A Theory of Rational Jurisprudence*

Scott Baker † and Claudio Mezzetti ‡

† School of Law, Washington University in St. Louis
‡ Department of Economics, University of Warwick

October 15, 2010

Abstract

We develop a dynamic model of judge-made law, assuming judicial resources are scarce and the proper legal rule is unknown, but can be learned. Judges mechanically apply precedent in some cases to save resources. Other cases are fully investigated to reduce the chance of error. With positive probability, a rational court fails to treat like cases alike. The law converges to a mixture of efficient and inefficient rules, with the degree of inefficiency correlated with the ratio of decision to error costs. The size of each jurisprudential step depends on costs and the amount of uncertainty about the law.

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

JEL Classification Numbers: K10, K40.

1 Introduction

In common law systems, judges develop law case-by-case. Judicial decisions set the rules for property, contract, and tort. As such, judge-made law shapes production decisions,

*Many people have provided helpful comments on this project. We particularly thank participants at workshops at Tilburg University, the University of Chicago, Northwestern University, Boston University, USC, the University of Amsterdam, Washington University, the 2010 American Law and Economics Association Meetings, and the 2009 Triangle Workshop on Law and Economics.

investment decisions and exchange decisions. Judge-made law matters outside of market contexts too: It sets the ground rules for the political process; it defines the state's reach into private lives. While central to the study of law, economists rarely study the method by which judge-made law is formed. What properties can we expect from creating law case-by-case? Will the judiciary treat like cases alike? Will the law get better over time? These questions form the basis of our inquiry.

Judges pay attention to precedent, even though there are no meaningful sanctions for deviations from prior case law. Why? Indeed, it is not obvious what it means to "follow" precedent. No two cases are the same. The judge's interpretation of the prior case law will determine whether the prior precedent controls or not. But this interpretation is a choice and, as such, should be governed by maximizing some objective function given constraints.

Judges claim argument is central to deciding cases. Legal arguments often involve a discussion about whether the current case is sufficiently close to the prior case so that the resolution of the past case should either control or provide persuasive authority. Economists have built models of persuasion (see, for example, Glazer and Rubinstein, 2004), but arguments based on precedent are different in kind, more akin to reasoning by analogy. To our knowledge, no one has formally examined situations where the strength of the argument turns on some notion of "closeness" between the case at hand and prior cases. This is surprising, because "arguing from precedent" is a common practice outside of courts too.

Using a dynamic programming model, this paper builds a theory of judge-made law from the ground up. We start with two assumptions: (1) There is uncertainty about the consequences of legal rules; (2) Judicial resources are scarce. The theory yields both surprising and intuitive results. First, in our model reliance on precedent arises endogenously. Judges follow precedent not because deviations are punished, but rather as a mechanism for conserving scarce judicial resources. A judge decides how "close" a case is to the prior case and then decides whether the prior case is sufficiently informative as to be controlling. This interpretative choice 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 instead of relying on the precedent as a proxy for what to do.

In identifying the tradeoff between these two costs, the model advances a theory of interpretation. This theory provides a justification for the practice of arguing from precedent; 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 then explore the consequences of this theory for the evolution of law. First, under this theory, a court with positive probability will fail to treat like cases alike; that is, cases with "identical" relevant facts will 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. In the model, strict adherence to the likeness principle inhibits judicial learning and the cost-justified updating of legal rules.

Second, we show that the judge-made law will, in general, converge. This convergence will be of two types. If decision costs are small relative to 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 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. This suggests reasons to doubt that the judge-made law is efficient or could be made correct across all possible activities. Simply stated, as a consequence of the conservation of scarce judicial resources, a mixture of efficient and inefficient, but nonetheless sticky, legal rules emerges. The degree of inefficiency in the ultimate legal rules also depends on the ratio of decision costs to error costs. The lower this ratio the closer the law will converge to an efficient rule.

Finally, the paper sheds economic light on a major debate in jurisprudence among legal scholars and legal philosophers: when should a court move slowly in creating law and when should it move quickly? In the model, the answer to this question depends on decision costs, error costs, and the uncertainty in the law. When decision costs are low and error costs are high, the court rationally looks at a large proportion of cases that come to its attention; it relies sparely on precedent. As a result, the court makes fewer mistakes (which cost a lot), while spending more on decision costs (which cost little). When the residual degree of uncertainty in the law is small, the jurisprudential steps become larger. Stated differently, as the court learns more and more, a rational judiciary will construe each prior case as, in effect, deciding relatively more and more future cases. This result can be viewed as partial support for minimalism. Holding error costs and decision costs constant, the court moves the law more slowly the greater the amount of uncertainty as to consequences.

We assume a single court that lives forever, consisting of judges with identical policy preferences. Despite these strong assumptions, our results track the development of much judge-made law. One advantage is that we avoid making a series of reduced form assumptions about the policy preferences of the judges. All the model needs is uncertainty and scarce judicial resources. From that, many of the institutional features observed in judge-made law arise: faithful, but not blind adherence to precedent; inconsistent rulings and interpretation of precedent; violations of the likeness principle; a smattering of efficient and inefficient legal rules; and heterogeneity in the construction of precedent, over time and across areas of law.

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 shows 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. That section also contains proofs of the optimality of inconsistent interpretation of precedent. Section 5 studies how the size of the steps courts will take vary with time and other factors. Section 6 offers some extensions and concluding remarks.

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[1], they create a model of judge-made law. Judges, in their model, cannot overrule prior cases, but rather they can distinguish them. In particular, the judge searches for a different dimension along which to consider the case. 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 and overcome

[1] See La Porta et al. (1998) and La Porta et al. (2004).

the cost of a biased 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, the predictions outlined in the introduction arise. In short, we get insights consistent with the institutional features of judge-made law assuming that all judges balance the same costs and benefits, rather than being motivated by different policy agendas.

Two other significant literatures relate to this work. The first is from law and economics scholars; the second from the political scientists. 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. In most models, judges are passive and selection of cases by the litigants drive the development of law (Rubin, 1977; Priest, 1977; Bailey and Rubin, 1994; Hyton, 2006). The first model, Rubin (1977), assumed that parties litigate cases involving inefficient rules and settle cases involving efficient rules. The reason for this choice is that the gains from overturning an inefficient rule are higher. Because of this litigation strategy, the courts see more inefficient than efficient rules and, even if courts act randomly, the law dove-tails toward efficiency. Both 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. The opposite occurs in our model. Case flow is random and judicial choices – which cases to spend effort examining – determine the evolution of the law.

There are a few important exceptions to this pattern in the law and economics literature. 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 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 decisions is the "precedent" in their model. Distinct from us, judges don't interpret prior case law in their model, one of the critical feature of judging.

The political scientists assume that judges, like legislators, make decisions to advance their preferred policy objective. The judge has some preference, the question is whether she will be able to implement that preference given the other actors in the system. The constraints analyzed differ and depend on the question being asked.[2] Learning doesn't occur and the informational value of precedent is placed to one side. Because the judge always knows what she wants, the issue of wanting to learn from the experiences of judges in prior cases is assumed away. In contrast, we show how reasoning by analogy and the shifting interpretation of precedent can be seen as a method of efficiently managing resources to learn about the proper structure of legal rules. 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 and the information from new cases with new facts. If the new facts indicate that, as stated, the legal doctrine no longer serves its function, the common-law judge can reformulate the rule or create an exception.[3]

To capture this process, suppose that the judge-made law is attempting to regulate a set of activities, $x \in [0, 1]$.[4] Activity x carries costs and benefits. It might, for example, be the degree of care and attention of a motorist. Suppose that there is a threshold $\theta \in [0, 1]$, below which an activity is socially valuable and above which it is not. If the motorist drives carefully and with full attention, he poses little risk to others and his activity should be permitted. If the motorist drives recklessly or is distracted by other activities (e.g., speaking on a cell phone), he poses a high accident risk and should be deterred. The benefits exceed costs if and only if $x < \theta$. The threshold point θ is initially unknown.[5] We model it as random variable distributed according to $F(\theta)$ with positive

[2] 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 constrained by other judges sitting on the panel, see Spitzer and Talley (2009). For a model where judges face constains 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).

[3] Levi (1948 p. 8-27) contains the classic discussion of common law reasoning.

[4] The choice of the interval $[0, 1]$ as the set of feasible activities is just a convenient normalization.

[5] For a constitutional law example of courts searching for an unknown θ, consider the series of cases examining what counts as a punitive damage award so excessive as to violate substantive due

density $f(\theta)$ on $[0, 1]$.

The infinitely-lived court would prefer to allow beneficial activities and deter harmful activities. To do so, it needs to learn about the parameter θ.[6] Cases provide learning opportunities. Each period, the judge decides how to allocate his effort. Effort decisions are simple: the judge can either fully investigate a case or summarily examine it. Full investigation means that the judge spends resources uncovering the relationship between the new facts in the case and broader social policies. The judge might hear oral argument, ask for additional briefing from the lawyers, read the scholarly literature on the topic, etc.

Full investigation costs C and is rewarded: The judge discovers whether activity x is, or is not, permissible.[7] In learning x's relationship to θ, the judge also learns something about θ, the ideal scope of the legal rule. The judge reports what he has learned in an opinion.

Summary examination of a case saves judicial resources. It costs zero, but might result in an incorrect decision; that is, the court might erroneously declare an impermissible activity permissible or a permissible activity impermissible. We denote the social loss from error as L.

Let W_t be the highest case that the court has fully investigated up to t and found permissible, while R_t is the lowest case the court has found not permissible. Prior case law, in other words, teaches that activities in the interval $[0, W_t]$ are socially valuable

process. See Pacific Mut. Life Ins. Co. v. Haslip, 499 U.S. 1 (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 (1993) (noting that an award of 10 to 1 would not "jar one's constitutional sensibilities"); B.M.W., Inc. v. Gore, 517 U.S. 559 (1996) (finding that an award of 500 to 1 was grossly excessive); State Farm Mutual Automobile Ins. Co. v. Campbell, 538 U.S. 408 (2003) (finding that an award of 145 to 1 was grossly excessive). For other examples of this kind of evolution of law, see Niblett (2010).

[6] Because all judges have the same preferences, we use the term "judge" and "court" interchangably. In addition, we ignore the relationship between courts of different levels, like the Supreme Court and the appellate courts, or the appellate courts and the district courts.

[7] It simplifies the analysis to assume that there is no noise or mistake in the discovery process, and that the judge correctly determines whether a case is permissible upon its full investigation. Allowing for mistakes by the court wouldn't affect the main results over the long term, so long as the average decision was informative. Previous judicial errors would cancel out, enabling the court to extract important information relevant to updating from the prior cases.

and hence permissible, while activities in $[R_t, 1]$ are not. 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 $[W_t, R_t]$.

The timing each period is as follows:

1. A new activity x is randomly selected from the interval $[0, 1]$ according to the distribution $G(x)$ with positive density $g(x)$.[8] The activity x is brought as a case to the attention of the court.[9]

2. The court consults the prior case law; that is, the judge looks at W_t and R_t.

3. The court decides how broadly or narrowly to construe the past precedent. This decision determines the interval of cases $[a_t, b_t]$ that the court fully investigates; cases outside this interval are summarily examined.

4. The case is decided. Cases in $[0, a_t]$ will be considered permissible; cases in $[b_t, 1]$ impermissible; the permissibility of cases in $[a_t, b_t]$ is observed upon full investigation. If the court fully investigated the case, the court reports the result of its investigation in an opinion.

5. If an opinion has been issued, beliefs about θ are updated.

The "precedent interpretation" choice – stage 3 – sets two bounds, a_t and b_t. These bounds partitions the interval $[W_t, R_t]$ into three areas. The first area is the interval $[a_t, b_t]$; in this interval, the judge fully investigates the case.[10] The second area 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

[8] In specifying that the court always draws facts from the same distribution, we abstract away from the law's impact on primary 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 on their primary activity, or face a small probability of getting caught and sued.

[9] We do not explicitly model the decision by litigants to bring cases to court. Since the court only summarily examines cases outside the area $[W_t, R_t]$ where the law is certain, we could equivalently assume that these cases are settled out of court.

[10] Startz and Yoon (2009) and Niblitt (2010) also have models where the law is partially unsettled. In the Startz and Yoon model, it is tougher to convince a court to rule in a litigant's favor the farther the case is from the precedent bound. In Niblett's model, the court is assumed to rule as narrowly as possible each period. In neither model, does the court learn or optimize its resource use over time.

activities are close enough to prior case law (W_t and R_t) as to be decided by application of precedent alone without spending any effort. We can think of the size of the intervals $[W_t, a_t]$ and $[b_t, R_t]$ as a measure of how expansively the judge reads the prior precedent, or, alternatively, of the risk of relying on precedent. The larger these intervals, the more broadly the judge will extrapolate from the past precedent to cases with different facts without taking a fresh look at them, and hence the higher the chance the prior cases will be off point.

Note that expending effort benefits future judges because investigation leads to an opinion that adds to the precedent stock. If the court does not fully investigate at time $t \geq 1$, then, nothing new has been learned. The precedent stock remains the same. As such, $[W_{t+1}, R_{t+1}] = [W_t, R_t]$ and $[a_{t+1}, b_{t+1}] = [a_t, b_t]$.

To further motivate the model, consider the following example. Suppose the precedent stock is a case involving preschool children injured while trespassing and playing on an owner's sand pit with a large water pool in the middle.[11] Say the fence around the sand pit was poorly maintained, and the judge determined that the owner was liable. In the language of the model, the court learns, say, that $\theta \leq 0.8$. The next case involves a group of teenagers injured on a sand pit after drinking large quantities of alcohol.[12] The judge might interpret the first case as covering all "sand pit" cases and, as such, find the owner liable by summary examination. In so doing, the court might set b_t far from $R_t = 0.8$, say at 0.1. Broadly construing the precedent saves on resources, but might be wrong. The ideal rule might hold teenagers responsible for their actions, while letting preschoolers off the hook. On the other hand, suppose that the second case involves injury to a six year old. The court might reason that six year olds are close in age to preschoolers and, as such, rely on the preschool precedent. The court is less likely to make a mistake by proceeding in this fashion. Relying on precedent thus saves decision costs without unduly increasing error costs. Each period, the judge makes this interpretative choice. He balances the two costs, while understanding that full investigation today benefits future judges because it leads to an opinion.

[11] See, for example, Barklett v. Heersche, 462 P.2d 763 (1969).

[12] To the extent cases involves multiple dimensions of facts, we assume that the facts can be collapsed and bundled in a single dimension, x. In the above example, x might refer, at a general level, to the ease at which a party can avoid the accident. In law and economic terms, the court will be searching for the "least cost" avoider of the accident. On the relationship between the least cost avoider inquiry and the general models of tort law, see Shavell (2004, pp.189-190).

We can now write the dynamic optimization problem that the court must solve. Let δ be the discount factor and let $V(W_t, R_t)$ be the court's value function at time t, as a function of the state variables W_t, R_t. When writing an opinion in period 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 maximize its expected discounted payoff:

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

The first term is the expected cost of having to decide a case in period t. For example, if $a_t = b_t$, then the court reads the precedent as deciding the law for all activities. As such, the court 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 second and third terms reflect the expected one-period error losses. Consider the second 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. Given the lower bound, this case will be ruled by precedent as a permissible activity. The expression in square brackets is the probability the court attaches to the event that this case x_t should instead be ruled as an impermissible activity. The third term follows from a similar analysis on the upper region of the interval of uncertainty; here precedent induces the court to rule the case as impermissible, but, in fact, the case should be declared permissible.

The fourth 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. The expected value function can be written explicitly as the sum of three components:

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

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

$[a_t, b_t]$. 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. First, only the size $R_t - W_t$ of the interval $[W_t, R_t]$ matters to the court when deciding the interpretative bounds. Furthermore, because of symmetry of the distributions, it is optimal to set $a_t - W_t = R_t - b_t$ in all periods t. We will use this uniform version of the model in the remainder of the paper; in the appendix we show that the results and insights extend to the more general version.

Using (1) and (2), we can write the court's objective function for the uniform version:

$$\begin{aligned} V(R_t - W_t) \quad = \quad & \max_{W_t \leq a_t \leq b_t \leq R_t} \left\{ -C\left(b_t - a_t\right) - \frac{L}{2}\left(\frac{\left(a_t - W_t\right)^2}{R_t - W_t} + \frac{\left(R_t - b_t\right)^2}{R_t - W_t}\right) \right. \\ & + \delta V(R_t - W_t)\left[1 - b_t + a_t\right] \qquad (3) \\ & \left. + \delta \int_{a_t}^{b_t} V(R_t - x_t)\frac{R_t - x_t}{R_t - W_t}dx_t + \delta \int_{a_t}^{b_t} V(x_t - W_t)\frac{x_t - W_t}{R_t - W_t}dx_t \right\}. \end{aligned}$$

Lemma 1 in the appendix proves that the value function $V(R_t - W_t)$ exists and is unique.

3 The Value of Precedent and the Convergence of Doctrine

The model captures 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, p. 249).

As noted the court trades off the cost of having to spend resources examining a case anew against errors from inaccurate decision-making, keeping in mind the learning benefit to future judges from viewing each prior opinion as narrowly tailored to its facts. The most narrow construction of precedent in period t is to set $[a_t, b_t] = [W_t, R_t]$. In so doing, the court maximizes the learning that case load provides. The court looks at every possible case where the resolution is uncertain; each of these cases carries a bit of information. Thus, if the effort cost C of looking at a case is sufficiently small relative to the error loss L, shouldn't the court use all the information potentially available in each period, setting the bounds as $[a_t, b_t] = [W_t, R_t]$? 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.

Differentiating the court's objective function (3) with respect to a_t and b_t (and ignoring the constraints) gives the first order conditions

$$0 = C - L\frac{a_t - W_t}{R_t - W_t} + \delta V(R_t - W_t) - \delta V(R_t - a_t)\frac{R_t - a_t}{R_t - W_t} - \delta V(a_t - W_t)\frac{a_t - W_t}{R_t - W_t} \quad (4)$$

$$0 = -C + L\frac{R_t - b_t}{R_t - W_t} - \delta V(R_t - W_t) + \delta V(R_t - b_t)\frac{R_t - b_t}{R_t - W_t} + \delta V(b_t - W_t)\frac{b_t - W_t}{R_t - W_t}, \quad (5)$$

from which it is easy to see that at an optimal solution it is $a_t - W_t = R_t - b_t$, and hence $R_t - a_t = b_t - W_t$. Moreover, if the constraint $a_t \leq b_t$ binds, then $a_t = b_t = (R_t - W_t)/2$. Let $\Delta_t = R_t - W_t$ and define the precedent parameter λ_t,

$$\lambda_t = \frac{a_t - W_t}{\Delta_t}.$$

We can think of $2\lambda_t$ as the precedent ratio, the proportion of cases in the interval of uncertainty Δ_t that will be ruled by precedent in period t. Imagine $W_t = 0.1$, $R_t = 0.9$, $a_t = 0.3$ and $b_t = 0.7$. In that situation, $2\lambda_t = \frac{0.4}{0.8} = 1/2$. Suppose the court extrapolates more from precedent, setting $a_t = b_t = 0.5$, the precedent ratio becomes $2\lambda_t = \frac{0.8}{0.8} = 1$. As the bounds shrink relative to the level of uncertainty, λ_t increases, meaning that the court decides more and more cases by reference to precedent alone.

Write the first order conditions in terms of λ_t as:

$$0 = C - L\lambda_t + \delta V(\Delta_t) - \delta(1 - \lambda_t) V((1 - \lambda_t)\Delta_t) - \delta\lambda_t V(\lambda_t\Delta_t). \quad (6)$$

It is now immediately apparent that the court will never set the interpretative interval $[a_t, b_t]$ to coincide with the uncertainty interval $[W_t, R_t]$. The court always relies on precedent and never expends effort on all cases in the interval of uncertainty Δ_t.

Proposition 1. *In each period t the court chooses a positive precedent ratio, $\lambda_t > 0$ (hence $a_t > W_t$ and $b_t < R_t$).*

Proof. Suppose, to the contrary, that $\lambda_t = 0$, then (6) requires

$$0 = C,$$

which cannot hold as long as $C > 0$. ■

The value of precedent, *stare decisis*, emerges endogenously in our model. 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 $a_t = W_t + \varepsilon\Delta_t$ rather than $a_t = W_t$, where ε is a small number. Note that the probability that x is drawn from the interval $[W_t, a_t]$ is ε, while the probability that both x and θ are in the interval and $x < \theta$ is $\varepsilon^2/2$. It follows that the marginal benefit of relying on precedent, the expected saving on the cost of effort, is εC, while the marginal cost of relying on precedent is the expected social loss of an error, $\varepsilon^2 L/2$. No matter how small is C and how large is L, it is always $\varepsilon C > \varepsilon^2 L/2$ for a sufficiently small ε.

Maximally wide interpretative bounds, $\lambda_t = 0$, 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 the example given earlier, the court investigates the case of six year old children trespassing and getting injured in a sand pit with a poorly maintained fence, in spite of having a similar case with preschoolers on the books. 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 cases. Instead the court is spending judicial resources investigating the merits of every case that comes in the court house door.

The remaining part of this section studies how much time courts should spend refining doctrine. Legal academics, policy-makers, and advocates often critique the law articulated by courts as imperfect or wrong-headed. 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 shared the underlying values of those critiquing the decisions.

Note that the first order condition (6) says that a myopic court, one that only cares about the present ($\delta = 0$), would set $\lambda_t = C/L$. However, in writing the first order condition (6) we have disregarded the constraint $a_t \leq b_t$, which is equivalent to $\lambda_t \leq 1/2$.

It follows that if $C/L > 1/2$, then the constraint binds and it is optimal for a myopic court to set $\lambda_t = 1/2$, or equivalently $a_t = b_t$. When the cost of effort C is more than half the loss of a case error, it is optimal for a myopic court to exert no effort and let all cases be ruled by precedent. A myopic court with $C \geq L/2$ refuses to learn, because the cost of learning and refining the doctrine outweighs the benefit.

What about a forward looking court that cares about the future? The next proposition shows that such a 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. 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; the court chooses $\lambda_t = 1/2$, equivalently $a_t = b_t$ ($\neq \theta$ with probability one), whenever $\Delta_t \leq \Delta^S$, where*

$$\Delta^S = \frac{4(1-\delta)(2C-L)}{\delta L}.$$

(2) *If $C \leq L/2$ the court eventually fully learns; $\lambda_t < 1/2$ whenever $\Delta_t > 0$ and* $\lim_{t\to\infty} a_t = \lim_{t\to\infty} b_t = \theta$, *while* $\lim_{t\to\infty} \lambda_t = C/L$

Proof. Suppose the court chooses $\lambda_t = 1/2$ (equivalently, $a_t = b_t$) for all $\Delta_t \leq \Delta^S$. Then, using (3), for $\Delta_t \leq \Delta^S$ the value function becomes

$$V(\Delta_t) = -\frac{L\Delta_t}{4(1-\delta)}, \tag{7}$$

and we can write the first order condition (6) as

$$\begin{aligned} 0 &= C - \frac{L}{2} - \frac{\delta L \Delta_t}{4(1-\delta)} + \frac{\delta L \Delta_t}{16(1-\delta)} + \frac{\delta L \Delta_t}{16(1-\delta)} \\ &= C - \frac{L}{2} - \frac{\delta L \Delta_t}{8(1-\delta)}. \end{aligned} \tag{8}$$

The right hand side of (8) is the marginal gain from a small increase in λ_t evaluated at $\lambda_t = 1/2$. It follows that it is optimal to choose $\lambda_t = 1/2$ if and only if $\Delta_t \leq \Delta^S$, where Δ^S is given by

$$\Delta^S = \frac{4(1-\delta)(2C-L)}{\delta L}.$$

It also follows that if $C < L/2$ then it is optimal to set $\lambda_t < 1/2$ for all $\Delta_t > 0$. Since in this case with probability $1 - 2\lambda_t > 0$ it is $\Delta_{t+1} < (1 - \lambda_t)\Delta_t < \Delta_t$, we have $\lim_{t\to\infty} \Delta_t = 0$ and hence $\lim_{t\to\infty} a_t = \lim_{t\to\infty} b_t = \theta$. It also follows from (6) that $\lim_{t\to\infty} \lambda_t = C/L$. ■

If decision costs are greater than half the error loss, the doctrine stabilizes with imperfections remaining in the law. The court sets $a_t = b_t$ without knowing the exact location of θ. The court defines the law for all activities, but realizes the doctrine might not apply well in some circumstances. That said, correcting those imperfections is not cost-justified.

As noted above, 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 is "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. This is true even if the judges care solely about efficiency as the relevant benchmark for legal rules. 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.[13] The next proposition, an immediate corollary of Proposition 2, specifies the relationship between the scope of the "inaccuracy" in doctrine and the parameters of the model.

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 loss 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.

From Proposition 2, it is also immediate that the court will not learn at all, it will

[13] 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.

set $\lambda_1 = 1/2$, or $a_1 = b_1 = 1/2$, as long as $\Delta^S \geq 1$, or equivalently

$$C \geq \frac{L}{2} + \frac{\delta}{8(1-\delta)} L.$$

When the cost of examining a case relative to the loss from error is above a threshold, the court sets the law once and for all at the error minimizing point 1/2 without ever attempting to fine tune it. This result does not arise because parties prefer a settled but imperfect law and the court places weight on that preference. Instead the court sets the rule immediately because making the law better is never worth its effort. The two concerns are related, however. Parties might prefer settled law because predictability facilitates planning. In our model, if parties can easily circumvent the legal rule through proper planning (like, for example, in contract law), setting the rule "wrongly" for a host of situations will result in relatively low error costs.

Many common law doctrines can be seen as well-settled and imperfect. In the interest of space, we give two examples, one from contract law and one from tort law. Under contract law, the impossibility doctrine specifies that a party need not perform its contractual obligation if performance has become impossible. Scholars have articulated several efficiency justifications for this rule. Posner and Rosenfield (1987) suggest that the doctrine is used to assign a risk to the party in the contract who is the superior risk bearer. Sykes (1990) suggests that impossibility might be used to mitigate the overreliance problem associated with expectation damages.[14] Neither of these theoretical justifications map perfectly onto the case law. Contract law recognizes an impossibility defense even when the breaching party is a large conglomerate with well-diversified shareholders (Sykes, 2002), casting doubt on the efficient risk-bearer justification. The cases do not track the reliance mitigation justification either. In applying the doctrine, the courts inquire whether cost of performance has become prohibitive. As scholars have noted, this concern does not have much to do with whether overreliance by the promisee is likely to be a problem (Sykes, 2002). In short, the impossibility doctrine lives, but it is not artfully tailored to reflect underlying economic considerations. This is consistent with our model since the benefits of fine tuning the doctrine are unlikely to be high, especially because parties can allocate many risks explicitly by contract.

Turning to 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's is found negligent

[14] On the overreliance associated with contract damage remedies, see Rogerson (1984).

and the negligence causes the injury, the defendant is liable for all the resulting damage. Alternatively, if the defendant 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. This happens if the marginal investment in safety both reduces the amount the defendant can expect to pay (the expected loss) and the chance he has to pay anything at all. In that case, the private gains to investments in safety are bigger than the social gains. 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. Yet, it is probably not worth the judiciary's time to account for these concerns since they involve estimating the probability distribution of damages the defendant thinks he will have to pay. Instead, a rough rule that works most of the time is preferable: that is, the rule is set and additional evidence about risks of over-deterrence (useful in some subset of cases) is not considered.

4 The Likeness Principle and the Interpretation of Prior Case Law

In this section we ask a more fundamental question about the evolution of law. Will this evolution be consistent with the rule-of-law value that like cases be treated alike?[15] To address this question, it is useful to spell out further the evolution of legal doctrine.

As noted above, we can think of the intervals $[W_t, a_t]$ and $[b_t, R_t]$ as the interpretation by the judge of the informativeness of prior precedent. The interpretation of prior precedent changes over time. The ability to characterize precedent one way at one time and another way at another time provides flexibility, which creates room for courts to incorporate new information into the law. As the next proposition shows, if $\lambda_t < 1/2$ and learning occurs at time t, then with positive probability next period the interpretation of precedent will vary in an apparently inconsistent way.

We define the interpretation of precedent to be inconsistent if either $a_{t+1} > b_t$ or $b_{t+1} < a_t$. In the former case, activity x with $b_t < x < a_{t+1}$ was deemed not permissible

[15] On the topic, the legal and philosphical literature is vast, see Fuller (1958), Hart (1958), Raz (1979), Tamanaha (2004), McCubbins et al. (2010). Many influential scholars have stressed that like cases should be treated alike. See, for example, Rawls (1971, p.237); Dworkin (1977, p.113), Whittingham (1999, p.169).

at time t, but permissible at time $t+1$. In the latter case, activity x with $b_{t+1} < x < a_t$ was declared permissible at time t, but not permissible at time $t+1$. Here identical cases – those with the exact same relevant facts – are treated asymmetrically.

Proposition 4. *If $\lambda_t < 1/2$, then with positive probability the interpretation of prior case law at time $t+1$ is inconsistent with the interpretation at time t.*

Proof. It suffices to show that $a_{t+1} > b_t$ with positive probability; that is, some activity level x with $b_t < x < a_{t+1}$ is deemed not permissible at time t, while it is deemed permissible at time $t+1$. By symmetry, an entirely analogous argument could be made to show that $b_{t+1} < a_t$ with positive probability. Let $\varepsilon^* = (b_t - a_t)/\Delta_t$. Since $\lambda_t < 1/2$, it is $\varepsilon^* > 0$ and $a_t < x(\varepsilon) = b_t - \varepsilon\Delta_t \leq b_t$ for all ε in the interval $I^* = [0, \varepsilon^*)$. Since $\theta > x(\varepsilon)$ with positive probability, with positive probability $W_{t+1} = x(\varepsilon)$ for some $\varepsilon \in I^*$. Note that, for all $\varepsilon \in I^*$

$$\begin{aligned} a_{t+1} &= W_{t+1} + \lambda_{t+1}[R_t - W_{t+1}] = x(\varepsilon) + \lambda_{t+1}[R_t - x(\varepsilon)] \\ &= b_t - \varepsilon\Delta_t + \lambda_{t+1}[R_t - b_t + \varepsilon\Delta_t] \\ &= b_t - \varepsilon\Delta_t + \lambda_{t+1}[\lambda_t\Delta_t + \varepsilon\Delta_t]. \end{aligned}$$

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

$$\varepsilon < \frac{\lambda_t\lambda_{t+1}}{1-\lambda_{t+1}} = \varepsilon^{**}.$$

Hence, it is a positive probability event that in period t the selected case is $x(\varepsilon)$ with $\varepsilon \in [0, \min\{\varepsilon^*, \varepsilon^{**}\})$ and $\theta > x(\varepsilon)$. In such an event $a_{t+1} > b_t$ and the interpretation of prior case law at time $t+1$ is inconsistent with the interpretation at time t. ■

On the doctrinal path, activity levels that were deemed impermissible become permissible, and levels that were permissible become impermissible. The doctrine doesn't follow a monotone pattern. Despite this fact, the court is taking the right jurisprudential approach. An elaboration of the sand pit example from Section 2 shows the logic behind why this might happen. Suppose that $[W_t, R_t]$ is $[0.2, 0.8]$; the court has discovered that the owner with a poorly maintained fence should be held liable for injuries to preschoolers (case 0.8), while the owner with a tall fence and warning signs should not be held responsible for injuries to drunken teenagers (case 0.2). Suppose the court sets $[a_t, b_t] = [0.4, 0.6]$, splitting the difference on its estimate of the posterior to save decision costs. The next case is brought to court, say that case is $x = 0.58$; a group of eleven year old boys has gone over the fence, in spite of the warning signs, and injured themselves. Say that the court finds that the owner put in place a fence of above-average height and safety and rules that he should not be held liable. As a result, the court's updated posterior is $[W_{t+1}, R_{t+1}] = [0.58, 0.8]$. Given these beliefs, the court now makes its interpretative choice. We know that the court will always rely on precedent and never set $[a_{t+1}, b_{t+1}] = [0.58, 0.8]$. So, say the court chooses $[a_{t+1}, b_{t+1}] = [0.64, 0.74]$. In such a way, the court engages in inconsistent decision-making. For cases like $x = 0.62$, say ten year old boys jumping over a fence of above-average height, the owner was deemed liable at time t, but he becomes not liable at time $t + 1$. What looks like inconsistent decision-making is really the court optimally gathering and using information from the case $x = 0.58$. Basically, if the court sticks to the old doctrine, the probability of error is much higher. The expected error cost is bigger, so the court recalibrates the interpretative bounds to reflect the learning contained in case 0.58.

Inconsistent interpretation of prior cases does not immediately imply a violation of the likeness principle; it does not imply that similar "real" cases that are brought to court will be decided differently. It could be that the inconsistencies in interpretations are small and quickly corrected, so that the probability that similar cases will in fact be decided differently by the court is zero. We now show that this is not so. The evolution of jurisprudence from a court that cares about errors and judicial resources is not always consistent with the lofty rule-of-law value that identical cases be treated alike.

We begin by defining formally the *likeness principle*. We say that the evolution of jurisprudence follows the likeness principle from period t onward if the following two conditions hold: (1) if activity x is brought to court and judged permissible at time t, then there exists an open interval of activities $(x - \varepsilon, x + \varepsilon)$ which are never judged not permissible in any period $\tau > t$; (2) if activity x is brought to court and judged not permissible at time t, then there exists an open interval of activities $(x - \varepsilon, x + \varepsilon)$ which are never judged permissible in any period $\tau > t$.[16]

If the likeness principle is violated, then in some period t a case x is brought to court and judged one way and one case very similar to x is brought to court in some period

[16] Because it is a probability zero event that the same activity x that was brought to court at t will also be brought in the future, it is necessary to look at a small interval of activities $(x - \varepsilon, x + \varepsilon)$ around x.

after t and judged the opposite way. With this definition in hand, we can now prove the following proposition.

Proposition 5. *If $\lambda_t < 1/2$, then with positive probability the optimal evolution of doctrine violates the likeness principle from period t onward.*

Proof. Since $\lambda_t > 0$, the interval $[b_t, R_t]$ has a non-empty interior and with positive probability at time t an activity x is selected in the interval $(b_t, b_t + \mu\Delta_t)$, where μ is a constant to be defined. The activity is judged as being not permissible. Since the court does not learn, it does not update and $a_{t+1} = a_t$, $b_{t+1} = b_t$. As described in the proof of Proposition 4, if $\lambda_t < 1/2$, then with positive probability at $t+1$ the selected activity is some $x(\varepsilon) = b_t - \varepsilon\Delta_t$ and the court will set $a_{t+2} = b_t - \varepsilon\Delta_t + \lambda_{t+2}[\lambda_t\Delta_t + \varepsilon\Delta_t] > b_{t+1} = b_t$. It follows that if $\mu < -\varepsilon + \lambda_{t+2}[\lambda_t + \varepsilon]$ then all small intervals of activities around x are judged as permissible if brought to court at $t+2$. Since it is a positive probability event that one such activity be selected and brought to court at $t+2$, the likeness principle is violated with positive probability. Note that a similar argument leading to the violation of the principle could be constructed starting from an activity x that is judged to be permissible at t. ■

Surprisingly, and contrary to the conventional wisdom of development economists and legal scholars, the proposition suggests that treating like cases alike is not always a good idea. 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 above, this interpretative approach taxes judicial resources, without enough of an offsetting benefit from what can be learned from the case.

As noted above, the two assumptions driving this result are (1) that judicial resources are scarce and (2) the courts aren't sure what the proper scope of the legal rule should be. Starting from that premise, the likeness principle must bend with positive probability to accommodate efficient learning.

Development economists recommend the adoption of rule-of-law values across countries. We do not question the general value of a rational and reliable legal system. What this paper show is that some violation of the likeness principle, one of the central ingredients of the rule of law, is socially optimal and should be expected from a rational, benevolent court system. One strength of our model is that its core assumptions are likely to hold across cultures, bringing into question some of the suggested rule-of-law reforms. Yes, predictability in the application of legal rules is important, but so is making those rules better over time and conserving judicial resources.

5 Big Versus Small Jurisprudential Steps

This section asks two final questions about the evolution of law. First, when should we expect a court to rely heavily on precedent in making decisions? Second, when should the court proceed slowly in making law, gathering lots of information from the cases before setting the doctrine? Proceeding in small steps in our model happens when the court construes the prior case law narrowly.

As we already pointed out, we can think of $2\lambda_t$ as the precedent ratio in period t, the ratio of the sum of the sizes of the two intervals $[W_t, a_t]$ and $[b_t, R_t]$ relative to the size Δ_t of the uncertainty interval $[W_t, R_t]$. It captures the court's relative reliance on precedent: That is, the percentage of cases that the court decides based on precedent among all cases for which the court doesn't know the answer for sure. Our next proposition explains the factors that influence the court's reliance decision. This decision depends on four factors: (1) the decision cost C, (2) the error loss L, (3) how much the court knows about the optimal rule, Δ_t and (4) future's value (as measured by the discount factor δ). We have the following result.

Proposition 6. (1) *The precedent ratio interpretation, $2\lambda_t$, is an increasing function of C and a decreasing function of L and δ.* (2) *The effect on λ_t of an increase in Δ_t is unclear, but λ_t is a decreasing function of Δ_t if either (i) $C > L/2$ and λ_t is "close" to $1/2$, or (ii) $C < L/2$ and Δ_t is "close" to zero.*

Proof. Rewrite the first order condition of the court's maximization problem (6) as

$$\Phi(\cdot) \equiv C - L\lambda_t + \delta V(\Delta_t) - \delta(1-\lambda_t) V((1-\lambda_t)\Delta_t) - \delta\lambda_t V(\lambda_t \Delta_t) = 0. \tag{9}$$

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

$$\begin{aligned}
\Phi_C &= 1 > 0 \\
\Phi_L &= -\lambda_t < 0 \\
\Phi_\delta &= \left[V(\Delta_t) - (1-\lambda_t)\, V((1-\lambda_t)\,\Delta_t) - \lambda_t V(\lambda_t \Delta_t)\right] < 0 \\
\Phi_{\Delta_t} &= \delta \left[V'(\Delta_t) - (1-\lambda_t)^2\, V'((1-\lambda_t)\,\Delta_t) - \lambda_t^2 V'(\lambda_t \Delta_t)\right],
\end{aligned}$$

where the third inequality follows from $V(\Delta)$ being decreasing in Δ. The sign of Φ_{Δ_t} is undefined. Since $V' < 0$, concavity of V would be sufficient to guarantee $\Phi_{\Delta_t} < 0$. Unfortunately, the value function need not be concave. However, when (1) $C > L/2$ and λ_t is close to $1/2$, we can use equation (7) to infer that $V(\Delta)$ is approximately linear with $V' = -L/4(1-\delta)$ and hence $\Phi_{\Delta_t} \simeq \delta V'/2 < 0$. Moreover, when (2) $C < L/2$ and Δ_t is close to zero $\Phi_{\Delta_t} \simeq 2\delta\lambda_t [1-\lambda_t]\, V'(0) < 0$. ■

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 extrapolate to a greater extent the costs and benefits from one case to another; reasoning by analogy becomes relatively more important. A reduction in the error loss L has a similar effect. It reduces the loss incurred when the court decides by applying precedent alone and hence induces the court to let precedent control. While the cost of looking at the facts of a particular case x is only incurred in one period, the potential benefits in terms of learning and reduction of future errors spread over the entire future. As a consequence, the effect of an increase in the discount factor δ, the value attached to future decisions, makes the court want to rely less on precedent, have a narrower interpretation, so as to induce greater learning and a smaller number of future errors.

In these intuitive results, we see an explanation of a number of judge-made doctrines. As an example of high error loss, consider constitutional rulings. The Supreme Court, the only court that can change constitutional rulings, avoids them if possible.[17] In the framework of our paper, the Supreme Court wants to avoid making – and thereby potentially making an error in – constitutional decisions because such decisions are stickier than other rulings, which can be modified by legislation.

[17] See *Ashwander v. TVA*, 297 U.S. 288 (1936). "When the validity of an act of the Congress is drawn in question, and even a serious doubt of constitutionality is raised, it is the cardinal principle that this Court will first ascertain whether a construction of the statute is fairly possible by which the question can be avoided." (Brandeis, J. concurring).

Contract law provides an example of low error loss. The mailbox rule states that acceptance of a contract occurs upon dispatch.[18] The courts have refined this broad rule rarely.[19] Why have the courts been stingy with exceptions to the rule, not allowing modifications for other circumstances (where, say, the offeree tried to retract the acceptance by phone, after he had dispatched the acceptance by mail)? Our model suggests that the cost of applying the mailbox rule to circumstances where it doesn't quite fit is small. Contracting parties, after all, can just draft around the rule. And so, it makes sense that the mailbox rule has broad applicability and is rarely subject to judicially-created exceptions.

Part (2) of Proposition 6 studies how the court's reliance on precedent varies with uncertainty in the law and hence with time, since uncertainty (Δ_t) decreases over time. The relationship between uncertainty and the interpretation of precedent is of interest to legal scholars and legal philosophers. One of the most prominent theories of jurisprudence suggests that courts should move slowly in making law, be minimalist (Sunstein, 1999). Sunstein makes this suggestion most forcefully in areas where the court lacks information about the consequences of its decisions (Sunstein, 1999 p.5).

Translated into our model, minimalism means that the court should gather a lot of information each period. That way, the court avoids mistakes. In addition, there is a link between the gains to be had from information acquisition and uncertainty. The court should gather relatively more information – construe prior cases even more narrowly – as the range of uncertainty increases.

Proposition 6 shows that the interpretation of precedent need not change monotonically over time. The amount of information the court gathers can sometimes be (relatively) large and other times be small. A court rationally will be sometimes minimalist and other times not. This is true both over time and across areas of law.

The model produces sharper predictions for times when the court's learning process is reaching its end. This may occur either because the court is about to stop learning in spite of not knowing the true value of the parameter θ (case (i) in part (2) of Proposition 6), or because the true value of θ is already almost perfectly known (case (ii)). In such situations, as learning approaches it stopping point, the model is consistent with

[18] See *Adams v. Lindsell*, 106 Eng. Rep. 250 (1818); Restatement (Second) of Contracts 63 (a).

[19] There is one primary exception. In the case of option contracts, acceptance occurs upon receipt. See *Santos v. Dean*, 96 Wash.App. 849 (1999); Restatement (Second) of Contracts 63(b).

Sunstein's view; the court relies on precedent more and more, comfortable that it knows enough about the optimal policy to avoid spending more resources gathering information.

One may object that rather than looking at the precedent ratio relative to the uncertainty in the law, one should look at the absolute ratio. Indeed, one may expect that as the uncertainty (Δ_t) decreases over time, the absolute precedent ratio will naturally decrease. Thus, contrary to Sunstein's view, one may expect our model to deliver initially broad readings of precedent (in an absolute sense) followed by narrower readings. Surprisingly, this is not necessarily so. Think of the actual size $2\lambda_t\Delta_t$ of the two intervals $[W_t, a_t]$ and $[b_t, R_t]$ as a measure of the "absolute precedent ratio". Then a reduction in Δ_t has a direct negative effect on the absolute ratio, as expected. However, such a negative effect may be more than offset by a positive effect on the relative ratio λ_t. Indeed, this will certainly be the case if Δ_t is close to zero or if λ_t is close to $1/2$. In short, it can be shown that Proposition 6 holds, unchanged, for the absolute ratio $\lambda_t\Delta_t$.

6 Extensions and Conclusion

Decision-making by precedent is not confined to courts. Our results are applicable to any situation where decision-makers learn from experience. Take a frontline 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 frontline 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. The list of applications of the general model can be extended to any involving case by case decision-making, such as decanal evaluation of student complaints, the parenting of children, and so on.

Turning back to judge-made law, we view our model as the foundation for a general theory of rational jurisprudence. Our analysis might be extended in a number of ways.

We mention two of them. First, we assume that judges share the same normative values. This is obviously not true, especially in the "hot button" cases. What's surprising is that the model has descriptive power, while maintaining such a strong assumption. For instance, we see optimal deviations from the likeness principle arise even when nothing about judicial preferences or the underlying environment has changed. The model might be modified to include judges with different preferences. 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. A second worthwhile extension is 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 (O'Brien, 2004). 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 show that Proposition 1 and 2, proven in the main body for the uniform model, extends to the general model.

Lemma 1. *The value function $V(\Delta)$ defined in (3) is uniquely defined.*

Proof. Dropping the subscript t, let $\Delta = R - W$ and $\Lambda = \lambda\Delta$. Exploiting symmetry, we can rewrite the court's dynamic optimization problem (3) as

$$\begin{aligned} V(\Delta) \;\; = \;\; & \max_{0\leq\Lambda\leq\Delta/2} -C(\Delta-2\Lambda) - L\frac{\Lambda^2}{\Delta} \\ & +\delta V(\Delta)\left[1-\Delta+2\Lambda\right] + 2\delta\int_{\Lambda}^{\Delta-\Lambda} V(x)\frac{x}{\Delta}dx. \end{aligned}$$

Define the operator T, mapping the metric space S of continuous, real-valued functions W on $[0,1]$ into itself:

$$\begin{aligned} TW(\Delta) \;\; = \;\; & \max_{0\leq\Lambda\leq\Delta/2} -C(\Delta-2\Lambda) - L\frac{\Lambda^2}{\Delta} \\ & +\delta W(\Delta)\left[1-\Delta+2\Lambda\right] + 2\delta\int_{\Lambda}^{\Delta-\Lambda} W(x)\frac{x}{\Delta}dx. \end{aligned}$$

This mapping produces a new guessed value function TW starting from a guessed value function W. Let the metric on S be $\rho(W^0, W^1) = \max_{\Delta\in[0,1]} \left|W^0(\Delta) - W^1(\Delta)\right|$. We now show that T is a contraction mapping. 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 $W^0(\Delta) \leq W^1(\Delta)$ for all $\Delta \in [0,1]$. It is immediate that $TW^0(\Delta) \leq TW^1(\Delta)$ and hence monotonicity holds. (2) To see that discounting also holds, let a be a non negative constant. Then

$$\begin{aligned}
T(W+a)(\Delta) &= \max_{0\leq\Lambda\leq\Delta/2} -C\left(\Delta-2\Lambda\right)-L\frac{\Lambda^2}{\Delta} \\
&\quad +\delta\left[W(\Delta)+a\right]\left[1-\Delta+2\Lambda\right]+2\delta\int_{\Lambda}^{\Delta-\Lambda}\left[W(x)+a\right]\frac{x}{\Delta}dx \\
&= \max_{0\leq\Lambda\leq\Delta/2} -C\left(\Delta-2\Lambda\right)-L\frac{\Lambda^2}{\Delta} \\
&\quad +\delta W(\Delta)\left[1-\Delta+2\Lambda\right]+2\delta\int_{\Lambda}^{\Delta-\Lambda}W(x)\frac{x}{\Delta}dx \\
&\quad +\delta a\left[1-\Delta+2\Lambda\right]+\delta\frac{a}{\Delta}\left[\left(\Delta-\Lambda\right)^2-\Delta^2\right] \\
&= TW(\Delta)+\delta a.
\end{aligned}$$

This proves that T is a contraction, and hence it has a unique fixed point, the true value function V, which can be obtained by successive iterations starting from an arbitrary guess. ■

We now prove Propositions 1 and 2 for the general model.

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

$$\begin{aligned}
0 &= C-L\frac{F(a_t)-F(W_t)}{F(R_t)-F(W_t)}+\delta V(W_t,R_t) \\
&\quad -\delta V(a_t,R_t)\frac{F(R_t)-F(a_t)}{F(R_t)-F(W_t)}-\delta V(W_t,a_t)\frac{F(a_t)-F(W_t)}{F(R_t)-F(W_t)}-\frac{(\gamma_t-\mu_t^a)}{g(a_t)}
\end{aligned} \tag{10}$$

$$\begin{aligned}
0 &= -C+L\frac{F(R_t)-F(b_t)}{F(R_t)-F(W_t)}-\delta V(W_t,R_t) \\
&\quad +\delta V(b_t,R_t)\frac{F(R_t)-F(b_t)}{F(R_t)-F(W_t)}+\delta V(W_t,b_t)\frac{F(b_t)-F(W_t)}{F(R_t)-F(W_t)}+\frac{\left(\gamma_t-\mu_t^b\right)}{g(b_t)}.
\end{aligned} \tag{11}$$

Proposition 1′. *In each period t the court chooses $a_t > W_t$ and $b_t < R_t$.*

Proof. Suppose $a_t = W_t < R_t$. Then (10) becomes $C = \frac{(\gamma_t-\mu_t^a)}{g(a_t)}$, which can only be satisfied if $b_t = a_t$. Equation (11) becomes $C - L = \frac{\left(\gamma_t-\mu_t^b\right)}{g(b_t)}$, which can only be satisfied if $b_t = R_t$. This is contradiction. Hence it must be $a_t > W_t$ whenever $W_t < R_t$.

Similarly, suppose $b_t = R_t > W_t$. Then (11) becomes $C = \frac{(\gamma_t - \mu_t^b)}{g(a_t)}$, which requires $a_t = b_t$, while (10) becomes $C - L = \frac{(\gamma_t - \mu_t^a)}{g(a_t)}$, which requires $a_t = W_t$, a contradiction. Hence it must be $b_t < R_t$ whenever $R_t > W_t$. ■

Proposition 2′. (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. 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 $W_t \neq R_t$. Assume $a_t = b_t$ and $W_t \neq R_t$. Since, as we have shown in Proposition 1′, it is $\mu_t^a = \mu_t^b = 0$, adding up (10) and (11) we obtain

$$F(a_t) = F(b_t) = \frac{F(R_t) + F(W_t)}{2}. \tag{12}$$

Replacing such value, and $\mu_t^a = \mu_t^b = 0$, into (10) and (11) yields

$$2C - L + 2\delta V(W_t, R_t) - \delta V(a_t, R_t) - \delta V(W_t, a_t) - \frac{2\gamma_t}{g(a_t)} = 0 \tag{13}$$

$$-2C + L - 2\delta V(W_t, R_t) + \delta V(b_t, R_t) + \delta V(W_t, b_t) + \frac{2\gamma_t}{g(b_t)} = 0. \tag{14}$$

Since $V(W_t, R_t) < V(a_t, R_t)$ and $V(W_t, R_t) < V(W_t, a_t)$, the conditions above cannot be satisfied if $2C < L$. It follows that it cannot be $a_t = b_t$, and hence learning will never stop if $C < L/2$.

Now suppose $a_t \neq b_t$ for all values of $W_t < R_t$, and as a consequence, $\gamma_t = 0$. Using this and $\mu_t^a = \mu_t^b = 0$, by subtracting (11) from (10) we obtain

$$\begin{aligned} 0 \;=\; & 2C - L\left(1 - \frac{F(b_t) - F(a_t)}{F(R_t) - F(W_t)}\right) + 2\delta V(W_t, R_t) \\ & -\delta V(a_t, R_t)\frac{F(R_t) - F(a_t)}{F(R_t) - F(W_t)} - \delta V(b_t, R_t)\frac{F(R_t) - F(b_t)}{F(R_t) - F(W_t)} \\ & -\delta V(W_t, a_t)\frac{F(a_t) - F(W_t)}{F(R_t) - F(W_t)} - \delta V(W_t, b_t)\frac{F(b_t) - F(W_t)}{F(R_t) - F(W_t)}. \end{aligned} \tag{15}$$

Note that for any $\varepsilon > 0$, there exists W_t and R_t sufficiently close to each other, so that the right hand side of (15) is greater than $2C - L + \varepsilon$. It follows that if $C > L/2$, then (15) cannot hold for such values of W_t and R_t; hence it cannot be the case that $a_t \neq b_t$ for all values of $W_t < R_t$. Learning will eventually stop if $C > L/2$. ■

References

1. Bailey, Martin and Rubin, Paul. 1994. "A Positive Theory of Legal Change." *International Review of Law and Economics*, vol. 14, pp. 467-77.

2. Blackwell, David. 1965. "Discounted Dynamic Programming." *Annals of Mathematical Statistics*, vol. 36, pp. 226-235.

3. Calfee, John E. and Craswell, Richard. 1984. "Some Effects of Uncertainty on Compliance with Legal Standards." *Virginia Law Review*, vol. 70, pp. 965-1003.

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

5. Cooter Robert, Kornhauser, Lewis and Lane, David. 1979. "Liability Rules, Limited Information, and the Role of Precedent." *Bell Journal of Economics*, vol. 10, pp. 366-373.

6. Dari-Mattiacci, Giuseppe, Deffainis, Bruno and Lovat, Bruno. 2010 "The Dynamics of the Legal System." *Journal of Economic Behavior and Organization*, forthcoming.

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

8. Eskridge, William and Ferejohn, John. 1992. "The Article I Section 7 Game." *Georgetown Law Journal*, vol. 80, pp.523-564

9. Fuller, Lon. L. 1958. "Positivism and Fidelity to Law – A Reply to Professor Hart." *Harvard Law Review*, vol. 71, pp. 630-672

10. Glazer, Jacob & Rubinstein, Ariel. 2004. "On Optimal Rules of Persuasion." *Econometrica* vol. 72, 1715-1736.

11. Gennaioli, Nicola and Shleifer, Andrei. 2007. "The Evolution of the Common Law," *Journal of Political Economy*, vol. 115, pp. 43-68.

12. Hadfield, Gillian K. 1992. "Bias in the Evolution of Legal Rules." *Georgetown Law Journal*, vol. 80, pp. 583-616.

13. Hadfield, Gillian K. 2009. "The Dynamic Quality of Law: The Role of Judicial Incentives and Legal Human Capital." mimeo, USC law school.

14. Hart, H.L.A. 1958. "Positivism and Fidelity to the Separation of Law and Morals," *Harvard Law Review*, vol. 71, pp. 593-629.

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

16. Hyton. 2006. "Information, Litigation, and the Common Law Evolution." *American Law and Economics Review*, vol. 8, pp. 33-61.

17. Klerman, Daniel. 2007. "Jurisdictional Competition and the Evolution of the Common Law," *University of Chicago Law Review*, vol. 74, pp. 1179-1226.

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

19. La Porta, Rafael, Florencio Lopez-de-Silanes, Cristian Pop-Eleches, and Andrei Schleifer. 2004. "Judicial Checks and Balances." *Journal of Political Economy*, vol. 106, pp. 445-470.

20. Levi, Edward H. 1948. An Introduction to Legal Reasoning. University of Chicago Press: Chicago IL.

21. McNollgast. 1995. "Politics and the Courts: A Positive Theory of Judicial Doctrine and the Rule of Law." *Southern California Law Review*, vol. 68, pp. 1631-1683.

22. McCubbins, Matthew D., Rodriguez, Daniel B. and Weingast, Barry R. 2010. "The Rule of Law Unplugged." Univ. of Texas Law, Public Law Research Paper no. 154.

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

24. O'Brien, David M. 2004. Judges on Judging. 2d. ed. Washington D.C.: CQ Press.

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

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

27. Priest, George L. 1977. "The Common Law Process and the Selection of Efficient Rules." *Journal of Legal Studies*, vol. 6, pp. 65-82.

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

29. Posner, Richard A. and Andrew Rosenfield. 1977. "Impossibility and Related Doctrines in Contract Law: An Economic Analysis." *Journal of Legal Studies*, vol. 6, pp. 83-188.

30. Rawls, John. 1971. A Theory of Justice. Cambridge: Harvard University Press.

31. Raz, Joseph. 1979. The Authority of Law. Oxford: Oxford University Press.

32. Restatement (Second) of Contract Law. 1981. American Law Institute.

33. Rogerson, William P. 1984. "Efficient Reliance and Damage Measures for Breach of Contract." *Rand Journal of Economics*, vol. 15, pp. 39-53.

34. Rubin, Paul H. 1977. "Why is the Common Law Efficient?" *Journal of Legal Studies*, vol. 6, pp. 51-63.

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

36. Sunstein, Cass R. 1999. One Case at a Time: Judicial Minimalism on the Supreme Court. Boston: Harvard University Press.

37. Songer, Donald, Segal, Jeffrey A. & Cameron, Charles M. 1994. "The Hierarchy of Justice: Testing a Principal-Agent Model of Supreme Court-Circuit Court Interactions." *American Journal of Political Science*, vol. 38, pp. 673-696.

38. Spitzer, Matthew and Talley, Eric. 2009. "Panel Effects and Judicial Auditing." mimeo, University of Southern California.

39. Startz, Richard and Yoon, Albert. 2009. "Litigant Resources and the Evolution of Legal Precedent." mimeo.

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

41. Sykes, Alan O. 1990. "The Doctrine of Commerical Impracticability in a Second-Best World." *Journal of Legal Studies*, vol. 19, pp. 43-94.

42. Sykes, Alan O. 2002. "Impossibility Doctrine in Contract Law." In New Palgrave Dictionary of Economics and the Law, edited by Peter Newman, New York: MacMillion Press.

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

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

45. Zywicki, Todd. 2003. "The Rise and Fall of Efficiency in the Common Law: A Supply-Side Explanation." 97 *Northwestern Law Review*, vol, 97, pp. 1551-163.

Cases Cited

1. Adams v. Lindsell, 106 Eng. Rep. 250 (1818).

2. Ashwander v. TVA, 297 U.S. 288 (1936).

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

4. Barklett v. Heersche, 462 P.2d 763 (1969).

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

6. Santos v. Dean, 96 Wash.App. 849 (1999).

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

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