The Evolution of Common Law

Nicola Gennaioli

University of Stockholm

Andrei Shleifer

Harvard University

We present a model of lawmaking by appellate courts in which judges influenced by policy preferences can distinguish precedents at some cost. We find a cost and a benefit of diversity of judicial views. Policy-motivated judges distort the law away from efficiency, but diversity of judicial views also fosters legal evolution and increases the law's precision. We call our central finding the Cardozo theorem: even when judges are motivated by personal agendas, legal evolution is, on average, beneficial because it washes out judicial biases and renders the law more precise. Our paper provides a theoretical foundation for the evolutionary adaptability of common law.

I. Introduction

In a common-law legal system, such as that of the United States and the United Kingdom, many important laws are made not by legislatures but by appellate courts deciding specific cases and thus creating precedents. Judge-made law is dominant in commercial areas of law, such as contract, property, and tort law. Judge-made legal rules promote or undermine economic efficiency when the Coase (1960) theorem does

We are grateful to Olivier Blanchard, Filipe Campante, Edward Glaeser, Claudia Goldin, Oliver Hart, Elhanan Helpman, Fausto Panunzi, Torsten Persson, Richard Posner, Ilia Rainer, Alan Schwartz, Kathryn Spier, David Strömberg, an anonymous referee, the editor of this *Journal*, and especially Louis Kaplow for helpful comments.

[*Journal of Political Economy,* 2007, vol. 115, no. 1]

not apply. Yet compared to the vast body of research on legislative lawmaking, judicial lawmaking has been relatively neglected by economists.

In this paper, we present a simple model of lawmaking that emphasizes the role of judicial preferences. Our model addresses both positive and normative questions about the evolution of judge-made law. Under what circumstances does legal evolution occur? What form does it take? Is it on average beneficial? What is the relationship between polarization of judicial preferences, volatility of legal rules, and welfare? Does the law ultimately converge to efficiency?

At least three areas of scholarship have tackled these issues. First, free-market philosophers such as Hayek (1960, 1973) and Leoni (1961) praised judge-made law for its role in preserving freedom. To them, decentralized evolution of law through primarily apolitical judicial decisions is vastly preferable to centralized yet arbitrary lawmaking by legislatures.[1] Consistent with these ideas, La Porta, Lopez-de-Silanes, Pop-Eleches, and Shleifer (2004) find a positive relationship in a cross section of countries between economic freedom and a proxy for recognition of judicial decisions, as opposed to just legislation, as a source of law. Beck, Demirguc-Kunt, and Levine (2003, 2005) argue further in the Hayek tradition that judge-made law is more adaptable than statutes. They suggest that such adaptability might be important for financial markets, and they find evidence that recognition of judge-made law predicts financial development and might account for the finding of La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1998) of the superior development of financial markets in common-law compared to the civil-law countries.

Second, the legal realist tradition in American jurisprudence, which in contrast to the free-market philosophers emphasizes that judges make decisions based on their political and other beliefs, nonetheless concludes that judge-made law evolves for the better (e.g., Llewellyn 1960, 402; see also Holmes 1897; Cardozo 1921; Radin 1925; Frank 1930; Llewellyn 1951; Stone 1985; Posner 2005). Perhaps the most famous assessment of this evolutionary process is Judge Benjamin Cardozo's (1921, 177):

> The eccentricities of judges balance one another. One judge looks at problems from the point of view of history, another from that of philosophy, another from that of social utility, one is a formalist, another a latitudinarian, one is timorous of change, another dissatisfied with the present; out of the attrition of diverse minds there is beaten something which has a

[1] Glaeser and Shleifer (2002) and Ponzetto and Fernandez (2006) explicitly compare legal rules produced by judges and legislators.

> constancy and uniformity and average value greater than its component elements.

Third, the evolution of common-law rules and their convergence to efficiency have been taken up in law and economics. In his *Economic Analysis of Law* ([1973] 2003), Posner hypothesizes that common law tends toward efficiency, largely on the basis of the argument that judges maximize efficiency. Cooter, Kornhauser, and Lane (1979) find that decision making by welfare-maximizing but imperfectly informed judges improves the law over time. Priest (1977) and Rubin (1977) further suggest that disputes involving inefficient legal rules are more likely to be taken to court rather than settled, leading to the replacement of such rules by better ones over time. Cooter and Kornhauser (1980) formally show that the law tends to improve if inefficient rules are more likely to be replaced than the efficient ones.[2]

These diverse strands of research do not share a common framework for studying the evolution of judge-made legal rules and evaluating their efficiency. In such a framework, there must be judges, these judges must be able to make decisions based on their preferences, and questions about the evolution and the quality of law must still be possible to address. For even if judicial decisions are governed by ideologies and biases rather than maximization of efficiency, the evolutionary process may still improve the law. When does legal evolution warrant the optimistic assessments of free-market philosophers, legal realists, and law and economics scholars?

To address these issues, we present a model in which precedents evolve through a series of decisions by appellate judges. Our model relies on three assumptions. First, following the legal realists and a modeling strategy of Gennaioli (2004), we assume that judges hold biases favoring different types of disputants and that these biases vary across the population of judges. Political scientists document the importance of judicial attitudes in shaping appellate rulings. The U.S. Supreme Court judges sometimes vote on the basis of their ideological preferences and distinguish precedents incompatible with their political orientation (George and Epstein 1992; Brenner and Spaeth 1995; Songer and Lindquist 1996). Extreme judges are more likely to vote against the precedent than the centrist ones (Brenner and Stier 1996). Hansford and Spriggs

[2] Some law and economics scholars are more skeptical about the efficiency of common law. Hadfield (1992) argues that it is not necessarily the case that the less efficient precedents are more likely to be challenged. Zywicki (2003) believes that common-law evolution used to be efficient in the nineteenth century but is no longer so because judges are excessively influenced by their own preferences as well as by rent-seeking pressures from litigants. Hathaway (2001) discusses how the doctrine of *stare decisis* introduces path dependence into the law that is not conducive to efficiency.

(2006) conclude that the decision to follow a precedent depends on the ideological distance between the preferences of the deciding Supreme Court and the precedent itself.[3]

Second, following Radin (1925) and Posner (2003, 2005), we assume that changing precedent is personally costly to judges: it requires extra investigation of facts, extra writing, extra work of persuading colleagues when judges sit on panels, extra risk of being criticized, and so on. "Judges are people and the economizing of mental effort is a characteristic of people, even if censorious persons call it by a less fine name" (Radin 1925, 362). The assumption that, other things equal, judges would rather not change the law implies that only the judges who disagree with the current legal rule strongly enough actually change it. Posner (2003, 544) sees what he calls "judicial preference for leisure" as a source of stability in the law; we revisit this issue.

Third, we assume that the law evolves when judges distinguish cases from precedents rather than just overrule precedents. By "distinguishing" we mean the introduction of a new legal rule that endorses the existing precedent but adds a new material dimension to adjudication and holds that the judicial decision must depend on both the previously recognized dimension and the new one. Distinguishing cases is the central mechanism, or leeway, through which common law evolves despite binding precedents (Llewellyn 1960; Stone 1985).[4]

Using a model relying on these three assumptions, we examine the evolution of legal rules in the case of a simple tort: a dog bites a man (e.g., Landes and Posner 1987). In this analysis, two general principles stand out. First, legal change enables judges to implement their own biases and to undo those of their predecessors. Second, such change occurs more often when judges' preferences are polarized because judges are more likely to strongly disagree with the current precedent. Putting these principles to work, we find a cost and a benefit of judicial polarization. On the one hand, biased judges distort the law away from efficiency. On the other hand, by fostering legal evolution, diversity of judicial views improves the quality of the law because, irrespective of whether the judge changing the law is biased or efficiency-oriented, distinguishing brings new data into dispute resolution, thereby increasing the precision of legal rules. Consistent with this trade-off, we find that greater polarization of judicial opinions may lead to better law.

[3] Klein (2002) finds that judicial attitudes matter also for appellate court decisions. Legal academics increasingly accept the importance of judicial ideologies for rulings on politically sensitive issues (e.g., Rowland and Carp 1996; Revesz 1997; Pinello 1999; Sunstein, Schkade, and Ellman 2004).

[4] In Gennaioli and Shleifer (2007), we consider the evolution of law when judges can overrule prior precedents. This technique of legal change is uncommon and generally leads to volatility and unpredictability of legal rules. In our framework, overruling is not conducive to either convergence or efficiency.

Although the cost of judicial bias renders the conditions for full efficiency of judge-made law implausibly strict, in our model legal evolution is beneficial *on average*, even if judges are extremely biased. In line with Cardozo's optimism, judicial biases wash out on average, and the informational benefit of distinguishing improves the quality of the law.

These findings provide a theoretical foundation for the evolutionary adaptability of common law. They further suggest that such adaptability is more beneficial in the areas of law in which there is room for change and updating but the disagreement among the judges is not extreme. The relatively apolitical yet still changing areas of law, such as contract and corporate law, are the likely candidates for reaping the benefits from the decentralized evolution of judge-made law.

II. A Model of Legal Precedent

There are two parties, a dog owner O and a bite victim V, as well as a dog. The dog bit V, who seeks to recover damages from O. The dog was not on a leash, so to assess O's liability, one should determine whether O was negligent (and so is liable) or not (and so is not).

Let P_{NP} be the probability that the dog bites V if O does not take precautions (he does not put it on a leash) and P_P the probability that the dog bites V if precautions are taken. Let C be the owner's cost of precautions (e.g., the costs of putting the dog on a leash). First-best efficiency requires that the dog owner takes precautions if and only if their cost C is lower than the reduction in the probability of a bite (weighted by V's harm, which we normalize to one).

We assume that damages are always set high enough to enforce precautions whenever the law holds O liable.[5] As a result, as indicated by the Hand formula, the first-best is implemented by holding O negligent and thus liable if $P_{NP} - P_P \geq C$ and not liable if $P_{NP} - P_P < C$. In this context, the question for the law is how to determine negligence from the facts of a case.

Many factual situations may influence the probability of a bite and thus whether O was negligent. We assume that only two empirical dimensions—the dog's aggressiveness and the location of the interaction between O and V—are *material* to determine liability in this legal dispute.

Variable $a \in [0, 1]$ measures the dog's aggressiveness. A dog with $a = 0$ is very peaceful (a golden retriever) and less likely to bite V than a dog with $a = 1$ (a pit bull). Variable $d \in [0, 1]$ measures the density of people in the location in which the dog is walked: if $d = 0$, the bite

[5] For instance, the court may award punitive damages. If damages are equal to harm, then O does not take over-precautions and, in addition, strict liability (rather than negligence) yields the first-best. Yet, even strict liability leads to over-precautions if O does not fully internalize the dog's cost of precautions.

occurred in a forest; if $d = 1$, it occurred on a playground. We assume that a and d are independently and uniformly distributed over the population of interactions between O and V. We further assume that

$$P_{NP} - P_P = \begin{cases} \overline{\Delta P} & \text{for } a + d \geq 1 \\ \underline{\Delta P} & \text{for } a + d < 1, \end{cases} \tag{1}$$

where $\overline{\Delta P} > C > \underline{\Delta P}$. Thus O is optimally held liable if and only if $a + d \geq 1$. Even owners of peaceful dogs are optimally held liable if they did not take precautions on a playground ($d = 1$); even owners of violent dogs are optimally excused if the dog was unleashed in a forest ($d = 0$).

In general, the social benefit of putting the dog on a leash is a function $\Delta P(a, d)$ increasing in a and d. We could allow for more general functions, but to clarify our analysis of legal change we assume that $\Delta P(a, d)$ depends only on $a + d$ and "jumps" at $a + d = 1$. The first restriction makes a and d symmetric for determining liability, which allows us to isolate the effect of legal change per se, abstracting from the particular dimension introduced into the law. The second restriction allows us to separate the *probabilities* of the different errors induced by a legal rule from their welfare *cost*.

A legal rule in this environment attaches a legal consequence (O liable, O not liable) to every case (a, d). In a sense, different legal rules put different substantive content into Hand's formula by specifying how $P_{NP} - P_P$ must be determined from a case's empirical attributes (a, d).

How do appellate judges make legal rules? We assume that, when no legal rule deals with dog bites at the beginning, the only factual issue that comes up through trial is the aggressiveness of the dog. As a result, the appellate judge who reviews the case sets the rule by choosing a threshold on a, which we call A. Owners of dogs more aggressive than A are held liable; owners of dogs less aggressive than A are not. We can think of A as the *ratio decidendi*—the principle of decision—of the case (Goodhart 1930; Stone 1985). At this point the judge cannot set a rule in which liability also depends on d, since a well-established principle of common law holds that judges consider only cases and factual dimensions that come before them (Llewellyn 1951).

Once A is set, a later judge dealing with a dog bite must respect *stare decisis*, or adherence to precedent, and so accept A. However, as soon as the issue of location of a bite is brought on appeal, the judge can still radically change the law by *distinguishing* the case from the precedent based on this previously neglected dimension d. True, this golden retriever is gentle, but it was unleashed on a playground. True, this pit bull is dangerous, but who would reasonably keep it on a leash in a forest? Effectively, the judge introduces d into adjudication, applying the previous precedent to only some of the cases in the (a, d) space,

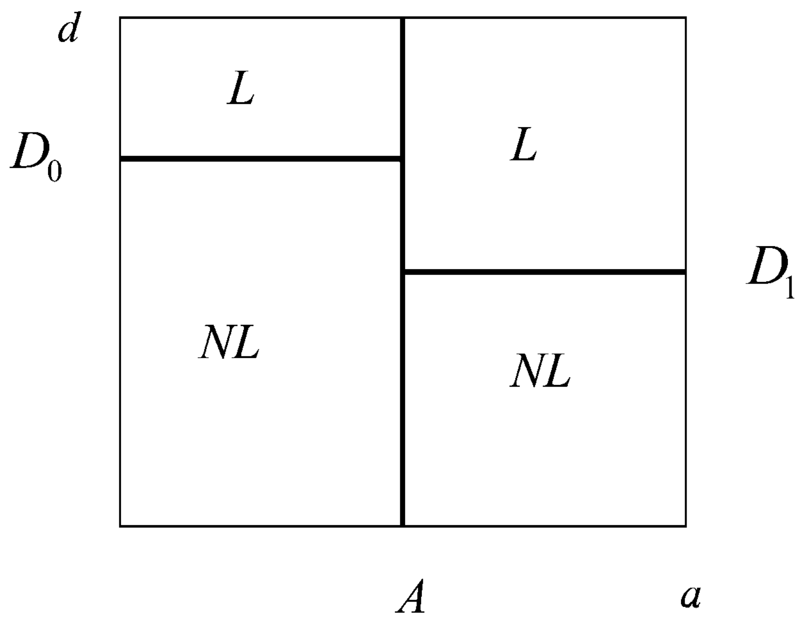

FIG. 1

but not to others. Such distinguishing is the main source of legal change in common-law systems (Llewellyn 1960; Stone 1985).

Because the judge respects the initial precedent, the best he can do is to choose two thresholds on density D_0 and D_1. The rule he establishes thus takes a two-dimensional threshold form, as illustrated in figure 1.

In figure 1, O is held liable in regions denoted by L but nonliable in those denoted by NL. Relative to a one-dimensional rule, a two-dimensional rule allows for liability of owners of peaceful dogs ($a \leq A$) in crowded locations ($d \geq D_0$) and nonliability of owners of aggressive dogs ($a > A$) in deserted locations ($d < D_1$). We expect $D_0 \geq D_1$. To take an extreme example of the power of distinguishing, if the first judge sets $A = 0$, the second judge can reverse it completely by saying that liability exists only in the most crowded locations, that is, by setting $D_0 = D_1 = 1$, which eliminates owner liability entirely. Although distinguishing by the second judge may allow him to render the law more precise (as when he sets $D_0 = D_1 = \frac{1}{2}$), the second judge can also distinguish strategically for the sole purpose of implementing his bias.

Sequential decision making by judges gives legal rules their threshold structure. Because only the dog's aggressiveness comes up through initial fact-finding, the first judge sets the threshold on aggressiveness beyond which O is liable. When future judges distinguish the case, they accept this initial threshold but, by bringing d into the law, create a two-dimensional threshold rule.

We assume that a judge distinguishing the case from precedent incurs a personal (fixed) effort cost k. The model's timing is as follows. At $t = 0$, the first judge establishes the aggressiveness threshold A. This precedent guides adjudication until another judge (if any) changes the rule at some t'. If at $t = t'$ a judge changes the rule, he sets two density thresholds, D_0 and D_1. In that case, the law is permanently fixed since there are no further material dimensions to introduce.

We now investigate the efficient—welfare-maximizing—rules that provide the normative benchmark for our analysis of legal change and judge-made law in Sections IV–VI.

III. Optimal Legal Rules

A dog owner finding himself in situation (a, d) decides whether to put the dog on a leash by considering the risk of liability under the prevailing legal rule at (a, d). First-best welfare, achieved under optimal precautions (i.e., O puts the dog on a leash whenever $a + d \geq 1$), is equal to

$$W^{FB} = -\tfrac{1}{2}\underline{\Delta P} - \tfrac{1}{2}C, \tag{2}$$

where the probability of a bite when precautions are taken is normalized to zero. In half the cases, precautions are not efficient and the parties bear the extra risk $\underline{\Delta P}$ of the dog biting the man; in the other half, precautions are efficient and cost C to society.

Judge-made law cannot achieve such high welfare because legal rules arrived at sequentially take threshold form. Consider a one-dimensional threshold rule A, which holds O liable if and only if $a \geq A$. Figure 2 represents it as a vertical bold line together with the first-best in the (a, d) space.

In the first-best, O is liable above the diagonal but not below. The one-dimensional rule A holds O mistakenly liable in region $L|NL$ and mistakenly not liable in region $NL|L$. In the former region, O takes excessive precautions, which cost $\Lambda^{\text{over}} = C - \underline{\Delta P}$ to society. In the latter region, O takes too few precautions, which cost $\Lambda^{\text{under}} = \overline{\Delta P} - C$ to society. We take these costs of over- and under-precautions as given and focus on how different legal rules affect the likelihood of different errors in adjudication. We also define $\lambda = \Lambda^{\text{over}}/\Lambda^{\text{under}}$ as the relative cost of over-precautions. For a given A, the error probabilities are given by $\Pr(L|NL) = \tfrac{1}{2}(1 - A)^2$ and $\Pr(NL|L) = \tfrac{1}{2}A^2$. The corresponding loss of social welfare (relative to the first-best) is

$$\Lambda(A) = \tfrac{1}{2}\Lambda^{\text{under}}[A^2 + \lambda(1 - A)^2]. \tag{3}$$

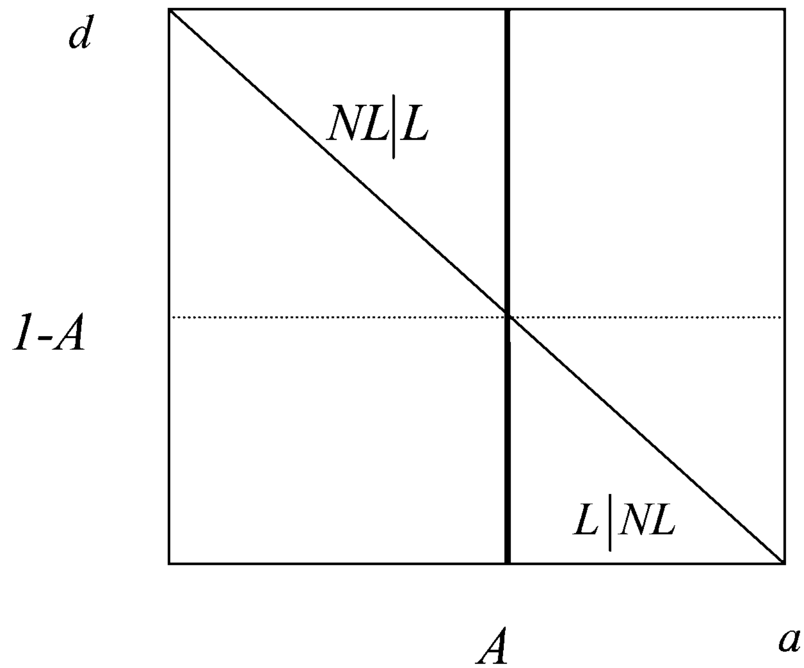

FIG. 2

If A is the initial precedent, social losses are $\Lambda(A)$—an average of the costs of over- and under-precautions under the error probabilities that A introduces. The higher A is (the more the initial rule favors O), the larger the loss from under-precautions but the smaller the loss from over-precautions.

Figure 3 depicts the two-dimensional legal rule with thresholds A, D_0, and D_1. The owner O is overpunished in region $L|NL$, with area $\Pr(L|NL) = \frac{1}{2}[(1 - D_0)^2 + (1 - A - D_1)^2]$, and underpunished in region $NL|L$, with area $\Pr(NL|L) = \frac{1}{2}[(A + D_0 - 1)^2 + D_1^2]$.

The social loss from the use of the two-dimensional legal rule is given by

$$\Lambda(A, D_0, D_1) = \Lambda^{\text{under}}\tfrac{1}{2}\{[(A + D_0 - 1)^2 + D_1^2] + \lambda[(1 - D_0)^2 + (1 - A - D_1)^2]\}. \quad (4)$$

By minimizing (3) with respect to A and (4) with respect to A, D_0, and D_1, we find the optimal one- and two-dimensional threshold rules, which are the normative benchmarks for our analysis.

PROPOSITION 1. (i) The optimal one-dimensional legal rule is given

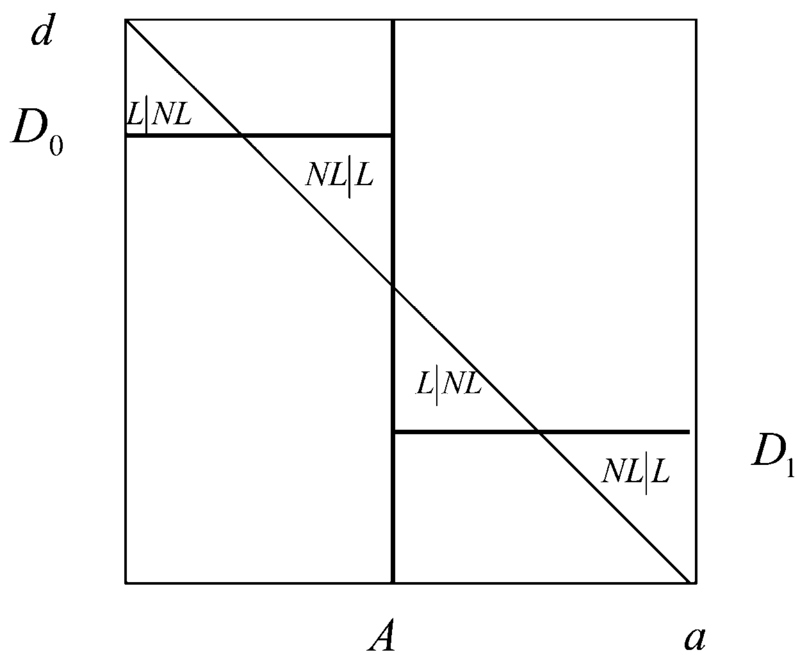

FIG. 3

by $A_L = \lambda/(1 + \lambda)$. (ii) The optimal two-dimensional legal rule is given by $A_F = \frac{1}{2}$, $D_{0,F} = (1 + A_L)/2$, and $D_{1,F} = A_L/2$.

The optimal one-dimensional rule responds to social costs. The higher the relative cost of over-precautions λ, the more lenient the optimal rule (the higher A_L). This is also true for the optimal two-dimensional rule. As seen in figure 3, if the cost of over-precautions is higher, then $D_{0,F}$ and $D_{1,F}$ should be raised so as to reduce the size of $L|NL$, the region in which O is mistakenly held liable. In addition, in the optimal two-dimensional rule, $A_F = \frac{1}{2}$ maximizes the precision benefit of introducing population density into the law. For extreme A_F (one or zero), the added dimension d is worthless: a single threshold on d (D_0 or D_1) describes liability over the entire (a, d) space, just as in a one-dimensional rule.

In our model, the efficiency of a rule depends on two factors: its overall imprecision $\Pr(NL|L) + \Pr(L|NL)$ and the ratio of different errors $\Pr(NL|L)/\Pr(L|NL)$. The optimal initial precedent and the optimal two-dimensional rule fare equally well in terms of this second factor (i.e., they induce the same $\Pr(NL|L)/\Pr(L|NL)$). Yet, by including d in adjudication, the optimal two-dimensional rule is more precise and thus more efficient.

With the results of this section in mind, we can move on to study judicial lawmaking. We ask when and how judge-made law evolves over time and evaluate the efficiency of legal evolution. By efficiency we mean ex ante efficiency, before judge types are revealed.

IV. Distinguishing

Like social welfare, the utility of a judge settling a dispute between O and V depends on the precision of the rule and on the ratio of different mistakes. We assume that a judge's objective diverges from efficiency because of his bias, which reflects his preference for V or O and induces him to sacrifice efficiency for a pattern of mistakes more favorable to the preferred party. Specifically, we assume that the utility of judge j is given by

$$U_j = -\beta_{V,j}\Pr(NL|L) - \beta_{O,j}\Pr(L|NL). \tag{5}$$

Judges dislike making mistakes, but they do not dislike the two types of mistakes equally. The terms $\beta_{O,j}$ and $\beta_{V,j}$ ($\beta_{V,j}$, $\beta_{O,j} \geq 0$) capture the preference of judge j for O and V, respectively: the larger $\beta_{O,j}$ is, the more he is eager to hold O not liable; the larger $\beta_{V,j}$ is, the more he is willing to hold O liable.

Under the assumed utility function, judges are unhappy with any error they make (albeit differentially for different errors). This judicial aversion to making mistakes leads to judicial self-restraint that is crucial for our results: even a judge heavily biased against dog owners would not introduce the most anti-owner liability rule available if this rule leads to mistakes he can avoid, including mistakes favoring bite victims. Such preferences allow us to emphasize—in line with the legal realists—that judicial bias is more problematic in the presence of uncertainty, when judges trade off different errors. We do not model the more extreme kind of favoritism in which the judge rules against dog owners who he knows for sure should not be efficiently held liable.

In our specification of judicial preferences, a judge's utility depends on the expected outcome arising from the application of a given rule, not from the resolution of a particular case. Such a judge would consider replacing a legal rule he dislikes even if the outcome of the specific case before him is the same under the new rule. A judge cares about having a rule in place that meets his idea of justice rather than about delivering a desired outcome in a specific dispute before him. This assumption is particularly appropriate for appellate judges, who establish legal rules.

There is a measure one of judges, who can be of three types: share γ of judges are *unbiased*, with bias $\beta_{O,j}/\beta_{V,j} = \lambda$ reflecting social welfare; the rest are equally divided among *pro-O* judges, with bias $\beta_{O,j}/\beta_{V,j} =$

$\lambda\pi$, and *pro-V* judges, with bias $\beta_{O,j}/\beta_{V,j} = \lambda/\pi$. Parameter π ($\pi \geq 1$) measures the polarization of judges' preferences: with a higher π, the preferences of pro-O and pro-V judges are more extreme (there is more disagreement among them). We assume that all judges have the same preference intensity and normalize it to one ($\beta_{V,j} + \beta_{O,j} = 1$ for all j).

A. *The Initial Precedent*

The first judge adjudicating a dispute between O and V establishes the initial precedent. We have assumed that, in this dispute, the issue of location never arises (and the judge cannot entertain legal issues that do not arise in the dispute). Suppose that this initial dispute comes up before judge i, with preferences $\beta_{V,i}$ and $\beta_{O,i}$. This judge then selects a threshold A_i to maximize[6]

$$-\tfrac{1}{2}\beta_{V,i}[A^2 + \beta_i(1 - A)^2], \tag{6}$$

where $\beta_i = \beta_{O,i}/\beta_{V,i}$ measures the pro-O bias of this judge. Judge i then sets

$$A_i = \frac{\beta_i}{1 + \beta_i}. \tag{7}$$

The subscript indicates that A_i is the initial precedent set with pro-O bias β_i. The result is intuitive: the more pro-O the judge is, the more lenient he is (the higher A_i is). The precedent A_i coincides with the efficient one-dimensional rule A_L only if $\beta_i = \lambda = \Lambda^{\text{over}}/\Lambda^{\text{under}}$, that is, if the judge's bias toward O reflects the relative social cost of overprecautions. If the case ends up in front of a pro-O judge ($\beta_i > \lambda$), too many aggressive dogs roam and bite with impunity; if instead the case ends up in front of a pro-V judge ($\beta_i < \lambda$), too many peaceful dogs are put on a leash. Depending on β_i, the initial precedent may turn out to be severely inefficient.

B. *The New Precedent*

Suppose that after some time a judge j has an opportunity to distinguish the initial precedent A_i by introducing d into the legal rule. Judge j's utility from setting thresholds $D_{0,j}$ and $D_{1,j}$ is

$$-\tfrac{1}{2}\beta_{V,j}\{[(A_i + D_{0,j} - 1)^2 + D_{1,j}^2] + \beta_j[(1 - D_{0,j})^2 + (1 - A_i - D_{1,j})^2]\}. \tag{8}$$

The first term of the expression represents the cost for judge j of mis-

[6] The judge is assumed to act myopically and thus does not act strategically with respect to future judges. The same qualitative results hold when judges are forward-looking.

takenly holding O not liable (i.e., ruling against V), and the second term is the cost for judge j of erroneously holding O liable. Let $A_j = \beta_j/(1+\beta_j)$ be the ideal threshold on the dog's aggressiveness that would be chosen by judge j if he were setting the initial precedent. From first-order conditions, we obtain

$$D_{0,j}(A_i) = 1 - (1 - A_j)A_i \tag{9}$$

and

$$D_{1,j}(A_i) = A_j(1 - A_i). \tag{10}$$

These reaction functions tell us that distinguishing exhibits path dependence: because of *stare decisis*, the way judge j introduces $D_{0,j}$ and $D_{1,j}$ into the law depends on the initial precedent A_i. To gauge the impact of such distinguishing, it is helpful to look at the probabilities of different errors after legal change has occurred:

$$\Pr(L|NL) = \tfrac{1}{2}(1 - A_j)^2[A_i^2 + (1 - A_i)^2] \tag{11}$$

and

$$\Pr(NL|L) = \tfrac{1}{2}A_j^2[A_i^2 + (1 - A_i)^2]. \tag{12}$$

As discussed in Section III, the efficiency of a legal rule depends on the ratio of the two errors $\Pr(NL|L)/\Pr(L|NL)$ and on the overall imprecision $\Pr(NL|L) + \Pr(L|NL)$ it induces. Expressions (11) and (12) show that after distinguishing, $\Pr(NL|L)/\Pr(L|NL) = A_j^2/(1 - A_j)^2$: the ratio between errors is fully determined, through A_j, by the *desired* bias of the second judge! When judge j introduces d into adjudication, he discretionally sets $D_{0,j}$ and $D_{1,j}$ so as to favor the party he prefers. As a result, there is no presumption that the final configuration of the law is less biased than the initial precedent. Owing to the very discretion embodied in distinguishing cases, legal change cannot eliminate this first effect of judicial bias: it cannot correct the ratio of different errors. In this sense, the eccentricities of judges do *not* balance one another, and legal evolution does not reduce the ability of biased judges to distort the ratio of errors away from efficiency.

On the other hand, the imprecision of the legal rule shows the potential of distinguishing:

$$\Pr(NL|L) + \Pr(L|NL) = \tfrac{1}{2}[A_j^2 + (1 - A_j)^2][A_i^2 + (1 - A_i)^2]. \tag{13}$$

Because $A_j^2 + (1 - A_j)^2 \leq 1$, distinguishing (weakly) *reduces* the imprecision of the law, which is equal to $\frac{1}{2}[A_i^2 + (1 - A_i)^2]$ under the initial precedent. This has two implications. First, even if judges are biased, legal evolution can beneficially increase the precision of the law. Notice, however, that expression (13) indicates that even from the standpoint of the law's precision, judicial bias is costly since it increases the *overall* likelihood of judicial error. The bias of the first judge reduces the precision of the initial precedent; that of the second judge reduces the precision benefit of distinguishing. For example, when $A_j = 1$, the precision benefit of distinguishing is nil.

Second, and in contrast to the finding on the ratio of errors, the *path dependence* resulting from sequential decision making now matters: the final legal rule is more precise the greater the precision of the initial precedent A_i. Expression (13) shows that the initial precedent dampens the impact of the second judge's bias on the precision of the law. Although a very pro-O judge may wish to introduce location just to excuse dog owners, he does not want to totally discard the information embodied in the initial legal rule. As a result, the waste of information associated with his exercise of discretion is limited. To see why this is the case, suppose that judge i sets legal rule $A_i = \frac{1}{2}$, whose imprecision is $\frac{1}{4}$. Then, a very pro-O judge j still *can* set $D_{0,j} = D_{1,j} = 1$, which would make imprecision jump to $\frac{1}{2}$, but he does not want to. The reason is that he can set $D_{0,j} = 1$ and $D_{1,j} = \frac{1}{2}$ and in this way avoid the error of excusing the owners of vicious dogs unleashed on a playground. He still keeps the area of false liability down to zero, but because he does not like making *any* errors, his decision is more efficient. In this way, the initial precedent helps constrain the impact of the bias of the second judge on the precision of the law. This discussion also shows that our assumption about judicial preferences actually matters; if judge j cared only about favoring dog owners without regard for making errors, he would set $D_{0,j} = D_{1,j} = 1$ regardless of what judge i did before him.

The ability of the initial precedent to soften the impact of the bias of the second judge depends on the bias of the first judge, since the moderation of the first judge i entails the relative moderation of judge j. If the first judge was extremist (i.e., if $A_i = 0$ or 1), then the second judge behaves *as though* no precedent is in place. For example, in light of precedent $A_i = 0$, a very pro-O second judge sets $D_{0,j} = D_{1,j} = 1$. Since judge j cares only about not erroneously holding dog owners liable, his introduction of d fully eliminates dog owners' liability. When judge i is so extreme, judge j is both able *and willing* to move from the regime of strict liability to the regime of virtually no liability by distinguishing the case on the basis of location. To summarize, we formally define extremism as the distance of a judge's preferred threshold A from $\frac{1}{2}$ and find that the following proposition holds.

PROPOSITION 2. Distinguishing increases the law's precision, but less so the more extreme the second judge is. The initial precedent softens the adverse impact of the second judge's extremism on the law's precision, and the more so the less extremist the first judge is.

C. Welfare Effects of a Precedent Change

After judge j distinguishes the initial precedent A_i, social losses are

$$\Lambda(A_i, D_{0,j}, D_{1,j}) = [A_i^2 + (1 - A_i)^2]\Lambda(A_j). \tag{14}$$

The term $\Lambda(A_j)$ stands for the social loss under the hypothetical assumption that the initial rule is chosen by judge j and captures the idea that legal change allows judges to regain their discretion. The term $A_i^2 + (1 - A_i)^2$ captures the precision benefit of distinguishing. If the initial threshold A_i were not binding in virtue of *stare decisis*, the social loss would be entirely determined by the preferences of judge j, as reflected in the hypothetical A_j.

Because distinguishing allows very biased judges to regain discretion, legal change is not *always* good. To see why, notice that under the initial precedent, social losses are equal to $\Lambda(A_i)$. When the bias of the second judge is more costly to society than the bias of the first judge (if $\Lambda(A_j) > \Lambda(A_i)$), legal change *reduces* welfare when the precision benefit of distinguishing is small. To illustrate, consider the extreme case in which judge i is infinitely pro-V and judge j is infinitely pro-O. The precision benefit of distinguishing is now absent, and legal change effectively allows the second judge to replace the initial rule of strict liability ($A_i = 0$) with his preferred rule of no liability ($A_j = 1$). If underprecautions are socially costlier than over-precautions ($\lambda < 1$), such legal change is harmful because it enables the pro-O judge to excuse careless owners of very aggressive dogs. More broadly, legal change is most likely to be detrimental when both the judge setting the initial precedent and the one distinguishing it are extremists and when the latter judge's extremism is more detrimental to social welfare than that of the former.

V. Judge-Made Law in the Long Run

Although judicial activism may make matters worse, to evaluate legal change overall we need to (*a*) study when and how it occurs and (*b*) average over all the possible paths of the law.

By comparing the utility judge j derives from retaining A_i completely (expression [6]) with the utility he obtains by introducing his preferred thresholds $D_{0j}(A_i)$ and $D_{1j}(A_i)$ into the law (expression [8]), we find that judge j distinguishes A_i when

$$(A_i - A_j)^2 + 2A_j(1 - A_j)A_i(1 - A_i) \geq 2k. \tag{15}$$

Intuitively, the smaller the cost k of changing the law, the higher the chance that legal change occurs. More important, the left-hand side of (15) illustrates that a judge changing the law obtains two benefits. First, legal change allows him to replace the precedent's bias with his own preferred bias. This benefit is captured by the term $(A_j - A_i)^2$ and suggests that greater disagreement between judges j and i leads the former to distinguish A_i more often. Second, there is an informational gain from distinguishing, namely, $2A_j(1 - A_j)A_i(1 - A_i)$. This gain is stronger for moderate judges ($A_j = \frac{1}{2}$) who care most about the precision of the law, but is small or absent if the first judge was an extremist (i.e., if $A_i = 0$ or 1). This gain may induce a judge to distinguish even a precedent set by a predecessor with identical preferences.

The idea that the extent to which a judge disagrees with the existing precedent shapes his incentive to change the law suggests that the polarization of judges' views is a key determinant of the long-run configuration of judge-made law. Indeed, the more polarized judges' views are (i.e., the larger π is), the greater the likelihood that a judge inherits a precedent he strongly disagrees with. The case of $\lambda = 1$ illustrates this intuition.

Proposition 3. If $\lambda = 1$, there exist two polarization levels $\tilde{\pi}_1$ and $\tilde{\pi}_2$ ($\tilde{\pi}_1 \leq \tilde{\pi}_2$) such that (i) if $\pi < \tilde{\pi}_1$, the initial precedent is never distinguished; (ii) if $\pi \in [\tilde{\pi}_1, \tilde{\pi}_2)$, only pro-$O$ and pro-V initial precedents are distinguished; and (iii) if $\pi \geq \tilde{\pi}_2$, all initial precedents are distinguished.

Polarization of judicial preferences determines whether d is eventually introduced into the law. If π is low, judge-made law immediately converges to the one-dimensional threshold on aggressiveness set by the first judge. At intermediate levels of polarization, pro-O and pro-V judges distinguish each other's precedents. The law converges to a two-dimensional legal rule unless an unbiased judge sets the initial threshold A, which then becomes permanent. At high levels of polarization, even unbiased precedents are distinguished.[7]

Our model yields the empirical prediction that legal change occurs more often the higher the dispersion of judicial preferences. Across areas of law, those with greater dispersion of judicial views (perhaps because they are more political) would see more legal change. These results are broadly consistent with the main finding of the political science literature, namely, that the Supreme Court distinguishes precedents incompatible with its political orientation and that extreme

[7] Trivially, the law always converges if the number of empirical dimensions defining a transaction is finite and—which is essentially the same—if transactions do not continuously change over time. This result hinges on the assumption that judges cannot introduce irrelevant dimensions into the law. In other words, the *materiality* of a dimension is a physical characteristic that even the most biased judges cannot subvert. As a result, a *stare decisis* doctrine constraining judges to distinguish the current precedent by using only material dimensions is successful in assuring convergence.

Because of its very informational benefit, distinguishing allows precedents to adapt to changing circumstances. Suppose that the previously immaterial dimension d becomes material to a transaction. Our model of distinguishing shows, for a given initial precedent A, *if* and *how* judges adapt the law to these changed circumstances.

judges are more likely to vote against the precedent than the centrist ones (Brenner and Stier 1996). But proposition 3 delivers another novel empirical prediction, namely, that legal rules are more complex (include more empirical dimensions) when judicial views are more dispersed.

The only novel feature arising when $\lambda < 1$ is that the unbiased initial precedents may no longer be the hardest to distinguish. For example, at intermediate levels of polarization, pro-O judges may prefer to distinguish an unbiased precedent to a pro-V one because the informational benefit of distinguishing the latter is too small when λ is very low. Still, the main thrust of proposition 3 is maintained: distinguishing increases with the polarization of judicial preferences.

VI. The Efficiency of Legal Evolution under Distinguishing

Having studied when and how legal evolution occurs, we now evaluate the efficiency properties of judge-made law. We start by asking when judge-made law converges to the optimal two-dimensional legal rule.

PROPOSITION 4. Judge-made law converges to efficiency if and only if all judges are unbiased, k is sufficiently small, and $\lambda = 1$.

Because biased judges distort the ratio between different errors away from efficiency, a population of fully unbiased judges is necessary for judge-made law to converge to the efficient two-dimensional rule. But two other conditions must be met. First, judges must be interventionist enough to introduce d into the law; otherwise a two-dimensional rule cannot even be reached. That is, k must be sufficiently low that unbiased judges distinguish the initial precedent A_L.

Second, and more important, even with two unbiased judges ruling sequentially, it must be that $\lambda = 1$; that is, the relative social cost of different errors should be one because of the law's path dependence, which introduces an externality across judges. When $\lambda \neq 1$, the initial precedent is set at A_L and not at $\frac{1}{2}$ (as in the optimal legal rule) because the first judge disregards the adverse impact of his choice on the long-run precision of the law. By virtue of *stare decisis*, the initial precedent is then respected, thus keeping the long-run law from full efficiency. Forward-looking behavior on this judge's part does not remove this inefficiency unless he is infinitely patient.

The key role of efficiency-seeking judges in the convergence of common law to efficiency is recognized by Posner (2003), although he does not explain just how stringent the conditions for full efficiency are. Nor does Posner discuss how path dependence of judge-made legal rules may prevent them from attaining full efficiency, even if all judges are unbiased.

How does social welfare depend on polarization? We find that the following proposition holds.

PROPOSITION 5. For every k, there exists a $\pi^*(k) \geq 1$ such that social welfare decreases with π if $\pi < \pi^*(k)$ or $\pi > \pi^*(k)$, but social welfare is maximized at $\pi = \pi^*(k)$ and $\pi^*(k) > 1$ for some k.

Proposition 5 points to a cost but also a benefit of judicial polarization. The cost of greater polarization of judicial views to the long-run efficiency of judge-made law comes from two effects. First, when judges are more biased, legal rules are less precise. Second, when judges are more biased, legal rules induce a ratio between different errors that is further away from efficiency. Both of these costs of polarization hold for both one- and two-dimensional legal rules.

However, proposition 5 also highlights a benefit of judicial disagreement because the efficiency of judge-made legal rules often jumps up at polarization level $\pi^*(k) > 1$. The intuition for this result hinges on our finding in proposition 4 that judges are more likely to distinguish a precedent when the disagreement among them is greater. If polarization is low, judges lack an important incentive to change the law, namely, the benefit of replacing the bias of their predecessor with their own. The law then does not evolve from its initial one-dimensional configuration. If judicial views are sufficiently polarized, judges find it worthwhile to distinguish the initial precedent, allowing the introduction of d into the law to improve its "precision." In a sense, polarization is the price to pay for judge-made law to adapt and become ever more precise. Of course, as proposition 5 shows, the precision benefit of distinguishing becomes smaller as polarization increases. Our model thus predicts an inverted U-shaped relationship between polarization and the efficiency of judge-made law.

Although proposition 5 indicates that some judicial polarization can improve the long-run efficiency of judge-made law, it does not tell whether legal change is generally good. For example, at high levels of polarization the introduction of d into the law may reduce welfare. When π is large, the precision benefit of distinguishing is small and the adverse impact of judicial bias on the ratio between different errors may undermine the desirability of legal change.

This observation leads to a key result of our paper. In line with Cardozo's intuition, we find that if we consider all possible paths of the law, then distinguishing is *on average* beneficial in many circumstances, even when judicial polarization is very high.

PROPOSITION 6 (Cardozo theorem). There exists a $\bar{k} > 0$ such that, for $k \leq \bar{k}$, distinguishing of precedents is on average beneficial. As a result, at every π, $k = 0$ is socially preferred to $k = \infty$.

Irrespective of judicial polarization π, judicial activism (i.e., a low k) renders legal change desirable on average. To see why judicial activism (and thus legal change) is beneficial even in the presence of biased judges, compare the ex ante social loss attained at $k = 0$ with that attained at $k = \infty$. If $k = 0$, judges are so activist that they always distinguish any initial precedent A_i, thereby leading to a second-period expected loss of $[A_i^2 + (1 - A_i)^2]E_j[\Lambda(A_j)]$. Averaging such losses across all paths of legal change (i.e., across initial precedents), we find an ex ante loss of

$$E_i[A_i^2 + (1 - A_i)^2]E_j[\Lambda(A_j)]. \tag{16}$$

If instead $k = \infty$, judges are passive and never distinguish the initial precedent, thereby leading to an ex ante loss of $E_i[\Lambda(A_i)]$. Since $E_i[A_i^2 + (1 - A_i)^2] \leq 1$, legal change is beneficial at every level of judicial polarization and $k = 0$ is socially preferred to $k = \infty$. The reason is that the introduction of d into the law brings an informational benefit that *on average* overpowers the cost of bias.

This result vindicates Cardozo's intuition for the presence of a "technological" force driving the evolution of precedent toward efficiency despite the vagaries of individual judges. When judges embrace legal change (as in the case of $k = 0$), their biases "wash out" on average, and the net gain for the law comes from the more accurate information (greater number of empirical dimensions) embodied in legal rules. At high levels of polarization, legal change may be detrimental along some paths, such as when a pro-O judge distinguishes away a pro-V precedent, thereby imposing on society a worse scenario of under-precautions. However, when judges embrace legal change, this event is just as likely as the one in which a pro-V judge distinguishes a pro-O precedent, thereby sparing society the cost of under-precautions for the lesser one of over-precautions. From an ex ante standpoint, the influence of bias along these two paths cancels out, and what remains on net is the greater precision of the law, which now uses two material dimensions rather than one. This creates a *tendency* for the law to become more efficient over time.[8]

Propositions 5 and 6 suggest that the evolution of common law would produce most socially efficient results in the areas of law in which there is room for change and updating, but in which the disagreement among the judges is not extreme. The relatively apolitical yet still changing areas of law, such as contract and corporate law, are the likely candidates for relatively efficient outcomes resulting from the decentralized evolution of judge-made law. In the extremely political areas of law, in

[8] Llewellyn (1960) and Stone (1985) argue that to implement their bias, judges sometimes distinguish precedents by using irrelevant dimensions. In our model, even if judges are allowed to introduce irrelevant dimensions into the law, the informational benefit of distinguishing is still present because judges prefer to distinguish using material rather than irrelevant dimensions. Both dimensions allow judges to bias the law optimally, but material ones yield the extra benefit of greater precision. Hence, the informational benefit remains a feature of distinguishing, at least until material dimensions are exhausted. This implies that some polarization in judicial preferences is still desirable.

contrast, the likelihood that the law gets stuck on a wrong trajectory is higher.

VII. Conclusion

When and how does the evolution of judge-made law take place? When does such evolution improve the law on average? Does it lead to convergence to efficient legal rules? We addressed these questions in a legal-realist model in which deciding judges face opportunities to distinguish the precedent from the case before them, but may be both biased and averse to changing the law.

We found that the conditions for ultimate efficiency of judge-made law are implausibly stringent. Moreover, a legal rule governing a particular situation may start off in a very inefficient place and, because of path dependence in judge-made law, remain highly inefficient despite future refinements. Yet even though full efficiency is hard to attain and some legal rules remain bad, there is a presumption that legal change raises welfare as it improves the informational quality of judicial decision making, at least when the cost of changing the law is low. Although common-law judges do not seek to improve the efficiency of legal rules but rather pursue their own agendas, their independent decentralized decisions have a benign side effect. The law better adapts to the underlying transactions (and to new circumstances) when activist judges distinguish cases. This basic finding on the evolution of common law is very supportive of the ideas of free-market philosophers such as Hayek and Leoni, of legal realists such as Llewellyn and Cardozo, and of law and economics scholars following Posner.

The benefit of legal change implies that, compared to the case of no judicial disagreement, some judicial disagreement is beneficial. Indeed, we found that judicial disagreement is an important factor fostering legal change and volatility in the law. The model predicts a faster pace of legal change in the more politically (or otherwise) divisive areas of law. On average with such change, legal rules become more efficient by becoming more complex, as measured by the number of tests or considerations entering into a legal decision. These predictions on the volatility and complexity of different areas of law can be tested using data on appellate court decisions.

Our model is a first step in the analysis of judge-made law and omits some important aspects of legal evolution. First, we consider legal change through distinguishing and disregard the possibility that courts occasionally simply overrule precedents. We examine the case of overruling in Gennaioli and Shleifer (2007), but the basic point is simple. Since overruling, unlike distinguishing, does not bring new material dimensions into the law, it leads to the volatility of legal rules without a tendency to improve the law over time. With overruling, there is no benefit of legal evolution.

Second, we have ignored several institutional features of appellate review that might affect our results. Unlike the previous researchers, we neglect the selection of disputes for judicial resolution rather than settlement. In addition, appellate judges sit on panels and make decisions collectively. These factors might be a force for moderation, although—precisely by inducing moderation—they might also slow down the pace of legal change, which is not necessarily efficient.

As a final note, we emphasize that ours is a theoretical analysis of the proposition that the evolution of common law is beneficial. We have tried to develop several testable implications of our analysis that suggest the areas of the law in which this benign conclusion is more likely to hold, in particular connecting efficiency to the volatility and complexity of legal rules. These predictions may be easier to verify empirically than the broad propositions about the efficiency of common law.

Appendix

Proofs

Proof of Proposition 1

The optimal one-dimensional threshold rule A_L is defined as

$$A_L = \arg\min_{A \in [0,1]} A^2 + \lambda(1 - A)^2.$$

The objective function is convex, and $A_L = \lambda/(1 + \lambda)$ $(A_L \in [0, 1])$ is found by solving the first-order conditions $(A_L - \lambda[1 - A_L] = 0)$. The optimal two-dimensional threshold rule (A_F, D_0, D_1) is defined as

$$(A_F, D_0, D_1) = \arg \min_{A, D'_0, D'_1 \in [0,1]^3} [(A + D'_0 - 1)^2 + (D'_1)^2] + \lambda[(1 - D'_0)^2 + (1 - A - D'_1)^2].$$

Again, the above objective function is convex in (A, D'_0, D'_1) (its Hessian is positive definite). Thus solving the first-order conditions for (A_F, D_0, D_1), namely,

$$\frac{\partial \Lambda}{\partial A} = (A_F + D_0 - 1) - \lambda(1 - A_F - D_1) = 0,$$

$$\frac{\partial \Lambda}{\partial Q_0} = (A_F + D_0 - 1) - \lambda(1 - D_0) = 0,$$

$$\frac{\partial \Lambda}{\partial Q_1} = D_1 - \lambda(1 - A_F - D_1) = 0,$$

yields $A_F = \frac{1}{2}$, $D_0 = (1 + A_L)/2$, and $D_1 = A_L/2$. Notice that $(A_F, D_0, D_1) \in [0, 1]^3$. QED

Proof of Proposition 2

When the initial precedent A_i is distinguished by judge j, the law's imprecision is

$$\Pr(NL|L) + \Pr(L|NL) = \frac{1}{2}[A_j^2 + (1 - A_j)^2][A_i^2 + (1 - A_i)^2].$$

In contrast, the imprecision of the initial precedent is $\frac{1}{2}[A_i^2 + (1 - A_i)^2]$. The precision benefit of distinguishing (i.e., the imprecision of the initial precedent minus that of the distinguished precedent) is clearly smaller the further A_j is from $\frac{1}{2}$ (i.e., the less moderate judge j is). In addition, the impact of judge j's extremism on the law's imprecision, namely,

$$\frac{\partial[\Pr(NL|L) + \Pr(L|NL)]}{\partial|A_j - \frac{1}{2}|},$$

is larger the less moderate judge i is, namely,

$$\frac{\partial^2[\Pr(NL|L) + \Pr(L|NL)]}{\partial|A_j - \frac{1}{2}|\partial|A_i - \frac{1}{2}|} > 0.$$

QED

Proof of Proposition 3

Call O a generic pro-O judge, V a generic pro-V judge, and U a generic unbiased judge. Then judge j distinguishes precedent A_i set by judge i (where $i, j = O, V, U$) when

$$h_{j,i}(\pi) \equiv \frac{\beta_i^2 + \beta_j^2}{(1 + \beta_i)^2(1 + \beta_j)^2} \geq 2k.$$

Given symmetry, we must consider only the following cases.

a. If pro-O ($\beta_j = \pi\lambda$) judge follows pro-O ($\beta_i = \pi\lambda$) judge, then

$$h_{O,O}(\pi) \equiv 2\frac{\pi^2\lambda^2}{(1 + \pi\lambda)^4}.$$

b. If pro-V ($\beta_j = \lambda/\pi$) judge follows pro-V ($\beta_i = \lambda/\pi$) judge, then

$$h_{V,V}(\pi) \equiv 2\frac{\lambda^2\pi^2}{(\pi + \lambda)^4}.$$

c. If unbiased ($\beta_j = \lambda$) judge follows unbiased ($\beta_i = \lambda$) judge, then

$$h_{U,U}(\pi) \equiv 2\frac{\lambda^2}{(1 + \lambda)^4}.$$

d. If pro-O ($\beta_j = \pi\lambda$) judge follows pro-V ($\beta_i = \lambda/\pi$) judge, then

$$h_{O,V}(\pi) \equiv h_{V,O}(\pi) \equiv \frac{\lambda^2(\pi^4 + 1)}{(1 + \lambda\pi)^2(\pi + \lambda)^2}.$$

e. If pro-*O* ($\beta_j = \pi\lambda$) judge follows unbiased ($\beta_i = \lambda$) judge, then

$$h_{O,U}(\pi) \equiv h_{U,O}(\pi) \equiv \frac{\lambda^2(\pi^2+1)}{(1+\pi\lambda)^2(1+\lambda)^2}.$$

f. If pro-*V* ($\beta_j = \lambda/\pi$) judge follows unbiased ($\beta_i = \lambda$) judge, then

$$h_{V,U}(\pi) \equiv h_{U,V}(\pi) \equiv \frac{\lambda^2(1+\pi^2)}{(\pi+\lambda)^2(1+\lambda)^2}.$$

If $j \neq i$, $h_{j,i}(\pi)$ is increasing, $h_{j,i}(1) = h_{U,U}$, $h_{O,V}(\infty) = 1$, $h_{O,U}(\infty) = 1/(1+\lambda)^2$, and $h_{V,U}(\infty) = \lambda^2/(1+\lambda)^2$. The function $h_{j,j}(\pi)$ is decreasing, $h_{j,j}(1) = 2\lambda^2/(1+\lambda)^2$, and $h_{O,O}(\infty) = h_{V,V}(\infty) = 0$. Some rankings in the $h_{j,i}(\pi)$ are $h_{V,U}(\pi) = \min_{j\neq i} h_{j,i}(\pi)$, $h_{V,V}(\pi) = \min_j h_{i,i}(\pi)$, and $h_{O,U}(\pi) \geq \max_i h_{i,i}(\pi)$. Disagreement tends to be a stronger incentive for distinguishing than information ($\max_{j\neq i} h_{i,j}(\pi) \geq \max_i h_{i,i}(\pi)$). If $\lambda = 1$, we have $h_{O,V} \geq h_{O,U} \geq h_{V,U} \geq h_{U,U} \geq h_{O,O} \geq h_{V,V}$. Call $\pi_{j,i} \in [1, +\infty) \cup \{+\infty\}$ the level of π such that, for $\pi \geq \pi_{j,i}$, j overrules i (and vice versa). In this proof, we allow $\pi_{j,i} = +\infty$ even though it is always the case that $\pi_{j,i} < +\infty$ (and π can tend to $+\infty$ only in the limit). This allows us to set $\pi_{i,j} = +\infty$ when i and j never distinguish each other at every π. This way, we can accommodate in proposition 3 also the special case in which $k > \frac{1}{2}$. Because $\lim_{\pi\to\infty} h_{O,V}(\pi) = 1$, for $k > \frac{1}{2}$, there is no π at which the initial precedent is distinguished. In this case, we set $\overline{\pi}_{O,V} = \tilde{\pi}_1 = \overline{\pi}_{O,U} = \tilde{\pi}_2 = \infty$, and proposition 3 correctly yields the result that the initial precedent sticks forever. At the same time, because $h_{O,U} = h_{V,U} \geq \frac{1}{8}$, for $k \leq \frac{1}{16}$, every initial precedent (pro-*V*, pro-*O*, or unbiased) is distinguished. This case is accommodated by setting $\overline{\pi}_{O,V} = \tilde{\pi}_1 = \overline{\pi}_{O,U} = \tilde{\pi}_2 = 1$. More generally, because of the ranking in the incentive to distinguish, we have that, for every k, $\overline{\pi}_{O,V} \leq \overline{\pi}_{O,U} = \overline{\pi}_{V,U} \leq \overline{\pi}_{U,U} \leq \overline{\pi}_{O,O} \leq \overline{\pi}_{V,V}$. Call $\overline{\pi}_{O,V} = \tilde{\pi}_1$ and $\overline{\pi}_{O,U} = \tilde{\pi}_2$. Then the long-run configuration of precedent behaves as in proposition 3. QED

Proof of Propositions 4 and 5

The symmetry in judicial behavior implies that, unless $\gamma = 1$, the bias of the law is not efficient. With distinguishing, $\gamma = 1$ is not enough for efficiency. If $\gamma = 1$ and $k = 0$, the law converges to $A = \lambda/(1+\lambda)$, $D_0 = 1 - D_1$, and $D_1 = \lambda/(1+\lambda)^2$, which is efficient only if $\lambda = 1$. Finally, judges must introduce d into the law, that is, $k \leq \tilde{k} \equiv \frac{1}{2}h_{U,U}$. To see how polarization can be beneficial, consider the case in which $k > \tilde{k}$. If $k \leq \tilde{k}$, unbiased judges are sufficiently motivated to distinguish even if $\pi = 1$, so polarization here is likely to reduce welfare. When $k > \tilde{k}$, if $\pi = 1$, the law sticks at A_L and d is never introduced. For every k, define $\hat{\pi}(k)$ as the level of polarization such that $\min_{i\neq j} f_{i,j}(\hat{\pi}(k)) = 2k$. In words, $\hat{\pi}(k)$ is the minimal level of polarization at which all judges j distinguish A_i for $i \neq j$. Clearly, as $k \to \tilde{k}$, $\hat{\pi}(k)$ can be made close to one (but greater than one) and the expected social losses at $\hat{\pi}(k)$ can be made arbitrarily close to $[A_L^2 + (1 - A_L)^2]\Lambda(A_L) < \Lambda(A_L)$. Thus, for some $k > \tilde{k}$, there exists a $\pi^*(k) > 1$ such that social losses at $\pi^*(k)$ are smaller than social losses at one. Hence, under distinguishing, some polarization can be strictly beneficial. QED

Proof of Proposition 6

Define $\bar{k} = h_{i,j}(1) = 2\lambda^2/(1+\lambda)^4$. If $k \leq \bar{k}$, judges $j \neq i$ and $j = i = U$ distinguish A_i (they would do it at $\pi = 1$). For $j = i$, $i = O$, V, judges can distinguish or not. If pro-O judges distinguish, welfare may go down if $\Lambda(A_O) \geq E_i(\Lambda(A_i))$, which we assume without loss of generality. Hence, for $k \leq \bar{k}$, legal change is good if it is good when only pro-O judges distinguish their precedent. In this respect, two cases must be considered. If $\Lambda(A_V) \geq E_i(\Lambda(A_i))$ (i.e., the activism of pro-V judges on their precedent reduces welfare), then legal change is good for $k \leq \bar{k}$ because, even if we add to the activism of pro-O judges the harmful activism of pro-V judges on their own precedents, then social losses are exactly the same as under $k = 0$. If $\Lambda(A_V) < E_i(\Lambda(A_i))$, then legal change is good if

$$[\tfrac{1}{2}(1-\gamma)\theta_O + \gamma\theta_U + (1-\gamma)\theta_V - 1]E(\Lambda(A_i)) \leq \tfrac{1}{2}(1-\gamma)^2\theta_V\Lambda(A_V),$$

where $\theta_i = A_i^2 + (1-A_i)^2$. This inequality holds if legal change is good when $\Lambda(A_V) = E_i(\Lambda(A_i))$ and thus a fortiori if $\Lambda(A_V) < E_i(\Lambda(A_i))$. This is true because if $\Lambda_V = E(\Lambda_i)$, social losses are the same as at $k = 0$. As a result, under both circumstances, for $k \leq \bar{k}$, legal change is on average beneficial even if pro-O judges but not pro-V ones distinguish their own precedent. This immediately implies that the same holds in all other cases. As a result, for $k \leq \bar{k}$, legal change is on average beneficial for every π. It immediately follows that at every π, $k = 0$ is socially preferred to $k = \infty$. QED

References

Beck, Thorsten, Asli Demirguc-Kunt, and Ross Levine. 2003. "Law and Finance: Why Does Legal Origin Matter?" *J. Comparative Econ.* 31 (December): 653–75.

———. 2005. "Law and Firms' Access to Finance." *American Law and Econ. Rev.* 7 (Spring): 211–52.

Brenner, Saul, and Harold Spaeth. 1995. *Stare Indecisis: The Alteration of Precedent on the Supreme Court, 1946–1992.* New York: Cambridge Univ. Press.

Brenner, Saul, and Marc Stier. 1996. "Retesting Segal and Spaeth's *Stare Decisis* Model." *American J. Polit. Sci.* 40 (November): 1036–48.

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

Coase, Ronald H. 1960. "The Problem of Social Cost." *J. Law and Econ.* 3 (October): 1–44.

Cooter, Robert, and Lewis Kornhauser. 1980. "Can Litigation Improve the Law without the Help of Judges?" *J. Legal Studies* 9 (January): 139–63.

Cooter, Robert, Lewis Kornhauser, and David Lane. 1979. "Liability Rules, Limited Information, and the Role of Precedent." *Bell J. Econ.* 10 (Spring): 366–73.

Frank, Jerome. 1930. *Law and the Modern Mind.* New York: Brentano's.

Gennaioli, Nicola. 2004. "Contracting in the Shadow of the Law." Manuscript, Harvard Univ.

Gennaioli, Nicola, and Andrei Shleifer. 2007. "Overruling and the Instability of Law." *J. Comparative Econ.* (forthcoming).

George, Tracey F., and Lee Epstein. 1992. "On the Nature of Supreme Court Decision Making." *American Polit. Sci. Rev.* 86 (June): 323–37.

Glaeser, Edward L., and Andrei Shleifer. 2002. "Legal Origins." *Q.J.E.* 117 (November): 1193–1229.

Goodhart, Arthur L. 1930. "Determining the Ratio Decidendi of a Case." *Yale Law J.* 40 (December): 161–83.

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

Hansford, Thomas G., and James F. Spriggs II. 2006. *The Politics of Precedent on the U.S. Supreme Court.* Princeton, NJ: Princeton Univ. Press.

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

Hayek, Friedrich A. von. 1960. *The Constitution of Liberty.* Chicago: Univ. Chicago Press.

———. 1973. *Law, Legislation, and Liberty: A New Statement of the Liberal Principles of Justice and Political Economy.* Vol. 1. *Rules and Order.* Chicago: Univ. Chicago Press.

Holmes, Oliver Wendell. 1897. "The Path of the Law." *Harvard Law Rev.* 10 (March 25): 457–78.

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

Landes, William M., and Richard A. Posner. 1987. *The Economic Structure of Tort Law.* Cambridge, MA: Harvard Univ. Press.

La Porta, Rafael, Florencio Lopez-de-Silanes, Cristian Pop-Eleches, and Andrei Shleifer. 2004. "Judicial Checks and Balances." *J.P.E.* 112 (April): 445–70.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny. 1998. "Law and Finance." *J.P.E.* 106 (December): 1113–55.

Leoni, Bruno. 1961. *Freedom and the Law.* William Volker Fund Series in the Humane Studies. Princeton, NJ: Van Nostrand.

Llewellyn, Karl N. 1951. *The Bramble Bush: On Our Law and Its Study.* New York: Oceana.

———. 1960. *The Common Law Tradition: Deciding Appeals.* Boston: Little, Brown.

Pinello, Daniel R. 1999. "Linking Party to Judicial Ideology in American Courts: A Meta-Analysis." *Justice System J.* 20 (3): 219–54.

Ponzetto, Giacomo, and Patricio Fernandez. 2006. "Case Law vs. Statute Law: An Evolutionary Comparison." Manuscript, Harvard Univ., Dept. Econ.

Posner, Richard A. [1973] 2003. *The Economic Analysis of Law.* 6th ed. Boston: Little, Brown.

———. 2005. "Judicial Behavior and Performance: An Economic Approach." *Florida State Univ. Law Rev.* 32 (Summer): 1259–79.

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

Radin, Max. 1925. "The Theory of Judicial Decision: Or How Judges Think." *American Bar Assoc. J.* 11 (6): 357–62.

Revesz, Richard L. 1997. "Environmental Regulation, Ideology, and the D.C. Circuit." *Virginia Law Rev.* 83 (November): 1717–72.

Rowland, Charles K., and Robert A. Carp. 1996. *Politics and Judgment in Federal District Courts.* Lawrence: Univ. Press Kansas.

Rubin, Paul H. 1977. "Why Is the Common Law Efficient?" *J. Legal Studies* 6 (January): 51–63.

Songer, Donald R., and Stefanie A. Lindquist. 1996. "Not the Whole Story: The Impact of Justices' Values on Supreme Court Decision Making." *American J. Polit. Sci.* 40 (November): 1049–63.

Stone, Julius. 1985. *Precedent and Law: Dynamics of Common Law Growth.* Sydney: Butterworths.

Sunstein, Cass R., David Schkade, and Lisa Michelle Ellman. 2004. “Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation.” *Virginia Law Rev.* 90 (March): 301–54.

Zywicki, Todd J. 2003. “The Rise and Fall of Efficiency in the Common Law: A Supply-Side Analysis.” *Northwestern Univ. Law Rev.* 97 (4): 1551–1633.