.S. 454 (1981).

Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1 (1991).

State Farm Mutual Automobile Ins. Co. v. Campbell, 538 U.S. 408 (2003).

Thorton v. United States, 541 U.S. 615 (2004).

TXO Prod. Corp. v. Alliance Resource Corp., 509 U.S. 443 (1993).