Democracy, Judicial Attitudes and Heterogeneity: the Civil Versus Common Law Tradition.

Carmine Guerriero∗
University of Cambridge
November 19, 2008

Abstract
A key issue in the design of a legal system is the degree of centralism of the mechanism aggregating the citizens’ preferences over the level of deterrence. I study a world of culturally heterogeneous groups and I compare two institutions. In one the law is chosen by a legislator who bargains over policies with coalitions of the groups facing fixed bribing costs. In the other the law is selected by appellate judges randomly picked from the population and facing variable costs of overruling the precedent. The former—statute law—outperforms the latter—case law—whenever the legislator rejects bribes because the political process is sufficiently democratic and/or the cultural disagreement is limited. Also, institutions fostering the use of discretion by lower adjudicating courts, rising the overruling costs borne by appellate judges, increase the efficiency of case law.

Data on the history of the legal order of 156 countries confirm these predictions: in countries in which statute law and bright line rules of adjudication were transplanted, reforms toward case law and discretion-enhancing rules are found where the executive is less tightly constrained and cultural heterogeneity is higher. The results remain robust when instruments suggested by evolutionary biology (cross-cultural psychology) are used for the cultural heterogeneity (democracy) proxies. This exercise casts doubts on the supposed primacy of common law documented by the “legal origins” literature.

Keywords: legal origins, legal evolution, culture, democracy, economic development.

JEL classification: K40; Z1; H11; 010; P16.

∗ I am deeply grateful to Ilaria Caggiano, Pierluigi Guerriero, Giuseppe Macario, Kristin MacDonald, Francesco Parisi, Stefano Sacchetto and to seminar participants at Cambridge, at LBS, at UVA (ACLE), at St Andrews, at the 2008 La Pietra-Mondragone workshop, at the 2008 EALE meeting (Haifa), and at the 2008 winter meeting of the Econometric Society (Cambridge) for useful comments. Moreover, I am indebted to Florencio Lopez-de-Silanes and to Romain Wacziarg for the data provided and to the Squire law library staff for the help. Finally, I really don’t know how to thank Toke Aidt and Melvyn Weeks for several precious discussions on earlier drafts.

Carmine Guerriero, Faculty of Economics, University of Cambridge.
Sidgwick Avenue, CB3 9DD, Cambridge, UK.
E-mail: cg372@econ.cam.ac.uk.
© Copyright 2008 by Carmine Guerriero.
“A legal tradition [...] is not a set of rules of law [...] rather it is a set of deeply rooted, historically conditioned attitudes about [...] the proper organization and operation of a legal system. The legal tradition relates the legal system to the culture of which it is a partial expression.” *The Civil Law Tradition* [Merryman, 1969, page 2]

I. Introduction

The role of law in the society and the institutions entrusted with its creation are deeply influenced, in a great part of the world, by either the civil or the common law tradition. While the latter originated in the laws of England and has, then, been transplanted through colonization into England’s ex-colonies, the former has its roots in the Roman law and was exported by Napoleon to continental Europe, and, then, transplanted to Latin America, Africa and Asia. Structurally, the two traditions operate in different ways: while civil law relies on legal codes and bright line adjudication rules, common law recognizes a crucial role to precedents and allows a broader procedural discretion to lower adjudicating courts. Building on these differences and taking advantage of the transplantation experiment, a recent research project has provided evidence according to which French and socialist civil law countries display more corrupt and less efficient governments and courts, less secure investor rights, more extensive market regulation than common law countries do (for a review see La Porta et al., 2008).

Yet, legal scholars have recently documented not only the increasing relevance of judge decisions in civil law legal systems but also the adoption of statutes and bright line rules of adjudication in common law countries (see Zweigert and Kötz, 1998; Hertig, 2004). What is, therefore, the broader set of forces that justify, despite the above mentioned overwhelming evidence, the existence of such dissimilar institutions, and how do these forces interact with each other and the economic environment?
This paper lays out a theoretical framework for thinking about these issues, and explores its empirical implications using data on 156 countries that received their legal tradition externally by a more or less conscious transplantation process (transplanted, therein).

The model merges a recent body of research on endogenous lobbying (Felli and Merlo, 2006) with a lively literature on judicial activism and case law (Gennaioli and Shleifer 2007a, 2007b; Fernandez and Ponzetto 2008a, 2008b). I study a world of social groups separated by their cultural preferences over the prevailing level of deterrence. Under statute (case) law, the law is chosen by a legislator (an appellate judge randomly picked from the population) who bargains over policies with coalitions of the social groups who, in turn, face a fixed bribing cost (faces a variable cost of overruling the precedent).

While the central authority selects certain rules whose bias falls with the quality of the political process, appellate judges constantly innovate at the margin upon established precedents reaching long-run unbiasedness at the cost of everlasting volatility.

If the legislator is sufficiently benevolent or the cultural distance among competing groups is small, bargaining costs discourage bribing and statute law reaches efficiency without volatility. In this case statute law outperforms case law. Yet, a higher cultural polarization raises the legislator willingness to accept bribes, and it does exist a level of democracy below which case law outperforms statutes. The intuition remains true under a wide battery of different hypotheses about the functioning of two institutions. Also, if rules fostering the use of discretion by lower adjudicating courts rise the overruling costs borne by appellate judges, case law becomes more efficient and pure common law systems—where case law and discretion-enhancing rules are used—endogenously arise.

To test the model’s predictions, I have collected data tracking the evolution of the legal system of a cross section of transplanted countries. I record whether statute or case law was in place in the year of independence and in 2000 and four proxies for the level of
discretion in adjudication for the same two points in time. Consistent with the model, in countries in which civil law—statute law and bright line rules—was transplanted, reforms toward common law—case law and flexible adjudication—are more likely the weaker are the democratic institutions—constraints on the executive—and the broader is cultural heterogeneity—the genetic distance between the population who chose the law-making institution and that of the transplanted country and the ethno-linguistic fractionalization in the transplanted country. These results survive if I use instruments suggested by evolutionary biology (cross-cultural psychology) to capture the exogenous variation in heterogeneity (democracy). I also document a positive relation between the likelihood of reforms toward common law and the risk of coercion by strong elites. This last piece of evidence contradicts the intuition proposed by Glaeser and Shleifer (2002) in their analysis of the medieval rise of common (civil) law in England (France).1

The paper most closely related to mine is Fernandez and Ponzetto (2008a). Building on the same attitudinal model, they also compare statute and case law and obtain that the latter has reversion toward the optimal rule properties. Yet, the hypothesis that the coalition of bribing groups is random produces the result that statutes are essentially unpredictable ex ante and essentially suboptimal with respect case law in the long run. Such an assumption is empirically unreasonable (see Wright, 1996); besides, no attempt is made to evaluate both the impact of preferences heterogeneity and the role of discretion by lower adjudicating courts. This, however, is a general drawback of the literature. Indeed, although the comparative merits of statute and case law have been debated since antiquity, the present paper is the first one to link the welfare properties of

1 Glaeser and Shleifer (2002) study the trade-off between vulnerability of judges to private subversion and their insulation from the central power. In the 13th century, England was more peaceful than France and it was optimal for the former (latter) to adopt adjudication by independent juries (state-employed judges).
the two traditions to the cultural and political environment. Accordingly, another major contribution of my analysis is to provide a first empirical test of different theories that justify the adoption of the legal institutions typical of the two traditions as a result of maximizing behaviors. Finally, my results cast more than a doubt on both the supposed primacy of common law stressed by the “legal origins” literature and the lack of relation between growth and legal traditions documented by Acemoglu and Johnson (2005). Researchers interested in clarifying these issues should not only consider altogether the institutions—i.e., law making and adjudication rules—characterizing the two traditions, but should also take care of the endogeneity of these rules to the preferences dispersion and to the quality of the political process, both of which affect performances as well.

The rest of the paper is organized as follows. Section 2 compares the welfare properties of statute and case law as a function of the level of cultural heterogeneity and of the strength of democracy. Section 3 studies discretion in adjudication. Section 4 presents the empirical test and section 5 concludes. Appendix 1 (2) gathers proofs (tables).

II. Statute Law versus Case Law

The basic model follows the Gennaioli and Shleifer’s (2007a, b) analysis of tort law.

II.A Preliminaries

The harmful action.—Consider a society composed by a continuum of citizens whose mass is normalized to one. This society is interested in regulating a harmful action affecting two parties: an offender \( O \) and a victim \( V \). The former can take precautions at

---

2 Proponents of statute law (Hobbes, 2005; Bentham, 1891) stress the certainty of rules; supporters of case law (Hayek, 1960; Cardozo, 1921; Posner, 2007) underline the evolutionary properties of precedents.
a cost $C$. Precautions reduce the probability of an accident from $p_N$ to $p_P$ (which I normalize to zero). The harm suffered by the victim is normalized to unity. Precautions are socially optimal if and only if $p_N - p_P > C$. Finally assume that damages are so high that they induce the offender to take precautions when she is held liable.

Thus, the problem reduces to finding the level of liability conditional on the observable empirical facts. The conditional probability of an accident depends on the attributes $a \in [0,1]$ and $u \in [0,1]$ which are independently and uniformly distributed in the population of cases. Assume that $p_N - p_P = \Delta P > C$ if $a + u \geq 1$ and $p_N - p_P = \Delta P < C$ if $a + u < 1$. These last hypotheses separate the probabilities of different errors from their welfare costs. So precautions are socially optimal and the offender should be held liable if $a + u \geq 1$.

Assume also that only $a$ is observable. Thus, the law-maker can only fix an $A$ such that the offender is held liable if and only if $a \geq A$. I relax this assumption in subsection II.C. Imperfect information implies statistical errors, and the offender is mistakenly held liable if $a \geq A \wedge a < 1 - u$, which happens with probability $\int_a^1 (1 - a) da = (1/2)(1 - A)^2$, and she is mistakenly held not liable if $a < A \wedge a \geq 1 - u$ which happens with probability $\int_0^A da = (1/2)A^2$. Over (under) precaution brings fixed marginal cost $\Delta P = \Lambda - \Delta P$. Over (under) precaution brings fixed marginal cost $\Lambda = \Lambda - \Delta P$.

Define $\lambda = \Delta / \Lambda$ as the relative “technological” cost of over-precaution—the one that a benevolent expert would fix basing her choice on state-of-the-art scientific knowledge.

A world of biases.—The population is equally split in two groups $i \in I = \{L, H\}$ differing in the relative concern $\beta_i = \beta_{0,i} / \beta_{1,i}$ for a false positive. This symmetry can be broken without upsetting the main results that follow. Normalize the preference intensity in such a way that $\beta_{0,i} + \beta_{1,i} = 1, \forall i$. The unconcerned group has a pro-offender bias $\beta_L = \lambda \pi$ and the concerned one a bias $\beta_H = \lambda / \pi$. Accordingly, $\pi \in [1, \infty)$ represents a measure of
cultural heterogeneity: i.e., the extent of disagreement about the perceived danger of the action between citizens belonging to groups with opposite biases. The loss of welfare relative to the first best—i.e., $\Lambda_i^F^F (A) = (1/2)(\Delta p + C), \forall i$—for a citizen of group $i$ is

$$\Lambda_i (A) = (1/2) \left[ \beta_{o,i} (1 - A)^2 + \beta_{v,i} A^2 \right],$$

(1)

which can be fully characterized by group $i$’s favorite threshold $\hat{A}_i = \beta_{o,i} = \Gamma(\beta)$ with $\Gamma(x) = x(1+x)^{-1}$. The unconcerned group prefers the lenient $\hat{A}_i = \lambda \pi (1+\lambda \pi)^{-1} \equiv \bar{A}$ while, the concerned one the stricter $\hat{A}_i = \lambda \pi (1+\lambda \pi)^{-1} \equiv \bar{A}$. Consequently, (1) rewrites as:

$$\Lambda_i (A) = (1/2) \left[ \hat{A}_i (1 - A)^2 + (1 - \hat{A}_i) A^2 \right] = (1/2) \left[ (A - \hat{A}_i)^2 + \hat{A}_i (1 - \hat{A}_i) \right] \times (1/2) \left[ (A - \hat{A}_i)^2 \right].$$

(2)

Thus, under rule $A$, the loss of social welfare (relative to the first best) is

$$\bar{\Lambda}(A) = (1/2) \left[ \Lambda_i (A) + \Lambda_o (A) \right] = (1/2) \left[ \hat{A}_i (1 - A)^2 + (1 - \hat{A}_i) A^2 \right] \times (1/2) \left[ (A - \hat{A}_i)^2 \right],$$

(3)

and the optimal $\bar{A}$ equals the mean of the $\hat{A}$—i.e., $E(\hat{A}) = (1/2)(\bar{A}+\bar{A})$—and is weakly greater (strictly lower) than the technologically efficient $A^* = \Gamma(\lambda)$ for $\lambda \leq 1$ ($\lambda > 1$).

In interpreting the foregoing, several observations should be borne in mind. First, the hypothesis that citizens have pro-offender or pro-victim tastes captures the existence of long lived cultural biases linking the perceived saliency of a particular crime within a group to ethnic, religious, and cultural traits as well as to group-specific political ideas and tastes for vengeance (see Glaeser and Sacerdote, 2003). Second, the assumption that all the citizens are biased is not restrictive and just simplifies the algebra (see subsection

---

3 The first best implies that in half of the cases precautions are not taken and the expected cost of accident is $\frac{1}{2}p \Delta$ and that, in the other half, precautions are optimally taken and the only cost is $C$.

4 See Appendix 1 for a proof. Without preference normalization, for all $\lambda \neq 1$, $\hat{A}$ differs from $E(\hat{A})$ and $A^*$ and the lobbying game has more equilibria. Yet, the qualitative message of the model remains unaffected.
Third, the model applies not only to tort law but to every conceivable legal case as long as $a$ is interpreted as a noisy signal on the culpability of the parties in the hand of lower courts. Finally, the welfare function in (3) is only affected by group-specific tastes. Yet, as subsection II.C shows, if the technological efficiency of the rule—i.e., the distance between $A$ and $A'$—is considered the results of this section remain unaffected.

The timing is the following (see for a similar choice Fernandez and Ponzetto, 2008a):

1. At the Constitutional table ($t=0$) society chooses between statute and case law comparing the long run losses of welfare under the two institutions;
2. If statute law has been chosen, the legislator selects at $t=1$ her preferred rule $A_i$ at the end of a bargaining game (to be further discussed below) with a coalition of the two groups. Then, $A_i$ sticks forever. If, instead, case law has been chosen, a judge randomly picked from the two groups selects at $t=1$ her preferred rule $A_{c,1}$. This precedent guides adjudication until another judge overrules it turning, in the generic period $t$, $A_{c,t-1} = A_{c,1}$ into $A_{c,t}$ and possibly giving rise to a new round of changes. 

II.B Statute Law Versus Case Law

Let me start from the characterization of the equilibrium under case law.

Case law.—Stare decisis only binds in so far as it is costly for judges to change the precedent. As in Fernandez and Ponzetto (2008b), a judge inheriting precedent $A_{c,t-1}$

5 The set up also applies to those situations in which ex-ante contracting and ex-post observability lead to conflicts on the optimal allocation: constitutional design, contracting and insurance (see Gennaioli, 2006).
6 In the background, I also posit that: 1. adjudication is performed by lower courts acting as perfect agents of the law-maker—this is relaxed in subsection II.C; 2. to each adjudicated case follows a round of precedent change. This last assumption is essentially innocuous (see Fernandez and Ponzetto, 2008a).
and setting a $A_{t,j} \neq A_{t,j-1}$ can introduce any change in the law. Yet, the need to justify his decision implies a persuasive effort whose cost $(K/2)(A_{t,j} - A_{t,j-1})^2$ rises with both the relevance of the doctrine of *stare decisis* $K > 0$ and the magnitude of the legal revision. This last detail formalizes “the idea that the greater the effective deviation from precedent, the greater the difficulty to reconcile it with the rhetorical demands of *stare decisis*” (Fernandez and Ponzetto, 2008b). Judge $i$’s strictly concave problem in $t$ is:

$$
\hat{A}_{t,j} = \arg\max_A \Lambda_i(A_{t,j}) - (K/2)(A_{t,j} - A_{t,j-1})^2 = \arg\max_A -(1/2)(A_{t,j} - \hat{A})^2 - (K/2)(A_{t,j} - \hat{A}_{t,j-1})^2.
$$

A type $i$ judge minimizes both the loss she bears when $A_{t,j}$ differs from $\hat{A}_j$ and the one of justifying a $A_{t,j}$ far from $A_{t,j-1}$. The unique and global solution $\hat{A}_{t,j} = \Gamma(K)A_{t,j} + [1 - \Gamma(K)]\hat{A}$ implies that case law follows an auto-regressive process converging to the ergodic distribution with mean $\mathbb{E}(\hat{A}) = \mathbb{E}(\hat{A})$ and variance $\mathbb{V}(\hat{A}) = \mathbb{V}(\hat{A})(1 + 2K)^{-1}$ where $\mathbb{V}(\hat{A})$ is the variance of the groups’ favorite rules. The long run loss of social welfare is:

$$
\mathbb{E}(\hat{A}) \times (1/2)\mathbb{V}(\hat{A}) + (1/2)(\mathbb{E}(\hat{A}) - \hat{A})^2 = \mathbb{V}(\hat{A})^2/(2(1+2K))^{-1},
$$

where $\hat{A} = \bar{A} = \mathbb{E}(\hat{A})$ is the long run case law rule. Lemma 1 summarizes these results:

**Lemma 1** (Proposition 5, Gennaioli and Shleifer, 2007b): *Let $u$ be unobservable. Case law evolves as a first-order autoregressive process which converges to the ergodic distribution $N(\mathbb{E}(\hat{A}), \mathbb{V}(\hat{A})(1+2K)^{-1})$. The asymptotic variance $\mathbb{V}(\hat{A})$ rises (falls) with $\pi (K)$ and it is always smaller than the variance of the judges-citizens’ preferences $\mathbb{V}(\hat{A})$. Judge-made law develops as a process of incremental change such that each judge*

---

7 The evaluation of the impact of different doctrines of precedents—e.g., *jurisprudence constant*—remains as an important agenda for future research (see Depoorter et al., 2005).
marginally moves the rule inherited from precedents in the direction of his own bias. The prevailing legal rule always incorporates the perspectives of all previous courts as well as the current one, and heterogeneous biases tend to balance one another and induce reversion toward the mean of the judges-citizens preferences. This leads to long run unbiasedness at the cost of precedent volatility (see for a similar result Cardozo, 1921; Ponzetto and Fernandez 2008a and 2008b; Gennaioli and Shleifer, 2007b).  

**Statute law.**—Dating back to the rise of absolute monarchy (Hobbes, [1681] 2005), legislation has traditionally been interpreted as a sovereign act of the legitimate political authority. This conception became part of the common heritage of the Enlightenment and received its modern normative groundings from the establishment of democratic institutions. In a democracy, statutes reflect the society’s will the more accurately the stricter the constraints on the legislator—the government, the legislature, the president, etc.—are. Indeed, even if ultimately implemented by elected representatives, statutes are the outcome of a political process that involves also nonelected actors which represent the interests of groups of citizens particularly affected by the law.

To capture such a complex political arena, I embrace the endogenous lobbying model proposed by Felli and Merlo (2006) and I assume that \( A_i \) is selected by a legislator who maximizes a weighted average of social welfare and money obtained through bribes \( y \)

\[
U(A_i, y) = -(1 - \mu) \tilde{A}(A_i) + \mu |i| y,
\]

where \(|i|\) is the numerosity of cultural groups (two in this small world) and \( \mu \) an inverse measure of democracy (see, for a similar choice, Acemoglu and Robinson, 2006). The adjustment for \(|i|\) takes care of the mechanical bonus to democracy that an increase in

---

8 Fernandez and Ponzetto (2008a) and Gennaioli and Shleifer (2007b) clarify that this basic pattern survives when the selection of disputed for litigation is endogenous (but see also Depoorter et al., 2005).
the number of groups creates: this assumption simplifies the algebra without affecting any of the results reported below. The legislator chooses the lobbies that participate to the policy-making process. This assumption avoids the empirically unsatisfactory (see Wright, 1996) hypothesis typical of the menu-auction models (Grossman and Helpman, 2001) that all the lobbies are active. I also posit that the winning coalition pays a fixed non sunk “collective action” (see Olson, 1965) fee $\Psi$ to transfer the equilibrium bribe.\(^9\)

Together these last two model’s elements constitute the sense in which lobbying is endogenous. All in all, the lobbying game works as follows.

Each lobby $i \in I = \{L, H\}$ has utility $-((1/2)\Lambda_i(A_i) - y$, and can sign binding contracts with the legislator over $A_i$ in exchange for transfers $y$. The legislator has the option of not signing any contract and implementing the status quo $\tilde{A}$. Let $\Omega = \{\emptyset, \{L\}, \{H\}, \{L, H\}\}$ denotes the set of possible coalitions $l$. For any $A_l$ the legislator may implement instead of $\tilde{A}$, the generic $l$ is willing to pay $w_l(A_i, \tilde{A}) = \sum_{i \in A_l} w_i(A_i, \tilde{A}) = (1/2)\sum_{i \in A_l} [(\tilde{A} - \tilde{A})^2 - (A_i - \tilde{A})^2] + -\Psi$ which is the sum of the maximum individual rational bribes of each lobby $i \in I$—$w_i(A_i, \tilde{A})$—less the fixed bribing fee. The legislator, first, chooses an $A_l$ for any $l \in \Omega$

$$\hat{A}_l(l) \in \arg\max_{A_l} - (1 - \mu)\Lambda(A_l) + \mu\|W_l(A_i, \tilde{A})$$

(4)

and then she chooses her preferred bargaining coalition

$$\hat{l} \in \arg\max_{l \in \Omega} - (1 - \mu)\Lambda(\hat{A}_l(l)) + \mu\|W_l(\hat{A}_l(l), \tilde{A})$$

(5)

Therefore, a subgame perfect equilibrium of the bargaining game is given by a threshold

\(^9\) These are the costs of “establishing links with politicians, hiring professional lobbyists, building a communications network among members, designing a scheme of punishments for defaulting members” (Mitra, 1999). Only the algebra (the interpretation) but not the qualitative message of the model changes when the fixed cost is per-group (is borne ex-ante before the legislator moves).
function $\hat{A}_i(l)$, a transfer function $W_i(\hat{A}_i(l), \hat{A})$ and a coalition $\hat{I}$. Implicit in (4) and (5), there is the assumption that the legislator appropriates the entire willingness to pay of the winning bargaining coalition. Given the model’s focus on the ratio decidendi, the hypothesis is immaterial to the analysis. Moreover, I maintain that side transfers and the interim cost do not entail social wastes: relaxing the assumption makes the algebra more cumbersome but do not change the qualitative message of the model (see footnote 11).

First let me characterize the equilibrium policy:

Lemma 2.A: Let $u$ be unobservable. For any coalition $l \in \Omega$, there exists a unique optimal threshold level $A_i(l)$ that solves problem (4):

$$\hat{A}_i(l) = \left[ (1-\mu) \bar{A} + \mu \sum_{i=1}^{n} \hat{A}_i \right] (1-\mu + \mu\bar{A})^{-1}.$$  

(6)

If the legislator accepts to be bribed, the outcome of the bargaining is a compromise between the socially optimal rule $\bar{A}$ and the rules preferred by the lobbies included in the winning coalition. Given the quadratic loss function, $A_i$ takes the form of a weighted average further away from $\bar{A}$ the less democratic is society and the higher is the cultural heterogeneity. Each of these forces has, however, a double role in the equilibrium:

Lemma 2.B: Coalition $\{L, H\}$ is never chosen in equilibrium. Besides, there are levels of cultural heterogeneity $\pi$ and $\overline{\pi}$ with $\overline{\pi} \geq \pi$ which are both decreasing with $\mu$ and increasing with $\Psi$ and such that: 1. for $\pi \geq \overline{\pi}$ the legislator picks either $\{L\}$ choosing $\hat{A}_i(\{L\}) = (1-\mu) \bar{A} + \mu \bar{A}$ or $\{H\}$ choosing the rule $\hat{A}_i(\{H\}) = (1-\mu) \bar{A} + \mu \bar{A}$ (she is indifferent between the two) and for $\pi < \overline{\pi}$ the legislator does not accept any bribes; 2. for $\pi \geq \overline{\pi}$, $\{L\}$ and $\{H\}$ pay the fixed cost and for $\pi < \overline{\pi}$ they don’t. Thus, for $\mu > 0$ and $\Psi \in (0, \overline{\Psi})$, the long run loss of social welfare is $0$ for $\pi < \overline{\pi}$ and $\left( \mu^2 / 8 \right) \left( \bar{A} - \bar{A} \right)^2$, which increases
with both $\pi$ and $\mu$, for $\pi \geq \overline{\pi}$. Neither $\{L\}$ or $\{H\}$ will pay any fixed fee $\Psi > \overline{\Psi} = 1/16$.

The impact of tighter constraints on the legislator’s action—a lower $\mu$—is twofold: on one hand, in a more democratic society, the legislator finds it less appealing being bribed—i.e., $\overline{\pi}$ is higher—and, on the other hand, even if side payments are paid, the equilibrium threshold is nearer to the second best $\hat{A}$ the lower $\mu$ is. Also, in this last scenario, the distance between $\hat{A}$ and $\hat{A}$ is wider the higher the cultural distance $\pi$ is.

The optimal level of law-making centralization.—The welfare comparison between the long run losses of welfare driven by the two institutions is now straightforward. For $K < \infty$ case law can outperform statute law only if the policy chosen by the legislator is at least partially biased by bribing—i.e., if $\pi \geq \overline{\pi}$—and

$$\Lambda(\hat{A}) > E\left(\Lambda(\hat{A})\right) \Leftrightarrow \left(\mu^2 / 8\right) (\bar{A} - \bar{A})^2 > (\bar{A} - \bar{A})^2 \left[8(1+2K)\right]^{-1} \Leftrightarrow \mu^2 > (1+2K)^{-1},$$

which leads directly to the following proposition:

Proposition 1: Let $u$ be unobservable. Statute law is always strictly better than case law for $\pi < \overline{\pi}$ and $K < \infty$ and the two systems induce the same expected welfare for $\pi < \overline{\pi}$ and $K = \infty$. For $\pi \geq \overline{\pi}$ and for every $K$, there does exist a $\mu(K)$ with $\partial \mu(K) / \partial K < 0$, $\mu(\infty) = 0$ and $\mu(0) = 1$, such that statute law performs strictly better (weakly worse) than case law for every $\mu < \mu(K)$ (otherwise).

The intuition for this result is straightforward. When bribes affect the equilibrium rule, statute law assures certainty at the cost of a biased equilibrium rule. Case law, instead,

\[10\] These features stand if judges can be bribed. For $K > 0$, they accept money if $\pi$ is smaller than a bound lower than $\overline{\pi}$: case-law remains unbiased but its volatility grows so that statutes are still preferred. Also, the Constitutional Table’s decision is collusion proof, being the losing group’s willingness to pay for a reform of the optimal institution lower than that of the winning group. Finally, statutes become volatile if selected in each period by a legislator who accepts bribes: yet, the main proposition’s idea survives.
achieves the socially optimal rule but at the cost of volatility of the precedent.\textsuperscript{11} Both costs are a function of the extent of disagreement between the groups—\textit{i.e.}, the distance $\left(\bar{A}-\bar{A}\right)^2$. Therefore, for $\pi \geq \bar{\pi}$, whether one institution prevails on the other is only due to the comparison between the overruling cost and the benevolence of the legislator. Given a non perfectly democratic political process, a society which is able to enforce more strictly the doctrine of \textit{stare decisis} will prefer institutions giving more power to the decentralized legislative action of appellate judges.\textsuperscript{12} Similarly, given a less then infinite institutional relevance of \textit{stare decisis}, a society who better fares in assuring the probity of her representatives will embrace statute law.\textsuperscript{13} A fortiori, when the cultural distance is small—\textit{i.e.}, $\pi < \bar{\pi}$—statutes always outperform judge-made law.

This result conflicts with the analysis in Fernandez and Ponzetto (2008a) and their claim that case law is evolutionary superior to statute law. The deep reason is that they model statute law as a process of exogenous lobbying where the active ones are random. In this case, the threshold under statute law is volatile and the expected cost of this uncertainty is bigger than the one under case law provided that democracy is imperfect and \textit{stare decisis} is sufficiently relevant. Yet, a similar hypothesis is empirically unreasonable and

\textsuperscript{11} With fixed overruling costs—as in Gennaioli and Shleifer (2007a), there is a $\pi^*$ such that for $\pi \leq \pi^*$ the threshold fixed by the first judge sticks forever and for $\pi > \pi^*$ judges with opposite tastes overrule each other. Losses equal always $(1/8)(\bar{A}-\bar{A})^2$ and so case law is always at least weakly worse than statute law.

\textsuperscript{12} Costly lobbying only reinforces the proposition’s results, given that statutes become even less efficient when the legislator accept bribes at a high level of heterogeneity and/or at low level of democracy.

\textsuperscript{13} When citizen-candidates run to become the legislator, this section’s qualitative features stand provided that the winner is not too insulated from the opposing party—\textit{i.e.}, it puts a not too small weight on its losses of welfare. The two candidates gather the same number of votes and have the same probability of winning. Again statutes perform better if no bribe is paid or if $\mu$ is relatively lower than $K$. 

\footnotesize

14
comes very short in explaining which group is willing to bargain with the legislator. Next, I consider a series of extensions to the basic model and I assess the robustness of the results reported above. First, I characterize the equilibrium of the lobbying game in a larger world where there is also a third cultural group with unbiased preferences for \( A' \). Then, I explore whether Proposition 1 is upset if: 1. society trades off cultural biases with efficiency concerns; 2. appellate judges are forward-looking or (3.) can introduce new information distinguishing the precedent. The results coming from these exercises can be summarized saying that the qualitative message of the model—i.e., relatively heterogeneous and/or less democratic society should choose case law—remains true.

II.C Robustness

*Introducing an unbiased group.*—This time, the population is equally split among the unconcerned group with pro-offender bias \( \beta_L = \lambda \pi \), the concerned one with bias \( \beta_H = \lambda / \pi \) and a third group with no bias and favoring the technologically efficient \( \beta_E = \lambda \).

Again, the legislator panders to the extremist coalitions \( \{L\} \) and \( \{H\} \), which, again, pay the fixed cost for \( \pi \) sufficiently big. Yet, the presence of the unbiased group breaks the legislator’s indifference. The following lemma characterizes the new equilibrium:

**Lemma 3:** Let \( u \) be unobservable. For any coalition \( l \in \Omega \), there exist a unique optimal \( \hat{s}_l(l) \) defined by (6). For \( \lambda \leq 1 \) \((\lambda > 1)\), the legislator chooses \( \{L\} \) \((\{H\})\) and there are levels of cultural heterogeneity \( \tilde{\pi}_L \) and \( \tilde{\pi}_H \) for \( \tilde{l} = \{\{L\}, \{H\}\} \) decreasing with \( \mu \) and increasing with \( \Psi \), with \( \tilde{\pi}_L \geq \tilde{\pi}_H \), and such that: 1. for \( \pi \geq \tilde{\pi}_L \) \((\pi \geq \tilde{\pi}_H)\) the legislator praises \( \{L\} \) \((\{H\})\) and, for \( \pi < \tilde{\pi}_L \) \((\pi < \tilde{\pi}_H)\), she chooses \( \hat{A} \); 2. for \( \pi \geq \tilde{\pi}_L \), \( \{L\} \) \((\{H\})\) pays the fixed cost and for \( \pi < \tilde{\pi}_L \) \((\pi < \tilde{\pi}_H)\) it does not. Therefore, for
\(\Psi \in (0, \bar{\Psi}] \) and \(\mu > 0\), the long run loss of social welfare is 0 for \(\pi < \bar{\pi}_i\) and \((\mu^2/18)(A' + A - 2\bar{A})\) for \(\pi \geq \bar{\pi}_i\). None of the groups is willing to pay a bribing cost \(\Psi > \bar{\Psi} = 1/27\).

Thus, if the socially optimal rule is biased toward \(O\), which happens when \(E(\hat{A}) \geq A'\) or \(\lambda \leq 1\), the legislator favors \(\{L\}\).\(^{14}\) Coalition \(\{H\}\), instead, prevails if \(E(\hat{A}) < A'\). Again, the welfare comparison between the losses of social welfare under the two institutions is immediate. For \(K < \infty\) case law can outperform statute law only if \(\pi \geq \bar{\pi}_i\) and

\[
\lambda(\hat{A}_i) > E(\bar{\Lambda}(\hat{A}_i)) = \left[3(\bar{A})^2 + 3(A')^2 + 3(A')^2 - (\bar{A} + A')^2\right][18(1+2K)]^{-1}.
\]

This last inequality leads directly to the following proposition:

Proposition 2: Let \(u\) be unobservable. Statute law is always strictly better than case law for \(\pi < \bar{\pi}_i\) and \(K < \infty\), and the two systems induce the same expected welfare for \(\pi \geq \bar{\pi}_i\) and \(K = \infty\). For \(\pi \geq \bar{\pi}_i\) and for every \(K\), instead, it does always exist a \(\tilde{\mu}_i(\bar{K}, \pi)\) with \(\partial \tilde{\mu}_i / \partial K < 0\), \(\partial \tilde{\mu}_i / \partial \pi < 0\), \(\tilde{\mu}_i(\pi, \pi) = 0\) and \(\tilde{\mu}_i(0, \pi) = 1\), such that statute law performs better (weakly worse) than case law for every \(\mu < \tilde{\mu}_i\) (otherwise).

The result is in all similar to the one proposed in Proposition 1. Yet, this time, cultural heterogeneity has another (intensive) margin: as it increases, the autocracy threshold over which case law outperforms statute law falls. In this respect, it is worth to notice that the main model’s message remains true even for \(\Psi = 0\) (see the Appendix 1).

The rest of the section considers, for sake of simplicity, the small world with two groups studied above; however, all the results hold true if the unbiased group is introduced.

**Accounting for tastes versus efficiency.**—Let me now consider a world in which technological efficiency is weighted against preferences aggregation, in such a way that

\(^{14}\) For \(\lambda = 1\) the legislator is supposed to incline for \(\{L\}\): this simplification is clearly immaterial.
group $i$’s losses are \( [(1-T)/2] \left[ \hat{A} (1-A')^2 + (1-\hat{A}) (A')^2 \right] + (T/2) (A' - A)^2 = \Lambda^E(A') \) where \( T \) is the relative technological concern in the efficiency (notice the apex \( E \)) regime.

Group $i$’s favorite rule is \( \hat{A}_i^E = (1-T) \hat{A} + T A' \) so that:

\[
\Lambda^E(A') = (1/2) \left( A' - \hat{A}_i^E \right)^2 + (1/2) \left( 1 - \hat{A}_i^E \right) - (T/2) A' (1-A') \propto (1/2) \left( A' - \hat{A}_i^E \right)^2 .
\]

This time, the optimal threshold \( \hat{A}_i^E = E(\hat{A}_i^E) = (1-T) E(\hat{A}) + T A' \) is a compromise between each group’s preferred rule and the technologically efficient ratio-decidendi \( A' \).

Under case law, judge $i$’s objective function at time $t$ is \( -\Lambda^E(A^E_t) - (K/2) \left( A^E_t - A^E_{t-1} \right)^2 \); the unique and global solution is \( \hat{A}_t^E = \Gamma(K) A^E_{t-1} + \left[ 1 - \Gamma(K) \right] \hat{A}_t^E \), and judge-made law follows an auto-regressive process with ergodic distribution \( N\left(E(\hat{A}_t^E), V(\hat{A}_t) (1-T)^{(1+2K)^{-1}} \right) \). As a consequence, the loss of social welfare is proportional to \( (1-T)^2 V(\hat{A}) / 2(1+2K) \).

The legislator, instead, selects \( \hat{A}^E = \left[(1-\mu) \hat{A} + \sum \mu_i \hat{A}_i \right] (1-\mu+\mu_i)^{-1} \) and the equilibrium follows the same patterns described by Lemma 2.B when \( \hat{A} \) and \( \hat{A}_i \) are replaced by \( \hat{A}^E \) and \( \hat{A}_i^E \) (see Appendix 1). For \( K < \infty \), case law outperforms statute law only if \( \pi \geq \bar{\pi} \) and

\[
\bar{\lambda}(\hat{A}^E) > E\left( \hat{A} \left( \hat{A}^E \right) \right) \leftrightarrow \mu^2 \left( 1-T \right)^2 \left( A - \hat{A} \right)^2 \left( 1+2K \right)^{-1} \left( 1-T \right)^2 \left( \hat{A} - A \right)^2 \leftrightarrow \mu^2 > \left( 1+2K \right)^{-1} .
\]

Therefore, Proposition 1 applies unchanged. This result clarifies that, even when the investment—or growth—enhancing features of the law are considered, the optimal law-making institution is selected looking at the extent of cultural polarization and level of democracy typical of a society at a point in time.

Forward-looking judges.—As stressed by Gennaioli and Shleifer (2007b), another source of judicial moderation is forward looking judges: to reduce the possibility of future overruling, extremist judges may choose to set a less biased precedent today.
Such strategic behaviors lead to results similar to those illustrated in proposition 1 in the same environment studied by Gennaioli and Shleifer (2007b). In particular, let me assume that there are only two period, judges discount the future by a factor \( \delta \leq 1 \) and face a fixed cost of overruling equal to \( \bar{K} \). I solve the model by backward induction, starting from the final period two. In the second period judge overrules the precedent replacing it with his desired legal rule when \( (\tilde{A}_i - \hat{A})^2 > 2\bar{K} \); as a result, if \( -i \) is the group with preference opposite to \( i \), the first period judge maximizes:

\[
\hat{A}_i = \arg \max_{\hat{A}} - \left( \frac{1}{2} (A_{-i} - \hat{A})^2 - \frac{1}{2} (A_i - \hat{A})^2 + (1/2)(A_{-i}(-i) - \hat{A})^2 \right],
\]

where the dependence of second period rule from the identity of the decision maker is made explicit. Indeed, forward-looking concerns can induce a period one compromise that also a judge with similar preferences will overrule later on; more than that, if \( \pi \) and \( \delta \) are sufficiently high (\( \bar{K} \) sufficiently small) this is the only equilibrium. Indeed:

**Lemma 4:** Let \( u \) be unobservable, \( \bar{K} < 1/8 \), and \( \delta > 1/2 \). There exists a level of cultural heterogeneity \( \hat{\pi} \) such that: 1. for \( \hat{\pi} < \pi \) the first period judge selects \( \hat{A}_{-i} = \hat{A}_i \) which is not overruled; 2. for \( \hat{\pi} \leq \pi \leq \bar{\pi} \) a first period judge of type \( L \) (\( H \)) selects \( \bar{A} + \sqrt{2\bar{K}} \) \((\bar{A} - \sqrt{2\bar{K}})\) which is not overruled; 3. for \( \pi > \hat{\pi} \) a first period judge of type \( L \) (\( H \)) selects a threshold \( \bar{A} + \sqrt{2\bar{K}} \) \((\bar{A} - \sqrt{2\bar{K}})\) which is overruled in the second period only by a type \( L \) (\( H \)) judge. The loss of social welfare is \((1/8)(\bar{A} - \bar{A})^2\) for \( \pi < \hat{\pi} \), \((1/2)[(1/2)(\bar{A} - \bar{A}) - \sqrt{2\bar{K}}]^2\) for \( \hat{\pi} \leq \pi \leq \bar{\pi} \), and \((1/8)(\bar{A} - \bar{A})^2 + (1/2)[(1/2)(\bar{A} - \bar{A}) - \sqrt{2\bar{K}}]^2\) for \( \pi > \hat{\pi} \).

For \( \pi \geq \bar{\pi} \), the long run (second period) loss under statute is \((1/8)\mu^2(\bar{A} - \bar{A})^2\) and:

**Proposition 3:** Let \( u \) be unobservable, \( \bar{K} < 1/8 \) and \( \delta > 1/2 \). Statute law is always strictly better than case law for \( \pi < \bar{\pi} \). For \( \pi \geq \min \left\{ \max \left\{ \tilde{\pi}, \hat{\pi} \right\}, \max \left\{ \bar{\pi}, \hat{\pi} \right\} \right\} \), it does exist a \( \mu(K, \pi) \)
with $\partial \hat{\mu}(K, \pi)/\partial K > 0$ and $\partial \hat{\mu}(K, \pi)/\partial \pi < 0$ such that statute law performs strictly better (weakly worse) than case law for $\mu < \hat{\mu}$ (otherwise). Notice that if $8\Psi \leq \mu \tilde{K}$ then $\hat{\pi} \geq \overline{\pi}$.

Thus, even with forward-looking judges the qualitative result of the model remains true. Distinguishing.—Stare decisis requires judges to abide by the holding of the first court, but still allows them to distinguish the precedent by introducing a new dimension $b$ into consideration. Gennaioli and Shleifer (2007a) assume formally that binding precedent confines a distinguishing judge to choose two thresholds $\hat{B}$ and $\overline{B}$ that respect the precedent $\hat{A}$ in the sense that the offender is liable if and only if $\hat{a} < \hat{A}$ and $b \geq \hat{B}$, or $a \geq \hat{A}$ and $b \geq \overline{B}$. They also maintain that the new dimension must provide previously unobserved information about the optimal threshold: i.e., $b$ coincides with $u$.

Fernandez and Ponzetto (2008b) challenge this view reporting anecdotal evidence according to which judges use to distinguish previous cases on the basis of utterly immaterial dimensions. Taken to the extreme, under the Gennaioli and Shleifer’s (2007a) premises, the introduction of immaterial dimensions can lead to the collapse of judge-made law. In reality, the constraints introduced by binding precedents are less extreme than those advocated by Gennaioli and Shleifer (2007a), and a “judge can claim that the previous rule applies to an arbitrarily small portion of the infinitely larger space that he is now mapping: he thereby “reduces the range of the binding ratio decidendi to a vanishing point” (Stone 1964).” This ability to create “complex balancing tests based on marginal trade-offs between different factors” (Fernandez and Ponzetto, 2008b) can be easily captured in the model of subsection II.A supposing that eventually the second material dimension becomes observable and appellate judges can implement the optimal two-dimensional rule $B = 1 - a$ at a fixed cost $\tilde{K} > 0$. Next, I study the effect of this flexibility advantage of case law on the message proposed by Proposition 1.
In order to isolate the effect of distinguish, I shut completely down overruling and I assume accordingly that $K$ is equal to 0. The following lemma summarizes my findings:

**Lemma 5:** Given $\tilde{A}$ and $K=0$, suppose that $u$ becomes observable, then there are two levels of the distinguishing cost $\tilde{K}$ and $\bar{K}$ with $\tilde{K} > \bar{K} \geq 0$ and a $\tilde{\pi}(\bar{K})$ such that for every $\tilde{K} \in [\tilde{K}, \bar{K})$, if $\pi \geq \tilde{\pi}(\bar{K})$ then the optimal two-dimensional rule $B = 1 - a$ is introduced with probability one in the long run and the first best prevails.

Conflicts and overruling persist so long as ignorance on the second dimension does; when $u$ becomes observable all judges agree on the reduction at zero of both the false positive and negative irrespective of their own biases. This matches the basic idea expressed by Posner ([1978], 2007) that common law tends to achieve efficient rules. Nonetheless, this is only possible when cultural heterogeneity is sufficiently high to justify the payment of the fixed cost. This observation leads to the following:

**Proposition 4:** Given $A$, let $K=0$ and $\tilde{K} \in [\tilde{K}, \bar{K})$. Suppose that $u$ becomes observable, then for $\pi < \tilde{\pi}$ ($\pi \geq \tilde{\pi}$) statute law outperforms case law (the opposite is true).

Therefore, also under distinguishing, the qualitative message of the model stands.\(^\dagger\)

### III. Rules versus Discretion in Adjudication

Many scholars in comparative law see the reliance on specific “bright line” rules rather than broad principles for adjudication, as a defining feature of civil law systems (see Merryman, 1969; Damaška, 1986; Zweigert and Kötz, 1998; Glendon et al., 1999).

\(^\dagger\) This is the case even when the flexibility advantage of case law is milder and the legislator can observe and introduce the refined two dimensional rule in period $t = 1$. In this case, given that each group strictly prefers to introduce the informative innovation, case law will be only weakly better for $\pi \geq \tilde{\pi}$.
In the previous section, the model has been solved under the maintained assumption that bright line rules of adjudication were always in place—i.e., assuming that the rule selected by the law-maker is strictly respected by lower adjudicating courts. Under discretion, instead, lower courts can adjudicate the case according to a different de facto rule by, for instance, manipulating the evidence produced in a trial (Gennaioli, 2006; Gennaioli and Shleifer, 2008), provided that procedural rules allow judges to depart from the de jure ratio decidendi. In this subsection, I show that this procedurally accepted form of discretion emerges as an efficiency attempt to blunt the losses driven by the volatility of precedents when case law is in place.

Consider the following mild form of discretion: a lower court can use any de facto rule whose distance from the de jure $\hat{A}_j$, with $j = \{c, s\}$, is weakly lower than $\alpha \leq (1/2)(\bar{A} - \underline{A})$ bearing a fixed justification cost $\theta = \bar{\theta}(\pi) + \varepsilon$, with $\bar{\theta}(\pi) = (\alpha/2)[(\bar{A}(\pi) - \underline{A}(\pi)) - \alpha]$ and $\varepsilon$ infinitesimally small. As Appendix 1 shows, this assures that a lower court of type $H$ ($L$) will simply use $\hat{A}_j - \alpha$ ($\hat{A}_j + \alpha$) when $\hat{A}_j > E(\hat{A})$ ($\hat{A}_j < E(\hat{A})$) and will stick with the de jure rule if $\hat{A}_j \leq E(\hat{A})$ ($\hat{A}_j \geq E(\hat{A})$); besides $E(\hat{A})$ is the best rule that a lower court can obtain under discretion.\textsuperscript{16} Introducing discretion entails the reform cost $\Theta > 0$.

The latter captures, for instance, society’s aversion to discretionary adjudication. Also, I posit that under case law, in each period $t$, first an appellate court fixes a new precedent and then a lower court eventually uses discretion; moreover, a type $i$ appellate judge cares about the de jure rule and—in a naïve way—about the extra-volatility due to the discretion created by lower courts. Indeed, a judge of type $i$ suffers a loss increasing with the distance from $A_j$ and the rule that always avoid discretion—i.e., $E(\hat{A})$:

\textsuperscript{16} This assumption spares tiresome qualifications about possible boundaries solutions.
\[ \hat{A}_{i,j} = \arg \max_A -\frac{1}{2} (A_{i,j} - \hat{A})^2 - \frac{K}{2} (A_{i,j} - \hat{A}_{i,j-1})^2 - \frac{\rho}{2} (A_{i,j} - \mathbb{E}(\hat{A}))^2, \]

where \( \rho \) measure the judicial aversion to discretionary adjudication.

Consider, first, case law. Judge \( i \) minimizes the sum of three losses. The first two are the usual ones, the third one is the loss driven by lower courts discretion. Thus, she sets

\[ \hat{A}_{i,j} = K(1+K+\rho)^{-1} A_{i,j-1} + (1+K+\rho)^{-1} \left[ \hat{A} + \rho \mathbb{E}(\hat{A}) \right], \]

which implies that case law converges to the ergodic distribution with mean \( \mathbb{E}(\hat{A}) \) and variance \( \mathbb{V}(\hat{A}) \left[ (1+2K)+(\rho^2+2\rho+2\rho K) \right]^{-1} \).

Thus, the equilibrium de facto rule \( \hat{A}_{i,j}^0 \) equals \( \hat{A}_{i,j} + \hat{\alpha} \) where \( \hat{\alpha} \) is the random variable equal to \( -\alpha \) (\( \alpha \)) when \( \hat{A}_{i,j} > \mathbb{E}(\hat{A}) \) \( \left( \hat{A}_{i,j} < \mathbb{E}(\hat{A}) \right) \)

and the low court type is \( H \) \( (L) \) and 0 otherwise. The ex-ante expected value of \( \hat{A}_{i,j}^0 \) is

\[ \mathbb{E}(\hat{A}_{i,j}^0) = \mathbb{E}(\hat{A}_{i,j} + \hat{\alpha}) = \mathbb{E}(\hat{A}_{i,j}) + (1/2) \mathbb{P}(\hat{A}_{i,j} > \mathbb{E}(\hat{A})) \left( -\alpha \right) + (1/2) \mathbb{P}(\hat{A}_{i,j} < \mathbb{E}(\hat{A})) \alpha = \mathbb{E}(\hat{A}) + (1/4)(\alpha - \alpha) = \mathbb{E}(\hat{A}). \]

Moreover, the ex-ante expected value of the de facto variance is given by:

\[ \mathbb{V}(\hat{A}_{i,j}^0) = \mathbb{V}(\hat{A}) \left[ (1+2K)+(\rho^2+2\rho+2\rho K) \right]^{-1} + (1/2) \mathbb{P}(\hat{A}_{i,j} > \mathbb{E}(\hat{A})) \alpha^2 + (1/2) \mathbb{P}(\hat{A}_{i,j} < \mathbb{E}(\hat{A})) \alpha^2 = \mathbb{V}(\hat{A}) \left[ (1+2K)+(\rho^2+2\rho+2\rho K) \right]^{-1} + \alpha^2/2, \]

All in all, the volatility of judge-made law under the pure common law—i.e., case law plus discretion in adjudication—is lower than the one under the mixed system—where case law is put side by side with bright line rules, provided that \( \rho \) is bigger than a threshold (obtained in Appendix) \( \bar{\rho} \) which is only a function of \( K \). As a consequence, discretion will always be introduced whenever \( \Theta \) is small enough—lower than the \( \bar{\Theta} \) found in Appendix 1. Consider now statute law. Problem (4) can now be written as:

\[ \hat{A}_{i,j}^0 \in \arg \max_{\hat{A}_{i,j}^0} - \left[ (1-\mu) \frac{1}{2} \right] (\hat{A}_{i,j}^0 + \hat{\alpha} - \hat{A})^2 - \frac{\mu}{2} \mathbb{E}(\hat{A})^2 - \frac{\mu}{2} \mathbb{E}(\hat{A})^2 \left( \sum_{\omega} (\hat{A} + \hat{\alpha} - \hat{A})^2 - (\hat{A}_o + \hat{\alpha} - \hat{A})^2 \right). \]
Both the legislator and the bribing coalitions always foresee the lower courts’ action and, as a result, the former is always able to neutralize discretion choosing the rule

\[ \hat{A}^*(l) = \left[ (1 - \mu) \hat{A} + \mu \sum_{i=1}^2 \hat{A}_i \right] (1 - \mu + |\mu|) \alpha. \]

The *de facto* threshold \( \hat{A} + \hat{A} \) equals the rule prevailing when discretion is forbidden, and Lemma 1, 2.A and 2.B continue to apply. Clearly, the pure civil law system is optimal for every \( \Theta > 0 \). The following proposition summarizes this section’s findings:

**Proposition 5:** Let \( u \) be unobservable, \( \theta = \bar{\theta} + \epsilon \), \( \rho < \rho(\mathcal{K}) \), and \( \Theta \leq \bar{\Theta} \). If case (statute) law is in place, discretion always outperforms bright line rules (otherwise).\(^{17}\)

*From theory to evidence.*—Provided that the restrictions on \( K, \hat{k}, \delta \) and \( \theta(\pi) \) hold, the above propositions can be restated as testable predictions as follows:

**TESTABLE PREDICTIONS:** The likelihood of a reform toward case and, in general, common (statute and, in general, civil) law will increase (decrease) with the degree of cultural heterogeneity and fall (rise) with the quality of democratic political institutions.

## IV. Civil versus Common Law: Evidence

Next, I will test these predictions, starting from the statute versus case law comparison.

### IV.A Data and Strategy

I require both a sample of reforms from statute to case law (or the way round) and proxies for the extent of cultural heterogeneity and the quality of political institutions.

*Law-making reforms.*—As previously highlighted, case law originated from the laws of

---

\(^{17}\) For \( \Theta > \bar{\Theta} \) the model produces a hybrid tradition where case law is put side by side to bright line rules. Germany, Switzerland and the Scandinavian countries are clear examples of such a type of legal system.
England and was transplanted through conquest or colonization to England’s ex-
colonies—e.g., U.S.A., Australia, Canada, and many countries in Africa and Asia.

Statute law, as embedded by the French “Code civil des Française”, brought into force
in 1804, and the related Austrian (1811), Portuguese (1867), Spanish (1889) and Soviet
Union (1922) Civil codes has, instead, been exported first to Latin America, Africa and
Asia and then to the former Sovietic republics (see Zweigert and Kötz, 1998).

Differently from these two groups, Denmark, Germany, Sweden and Switzerland have
approved substantive pieces of codification entrusting, at the same time, a key role to
judge-made law as a key instrument for the maintenance of those parts of the statutes
becoming with time obscure and/or inadequate.18 This last experience was especially
relevant for a group of countries—notably China, Ethiopia, Thailand and Japan—which,
just after independence—mostly in the aftermath of the Second World War—faithfully
modeled their legal system on the shape of these European examples.

Starting from the peculiar historical features of these transplantation waves, Berkowitz
et al. (2003) have introduced the distinction between countries that are origins—because
they have developed their legal order internally—and countries that are transplanted—
because they have received their legal order either through colonization or by adoption
of the legal tradition considered most advanced or fashionable at the time. At the same

18 As stressed by Zweigert and Kötz (1998, pp. 153) “the maintenance of the general structure of the BGB
[German Civil code, 1900] is really the work of courts”. The latter have heavily relied on the BGB
general clauses (#157, 242 and 826) to adequate the codes to social and technological innovations in the
name of good faith and the defense of private transactions. Similarly, article 1 of the Swiss Civil Code
(1881) claims that, “where no provisions are applicable, the judge shall decide according to the […] rules
which he would lay down if he had himself to act as legislator. Herein he must be guided by […] legal
that have embraced the German and the Swiss codes does not significantly affect the main results.
time, they have stressed for the last group the essential randomness of the initial matching between legal tradition and national legal systems at independence.

Yet, starting from independence, countries penalized by this match—for instance, culturally homogenous ex-English colonies with well-functioning institutions—had the chance of reforming their institutions toward their ideal legal order: the following empirical exercise takes exactly advantage of this unique transplantation experiment.

In order to accomplish the task, I made use of two data sources: 1. the first volume of the International Encyclopaedia of Comparative Law written by some of the most illustrious scholars of comparative law—Conard, David, Von Mehren, and Zweigert—between 1969 and 1995, and listed among the references as Abdesselem et al. (1995); 2. the appendix accompanying the Djankov et al.’s (2003) dataset on legal procedures, which is listed among the references as Acartürk et al. (2005). I considered as origins Austria, Denmark, France, Germany, Russia, Sweden, Switzerland and, of course, the United Kingdom and as transplanted the 156 countries studied in the above mentioned publications for which enough data on the quality of democratic institutions and on the extent of cultural heterogeneity are available. I collected data on the law-making process in place in each of these transplanted countries at independence and in the year 2000, distinguishing case and statute law according to the following criterion:

**Definition:** A country uses case law, at a given point in time, if the decisions of some appellate courts but not only of the Constitutional court are considered as a source of

---

19 When I restrict the sample to the ex-colonies only, the main results remain unaffected.

20 A more (less) restrictive criterion is to consider as using statute (case) law also the countries in which only the decisions of a subset of appellate courts (Constitutional-Cassation court) are treated as binding. If I switch to one of these two alternatives, the main section’s results remain unaffected. Yet, the analysis of different case law processes and of different doctrines of precedents is a fertile ground for future research.
law and treated as binding by lower courts. Otherwise, a country employs statute law. With these data, I can define the following two dummy variables: Statute_O (Statute) is equal to one if a country used statute law in the first year of independence (in 2000). Accordingly, the long run evolution of the law-making institutions in each country can be summarized by the indicator Case. The latter is equal to 4 if Statute_O equals one and Statute zero; 1 if the opposite is true; 3 if both Statute_O and Statute equal zero; 2 if Statute_O and Statute have value one.

There are 9 countries with a four score, 61 with a three, 69 with a two and 17 with a one. Switching the 2 for the 3 does not affect in any relevant way my results.21

Proxying cultural heterogeneity and democracy.—Testing the model’s predictions also requires measuring the extent of cultural heterogeneity and the strength of democratic institutions in the sample. With regard to the former, I make use of two variables. The first one captures the cultural distance between the principal population in the origins—the country which chose the law-making institution in the first place—and the one in the transplanted country.22 The second proxies for the cultural fractionalization within the transplanted country. Thus, while the former should take care of the inefficiency driven by the transplantation match, the second should uncover the impact of the heterogeneity in cultural values among groups of citizens on the evolution of the legal system.

Creating a meaningful proxy for the cultural distance between populations is not an easy task. Yet, Cavalli-Sforza et al. (1994)—CS hereafter—suggest that an index of genetic distance—the coancestry coefficient (Genetic_Distance)—is a natural proxy. In their study of nation formation in Europe, Desmet et al. (2006) have recently documented a

21 The timing of each reform is unobserved. Identifying the latter is a relevant agenda for future research.

22 If the distance is, instead, between the principal population in the transplanting country and the principal population in the transplanted country, the results shown below remain essentially unaffected.
strong relation between a measure of opinion poll distances based on answers from the section on perceptions of “life, family and religion and moral” of the World Value Survey and Genetic Distance. As an explanation to this pattern, it is illuminating to recall how CS justify the strong link existing between population’s linguistic roots and Genetic Distance: “both genetic and cultural contacts take place by the same routes [e.g., marriage]; they respond to the same geographic and ecological barriers; and they can also reinforce each other” (CS, p. 23).23

The coancestry coefficient is based on the existence of genetic or DNA polymorphisms (situations in which a gene or a DNA sequence exists in at least two different forms [alleles])24 and it is calculated as the sum of the differences in frequencies of these polymorphisms. Higher values of Genetic Distance imply wider genetic differences. Gene frequencies change over time mainly because of migration, natural selection and random genetic drift—i.e., random sampling of polymorphism at each generation.

Yet, accurately describing the evolution of selection and the different migration waves is an essentially unfruitful undertaking. To the latter extent, CS sampled only genetic markers unaffected by evolution from those aboriginal populations in place before 1500 (i.e., before that colonization had fuelled the greatest migration episode in human history). Therefore, Genetic Distance is immune to measurement errors and exogenous to all those unobserved historical patterns which affected cultural fractionalization and

23 Building on this intuition, Guiso et al. (2007) have used the coancestry coefficient as an explanatory variable for trust in trade gravity regression and Spolaore and Wacziarg (2008) have proposed Genetic Distance as a proxy of family-transmitted characteristics, including cultural and moral values.

24 For example, even if the ABO blood group alleles are present in all populations, the frequency of each allele varies substantially across populations so that the estimated frequency of the O allele is 61% within Africans and 98% in American Natives. These differences arise for many other genes or DNA sequences.
political institutions during and after the transplantation experience (see also Spolaore and Wacziarg, 2008). Data on genetic distance are available for 42 populations groups. Following Spolaore and Wacziarg (2008), I matched populations to countries using the population labels listed in Appendices 2 and 3 of CS and the historical information reported in the Encyclopedia Britannica (2008).25

The strong connection among the genetic pool, the socioeconomic distance, linguistic families and culture described above leads me to choose the ethnic and language fractionalization in a country as a proxy for the “within” heterogeneity. The most widely used measure of ethno-linguistic fractionalization is the probability that two individuals randomly chosen from the population differ in the characteristic under consideration. The results presented below are based on $Av_{\text{Elf}}$, 26 which is built from 1960 data coming from a Soviet ethnographic source (Atlas Narodov Mira, 1964).

Turning to democracy, I will make use of a proxy for the strength of the constraints on the executive—$\text{Constraint}_{\text{Executive}}$—available from the Polity IV dataset for most of the countries of the sample starting from 1800.27 $\text{Constraint}_{\text{Executive}}$ is defined as the average constraints on the executive over the period elapsing between independence and the year 2000, ignoring missing data. This index is mainly related to the existence of checks and balances among various layers of the decision-making process and not on

25 The resulting matching is equal to that used by Spolaore and Wacziarg (2008) except for a few entries (Egypt, Finland, Hungary, Kenya, Lybia, Malawi, Poland, Tanzania, Tunisia and Uganda) for which the histories of demography in the Encyclopedia Britannica suggest different classifications. CS provide a conceptually similar measure of genetic distance (Nei’s distance) which deliver very similar results.

26 The qualitative results are in all similar when I use the linguistic, ethnic and religious fractionalization indexes proposed by Alesina et al. (2003) or the absolute number of ethnic or linguistic groups.

27 Different measures of democracy available for the same period (see, for instance, Vanhanen, 2003; Marshall and Jaggers, 2004) deliver essentially similar results.
other civil liberties, like the rule of law or the freedom of the press. The latter could be correlated with unobserved country-specific characteristics driving, for instance, also the population tolerance for different cultural preferences (see Acemoglu et al., 2008).

*Conditional independence.*—Yet, as made clear by a recent literature on endogenous linguistic and ethnic group formation (see Ahlerup and Olsson, 2008; Michalopoulos, 2008), neither $Av_{Elf}$ nor $Constraint_{Executive}$ are immune to a series of other reverse causation and self selection biases. This literature formalizes the intuitions proposed by two major theories of social evolution: the primordial and the constructive views.

While the former contends that ethnicity is mostly a primordial by-product of extended kinship, the latter has proposed two more recent factors: 1. the rise of the stratified society—*i.e.*, the state and the nation with its formal legal order—deprived the extended kinships of their *raison d’être*, pushing for a more homogenous citizenship; 2. the European colonizers, instead, pursued a “divide-and-rule” approach in order to increase the chance of successful exploitation of the colonies (see also Michalopoulos, 2008). Besides, this latter exploitation was pursued to different degrees depending on the adaptability of the colonizer to the climate and to pathogen loads, which, in turn, have been identified as crucial determinants of species richness by a large legacy of biology and ecology research (see Ahlerup and Olsson, 2008). Thus, it is not difficult to think of unobserved social and physical factors fostering, at the same time, either preferences for

---

28 Cultural markers such as dialects arise to differentiate the members of the extended family or clan from “the others”, facilitating in this way the provision of public good and the preservation of the group physical and human capital comparative advantages (see Ahlerup and Olsson, 2008).

29 Acemoglu et al. (2001) document that Europeans refrained from developing extractive societies, and set up institutions protecting private property rights and placing effective constraints on politicians and elites only where they were more likely to successfully settle.
ethnic homogenization (especially after a colonial experience) or the quality of political institutions and the need of reforms of the legal system.

Next, I will first attack the above predictions maintaining the conditional independence assumption and, then, I will relax the latter hypothesis using instruments for Av_Elf and Constraint_Executive. Finally, I will consider other biological and geographical factors which most likely belong to the set of unobserved forces discussed above.

Empirical strategy.—Next, I need a model for more than two non-ordered alternatives where controls do not vary over alternatives. The natural choice is the multinomial logit

\[
\Pr(y_i = k | z_i) = \frac{\exp(\beta_i' z_i)}{\sum_{i=1}^{4} \exp(\beta_i' z_i)},
\]

where \(y_i\) is Case and \(z_i\) contains Genetic_Distance, Av_Elf, Constraint_Executive, the log of the real GDP per capita—GDP—and the size of the population—Population—averaged over the 1950-2000 period, and the number of years between independence and the year 2000—Time_Independence.\(^{30}\) The latter should crudely proxy for lower switching costs or stronger preferences for homogenization. Table 1 (2) reports sources (summary statistics) of all the controls. It is worth to notice that Genetic_Distance and Constraint_Executive are normalized in order to range between 0 and 1.

IV.B Statute Law versus Case Law: What do the Data Say?

Exogenous controls.—Panel A of Table 3 lists the marginal effects obtained when the \(z_i\) gathers only the cultural heterogeneity proxies. These figures give the percentage variation in the likelihood of the outcome considered when the control rises by one percentage point, and they are strongly consistent with the model’s predictions. A one-standard-deviation rise in the normalized genetic distance—0.29—implies a little more

\(^{30}\) Using the mean of GDP and population over the period 1900-2000 does not change the analysis.
than 4 (6 in column 3) percentage point increase in the likelihood of a reform from statute to case law and a little more than a 5 (5 in column 3) percentage point decrease in the likelihood of a reform from case to statute law. Also a one-standard-deviation rise in the ethno-linguistic fractionalization—0.30—is associated to a 3 percentage point increase (1.5 in column 3) in the likelihood of a reform from statute to case law and a 6 (6 in column 3) percentage point decrease in the likelihood of a reform from case to statute law. The effect of the “between” cultural heterogeneity is definitely stronger than the one of the “within” heterogeneity. Besides, only the former is statistically significant at a level nowhere lower than 5 percent. Finally, a longer spell of political autonomy raises the likelihood of both reforms while GDP (Population) significantly decreases (increases) the probability of reform toward statute law only. This last set of result is common to all the remaining tables and accordingly it is not reported.

Panel B adds to the picture Constraint_Executive. Here, the evidence is more mixed. Stricter constraints on the executive decrease the likelihood of both reforms. In column 4, a one-standard-deviation rise in the constraints on the executive—0.32—implies a little less than a 7 percentage point fall in the likelihood of a reform from statute to case law and a 7 percentage decrease in the likelihood of a reform from case to statute law. Yet, while the first effect tends to become more significant when cultural heterogeneity is taken into consideration, the second becomes insignificant.\footnote{An example of the results in column 4 is the high Genetic_Distance (0.54) low Constraint_Executive (0.19) Mexico that inherited statute law from Spain and introduced, after independence—1810, a set of “compulsory case law” institutions (see Abdesselem et al., 1968-1995: pp. M-68).}

\textit{Endogenous controls.}—In order to relax the conditional independence assumption, I require, first of all, a model for endogenous discrete variables and, then, instruments related to the endogenous controls but not to the legal system evolution.
Focusing on the first point, I use the IV probit with dependent variable *Case_Law*. The latter is equal to 1 when *Case* is equal to 4 and 0 otherwise. Given the differences in the underlining models, the marginal effects obtained from this model cannot be compared with those of the multinomial logit used above.

More complex is to select reasonable instruments. The primordial theory of social evolution provides a natural instrument for *Av_Elf*. In particular, Michalopoulos (2007) argues that, starting with the Neolithic agricultural revolution in the Near East—10,500 BP, heterogeneous land endowments generated region specific human capital, limiting population mobility and leading to the formation of localized ethnicities and languages. This intuition is empirically supported by cross-regional data. So, I use an index of richness of soil types—*Soil_Diversity*—proposed by Ahlerup and Olsson (2008). The variable aggregates qualitative measures of soil quality proposed by Zobler (1986) and unaffected by land-use or human intervention.\footnote{The qualitative characteristics from which the index is calculated refers to the texture of the soil: *i.e.*, to the quality in terms of erosion and water retention of a piece of land.} This property assures the exogeneity of *Soil_Diversity* to the evolution of the legal systems.

Turning to the quality of the political institutions, a recent literature informed by cross-cultural psychology studies (Licht et al., 2007; Tabellini, 2008) has linked the respect for the rule of law and the preservation of the democratic accountability to the cultural emphases on autonomy. “Societies whose cultures emphasize individual uniqueness and view individual persons as moral equals are likely to develop norms that promote societal transparency as a means for social coordination […]”. In contrast, societies [that] view the individual as an embedded part of hierarchically organized groups […] accommodate exercise of power from above” (Licht et al., 2007). As stressed above, language constitutes a stable factor constraining the development of cultural norms. A
key feature of languages is the rules on pronouns. Languages that require to include the first person pronoun in a sentence (have a set of second person pronouns differentiated according to the social distance of the speakers) help distinguishing the subject from the general context (instill in the subject the respect for social hierarchies). Therefore, I follow Kashima and Kashima (1998) and I make use of the first person pronoun drop license \( \text{Pronoun\_Drop} \) and of the existence of differentiated second person pronouns \( \text{Second\_Person} \) as instruments for cultural emphases on embeddedness.

Table 4 reports in the lower panel the first stage results and in the upper panel the second stage IV probit marginal effects.\(^{33}\) The results not only confirm the relevance of \( \text{Genetic\_Distance} \) and \( \text{Constraint\_Executive} \) in explaining reforms toward case law in the direction foreseen by the model but also suggest that the weak significance of \( \text{Av\_Elf} \) was mainly due to the failure of the conditional independence hypothesis. The coefficient attached to \( \text{Av\_Elf} \) is now significant in column 3 at the 1 percent level. This pattern is robust to the introduction of other relevant covariates.

Several recent papers have proposed that less unpredictable climate and more severe pathogen loads negatively affect both the species richness (Ahlerup and Olsson, 2008) and the quality of political institutions (Acemoglu et al., 2001). Columns 4 and 5 list the results when the rescaled latitude—\( \text{Latitude} \), the maximum difference in land altitude—\( \text{Altitude\_Diff} \)—and an index for the risk of malaria—\( \text{Malaria} \)—are considered. These controls proxy, respectively, for less unpredictable climate, higher soil diversity and a more dangerous pathogen load. Finally, columns 6 directly tests the Glaeser and Shleifer’s (2002) hypothesis adding a proxy negatively linked to coercion by elites—

\(^{33}\) The evidence is in all similar when I use a two stage multinomial logit obtained bootstrapping a routine that substitutes the fitted values from the first stages for the endogenous into (7). Moreover, interaction effects are usually not significant at a probability of the reform level of 0 and 0.5 (Ai and Norton, 2003).
Economic_Power. The index assumes higher values the more diffuse is the economic power within the population of a country (see Vanhanen, 2003). All the proxies except Economic_Power enter with a coefficient which is highly significant and displays the expected sign. The failure of the Glaeser and Shleifer’s (2002) intuition, however, deserves a further comment. Even if Economic_Power could also proxy for weaker political institutions, the results seem to confirm the doubts cast by a recent economic history literature looking at the medieval experiences of England and France (see Roe, 2007; Klerman and Mahoney, 2007).

For what, finally, concerns the quality of the instruments, the exogeneity test—shown for every specification for which Stata is able to calculate it—and the first stage $R^2$ confirm their quality. These covariates, moreover, enter into the first stages in a nice separable form whereby soil diversity mainly affects the within heterogeneity, and the lack of individualism mainly affects the strength of the constraints on the executive. Given the correlation between Av_Elf and Constraint_Executive—0.16, this assures that the attenuation bias induced by the variable with the greatest measurement error does not load on to the other one (see Acemoglu and Johnson, 2005). Since the first stages in this and subsequent tables are almost identical, I do not report them to save space.34

IV.C Rules versus Discretion in Adjudication: What do the Data Say?

Next I will attack the second set of testable predictions concerning with the evolution of the transplanted legal system toward a pure common or civil law tradition. Two are the data sources: 1. Acartürk et al. (2005); 2. several treaties (list available from the Author)

---

34 A mirror evidence is obtained when the dependent of the IV Probit equal 1 if Case is equal to 1 and 0 otherwise. Using the difference in precipitation or temperature (an index of intellectual power) in spite of Altitude_Diff (Economic_Power) does not affect the main results.
detailing the history of the civil procedure of the countries that have transplanted the legal order into the nations studied in Acartürk et al. (2005).

Djankov et al.’s (2003) analyze procedural rules governing the adjudication of simple legal disputes—the eviction of a non-paying tenant and the collection of a bounced check—for 109 countries in 2000. Among other proxies for procedural formalism, Djankov et al. (2003) compute the following four dummies which I use as proxies for lack of discretion in the adjudication of civil disputes: 1. Appellate_l&f_F equals one if issues of both law and fact can be reviewed in appeal and zero if only new evidence or issues of law can be reviewed, or if there is no appeal; 2. Judge_Law_F equal to one if judgment must be on law only, and zero when they may be based on equity grounds; 3. Written_F with value one if the evidence is mostly submitted to the court in written form—attachments, affidavits, or other—and zero if most of the evidence is presented at oral hearings; 4. Inquisitorial_F equal to one (zero) if the evidence gathering is inquisitorial (adversarial). These dimensions are consistently (see Merryman, 1969; Zweigert and Kötz, 1998; Glendon et al., 1999) considered as discriminating among the civil procedures typical of the hierarchical, inquisitorial and technical civil law tradition and those of the coordinate, adversarial and substantive common law one.

Similar indicators (Appellate_l&f_O, Judge_Law_O, Written_O, Inquisitorial_O) can be built for the first year of independence for 98 countries, starting from the history of

35 Judges can both freely request or take evidence that has not been offered, requested or introduced by the parties, and refuse to collect or admit requested evidence.

36 The link with the degree of discretion by lower adjudicating courts can be summarised as follows. While extensive supreme review and written records assure a credible check on discretionary lower courts; inquisitorial procedural requirements curbs the parties’ incentive to report noise and contrasting evidence which calls, in turn, for interpretation of the law (see Damaška, 1986: pp. 3-6). Finally, an explicit limitation to equitable judgements is of similar use (see Merryman, 1969: pp. 123-127).
civil procedure governing similar commercial cases in the transplanting country. I stack one over the other the Appellate_l&f_F, Judge_Law_F, Written_F and Inquisitorial_F indexes for the two procedures in order to obtain four new index with two observations for each of the 98 countries (see for a similar procedure Balas et al., 2008). Assuming, next, that Appellate_l&f_O, Judge_Law_O, Written_O, Inquisitorial_O approximate the two disputes in an similar way, I build four multiple outcomes indicators similar to Case—i.e., Appellate_l, Equity, Oral, Adversarial. The latter have two observations for each of the 98 transplanted in the sample.

Two crucial background hypotheses make this second empirical exercise more fragile than the one discussed above: 1. rules in place in a transplanted at independence are fairly well approximated by those in place in the transplanting country, which are the ones I observe; 2. the procedure for small commercial cases at independence and the rules governing the eviction of a non-paying tenant and the collection of a bounced check in 2000 are good proxies of the underlying “discretion in adjudication” concept.

Recovering legal traditions.—The analysis of these sub-indices generally point in the same direction. Yet, as Rosenthal and Voeten (2007) point out, this is not a convincing argument in favor of the approach usually embraced by the “legal origins” literature. Similar works by Djankov et al. (2003) and Balas et al. (2008) have chosen as study objective the sum of all the indexes. Such a methodology relies on some untested assumptions which could be defended only if we had particularly strong a priori about the relevance of each indicator. Indeed, the procedure assumes that each indicator is equally informative about the underlying concept, that the employed indicators measure a single dimensional concept and, finally, that measurement error in the observations is

37 For instance Appelate_l equals 4 if Appelate_l&f_O is 1 and Appelate_l&f_F is 0; 1 if the opposite holds true; 3 if both Appelate_l&f_O and Appelate_l&f_F equal 0; 2 if both dummies equal 1.
irrelevant (Rosenthal and Voeten, 2007). Yet, while several comparative law scholar emphasize *Appellate_l, Equity* and *Oral* (Merryman, 1969: pp. 52, pp.123-127; Glendon et al., 1982: pp. 127) as the institutions crucially differentiating the two traditions, others (Damaška, 1986; Zweigert and Kötz, 1998: pp. 271-272) stress that the defining feature of the procedure of each tradition is her adversarial or inquisitorial nature.

To the extent of solving the diatribe, I follow Rosenthal and Voeten (2007) and I use factor analysis tools in order to model the relationship among *Case, Appellate_l, Equity, Oral* and *Adversarial* and the latent “convergence toward common law” construct. Let $j = 1,\ldots,J$ denote countries and $i = 1,\ldots,I$ summarize the evolution of one of the characteristics of the legal system—*Case, Appellate_l, Equity, Oral* and *Adversarial*. Each observation $x_{i,j}$ is the value for country $j$ on aspect $i$ of its history of the legal system. If, as scholars in comparative law argue, the degree of convergence toward common law were the defining characteristic underlying these individual qualities, then we have $I$ imperfect indicators of a single unobserved variable. More generally, the $I$ indicators arise probabilistically from a set of $\tilde{I}$ unobserved fundamental historical aspects with $\tilde{I}$ smaller than $1$. The analysis is implemented mapping the observed matrix of indicators $X$ into an $I \times J$ matrix of latent variables $X^*$ with $x_{i,j}^*$ representing the log odds of country $j$ evolving in the characteristic and in the sense described by $i$.

Next, I can define the two-parameter multinomial logit item response model:

$$x_{i,j}^* = \beta_i + \eta_j \lambda_i,$$

where $-\beta_i$ represents the difficulty parameters—how unlikely the reform underlined by aspect $i$ is, the factor $\eta_j$ represent the latent construct and $\lambda_i$ the extent to which aspect $i$ discriminate between subjects evolving toward one or the other tradition. The factor loading for the *Case* is set in order to equal 1 in order to identify the scale of the factor.
Table 5 reports both coefficients and standard errors of the factor loadings and the likelihood ratio test that all these factor loadings do not differ significantly from 1 for a specification including or not the indicator *Oral*. The main message is that the latter does not discriminate between common and civil law countries while all the other indicators are all relevant and equally informative about the underlying “convergence toward common law” construct. This result confirms the key relevance of a preliminary study of the relative importance of each of the rule that sustain an institutional structure—in this case, the legal tradition.

Accordingly, I define the index *Convergence* as the sum of *Case*, *Appellate_1*, *Equity* and *Adversarial* and, next, the discrete indicator *Common* equal to 3 if *Convergence* is weakly greater than 13—at least three dummies equal four, equal to one if *Convergence* is weakly lower than seven—at least three indicators equal to one, and two otherwise. There are 60 observations equal to four, 60 equal to two, and 76 equal to one.

*Exogenous controls.*—An analysis in all similar to the precedent can be devised: this time, the dependent of (7) is *Common*. Panel A (B) of Table 6 lists the marginal effects when *Common* is controlled for the extent of cultural heterogeneity only (heterogeneity and democracy). A one-standard-deviation rise—0.32—in democracy implies a little more than a 24 percentage decrease in the likelihood of a reform from institutions typical of civil law to those typical of common law and more than a 33 percentage point increase in the likelihood of a reform in the opposite direction in column 4 of panel B. Cultural heterogeneity, instead, enters almost always with an unexpected sign. Yet, the relative coefficients are generally highly insignificant at the usual levels.

*Endogenous controls.*—This last pattern could be driven by a failure in the conditional independence assumption. Let me define, this time, *Common_Law* as the dummy equal to 1 when *Case* is equal to 3 and 0 otherwise. As columns 1 to 3 of Table 7 show, while
the evidence on Constraint_Executive remains robust; the one regarding the proxies for cultural heterogeneity becomes even more mixed. Both Genetic_Distance and Av_Elf enter with coefficients that are negative and significant. Yet, the magnitude of these coefficients is almost indistinguishable from zero and the sign of the one attached to Av_Elf becomes positive when Malaria or Economic_Power are considered. Therefore, the model’s failures in describing this feature of the convergence of the two traditions could be driven by the lack of relevant covariates proxying either for other unobserved determinants of cultural heterogeneity or for the society’s aversion to discretion. As the above model clarifies, the latter is responsible of the origin of hybrid systems.

All in all, it is reasonable to summarize the evidence saying that structure of the existing legal systems cannot be considered as randomly assigned but they are the result of long-lasting legal traditions and of the welfare maximizing choice of the society at whole.

V. Concluding Comments

This paper has moved some steps toward a theory of “endogenous legal systems” (see also Niblett et al. 2008; Guerriero, 2008). I have focused on a crucial aspect of the design of a legal system, namely the choice of those law-making and adjudication rules that all together can be classified as “civil” or “common” law traditions. Rather than reviewing my results, I highlight several avenues for further research in the field.

First, a relevant link exists (see Damaška, 1986) between law making institutions and the ex-post degree of insulation of political leaders (see Aghion, Alesina and Trebbi, 2004). To the latter extent, a key question to be answered is: how large the majority should be to pass laws given the degree of centralization of the law-making process in place in a country?
Second, the present paper leaves to the applied researchers the key task of rewriting the “legal origins” literature. In the lights of the actual analysis, the empirical literature on comparative legal and economic systems should not only consider all together the different institutions—\textit{i.e.}, law making and adjudication rules—characterizing the two traditions, but also take care of the endogeneity of such rules to the level of cultural heterogeneity and democracy in a society\textsuperscript{38}. These last two characteristics surely drive also the performances that we want to explain.

Finally, a crucial question related to the impact of the endogenous legal traditions on the efficiency of several—especially financial—economic activities is the one looking at the possible link between legal traditions and long-run economic growth. A recent piece of empirical research—Acemoglu and Johnson (2005)—has suggested that, contrary to the incomplete markets intuition (Coase, 1960), efficient contracting institutions, as driven by the “legal origins” of a country, have no first-order effect on the long-run economic growth, investment, and financial development. Again, the main insight of the present analysis is that this body of research should be revisited in order not only to consider the multifaceted design of a legal system but also to take into account in what respect the rules used by a society constitute an efficiency response to the economic and political environment and in what extent they are, instead, long-lasting transplantation constraints left in place by huge switching costs.

\textsuperscript{38} Also a clear matter of codification arises. Not only the La Porta et al.’s (1999) classification does not consider the 26 switches but three non Scandinavian countries are coded as having some form of civil law origin (three countries are assigned to the English law origin group) but case (statute) law was in use over the whole independence period. Also, in the section \textit{IV.C} sample, 92\% (68\%) of the countries coded by La Porta et al. (1999) as having an English (French) law origin have a value of \textit{Common} of one (three).
References


Appendix 1: Proofs of Lemmas and Propositions

Social optimally versus technologically efficient rules

It is straightforward to see that, for $\pi > 1$, $E(\hat{A}) \geq A'$ whenever:

$$(\frac{1}{2})(\bar{A} + A) \geq A' \leftrightarrow \Gamma(\lambda \pi) + \Gamma(\lambda / \pi) \geq 2 \Gamma(\lambda) \leftrightarrow$$

$$(\pi + \lambda)(\pi + \pi \lambda - 2 - 2 \pi \lambda) + (1 + \lambda \pi)(1 + \lambda) \geq 0 \leftrightarrow \lambda (\pi - 1)^2 (1 - \lambda) \geq 0,$$

which implies that $E(\hat{A}) \geq A'$ ($E(\hat{A}) < A'$) whenever $\lambda \leq 1$ ($\lambda > 1$). ■

Proof of Lemma 1

The distribution of judges’ preferences has expectation $E(\hat{A}) = (1/2)(\bar{A} + A)$ and variance $V(\hat{A}) = (1/2)(\bar{A} - E(\hat{A}))^2 + (1/2)(A - E(\hat{A}))^2 = (1/4)(\bar{A} - A)^2$. By the properties of the AR(1), case law converges to the distribution $\hat{A} = N(E(\hat{A}), V(\hat{A})(1+2K)^{-1})$. The comparative statics with respect $K$ holds by inspection while the one with respect $\pi$ is proved by the fact that $\frac{\partial V(\hat{A})}{\partial \pi} = (\lambda/2)(\bar{A} - A)\Gamma'(\lambda \pi) - \Gamma'(\lambda/\pi)(-1/\pi^2) > 0$. ■

Proof of Lemma 2.A and 2.B

From (2) and (3) it is immediate to see the objective function in (4) is strictly concave and the relative necessary and sufficient first order conditions are

$$(1 - \mu)(\hat{A}_l(l) - \bar{A}) + \mu [\|\hat{A}_l(l) - \sum_{i} \hat{A}_i]\] = 0.$$ (A1)

Then, the unique solution to (A1) is (6) which rewrites for every possible $l$ as

$\hat{A}_l(\emptyset) = \bar{A}$; $\hat{A}_l(L) = (1-\mu)\bar{A} + \mu A$;

$\hat{A}_l(H) = (1-\mu)\bar{A} + \mu A$; $\hat{A}_l(L, H) = [(1-\mu)\bar{A} + \mu(\bar{A} + A)](1+\mu)^{-1}$.

The equilibrium willingness to pay for each of the latter three coalitions is
\[
W_i(\hat{\lambda}_i(L), \hat{\lambda}) = \frac{1}{4}((\hat{\lambda} - \bar{\lambda})^2 - (1 - \mu)(\hat{\lambda} - \bar{\lambda})^2)^2 - \Psi = \frac{1}{4}(2 - \mu)(\hat{\lambda} - \bar{\lambda})^2 - \Psi = \frac{1}{16}\mu(2 - \mu)(\hat{\lambda} - \bar{\lambda})^2 - \Psi = W_i(\hat{\lambda}_i(H), \hat{\lambda}); \tag{A2}
\]

\[
W_i(\hat{\lambda}_i(L, H), \hat{\lambda}) = \frac{1}{4}\left((\hat{\lambda} - \bar{\lambda})^2 + (\hat{\lambda} - \bar{\lambda})^2 - \left((\hat{\lambda} - \bar{\lambda}) - \frac{\mu}{1 + \mu}(2\hat{\lambda} - \bar{\lambda} - \bar{\lambda})^2\right)^2 + \right.
\]
\[
\left. -\left((\hat{\lambda} - \bar{\lambda}) - \frac{\mu}{1 + \mu}(2\hat{\lambda} - \bar{\lambda} - \bar{\lambda})^2\right)^2\right)^2 - \Psi = \frac{\mu}{2(1 + \mu)}(2\hat{\lambda} - \bar{\lambda} - \bar{\lambda})^2 - \Psi.
\]

The legislator’s utility for each coalition \( l \) is

\[
U(\hat{\lambda}_i(\{\emptyset\}), \hat{\lambda}) = \frac{1 - \mu}{2}(\hat{\lambda} - \bar{\lambda})^2 = 0;
\]

\[
U(\hat{\lambda}_i(\{L\}), \hat{\lambda}) = (\mu^2/2)(\hat{\lambda} - \bar{\lambda})^2 - 2\mu\Psi = (\mu^2/8)(\bar{\lambda} - \bar{\lambda})^2 - 2\mu\Psi = U(\hat{\lambda}_i(\{H\}), \hat{\lambda}); \tag{A3}
\]

\[
U(\hat{\lambda}_i(\{L, H\}), \hat{\lambda}) = \frac{\mu^2}{2(1 + \mu)}(2\hat{\lambda} - \bar{\lambda} - \bar{\lambda})^2 - 2\mu\Psi = -2\mu\Psi.
\]

Being \( \hat{\lambda} = \bar{\lambda} - \bar{\lambda} \), coalition \( \{L, H\} \) will never pay the fixed cost \( \Psi > 0 \). The legislator is indifferent between \( \{L\} \) and \( \{H\} \) and both her utility in praising one of these two coalitions and the coalitions’ bribing schedules are strictly increasing with both \( \mu \) and \( \pi \), being \( \mu^2 \) and \( \mu(2 - \mu) \) increasing with \( \mu \) and from above \( \frac{\partial(\hat{\lambda} - \bar{\lambda})^2}{\partial\pi} > 0 \).

Therefore, for \( \Psi > 0 \) sufficiently small, it does always exist a level of \( \pi \) implicitly defined by \( \left(\mu/16\right)(2 - \mu)(\bar{\lambda}(\bar{\pi}) - \bar{\lambda}(\bar{\pi}))^2 = \Psi \), such that for \( \pi \geq \bar{\pi} \) both groups become organized and for \( \pi < \bar{\pi} \) they remain dormant. By the same token, it does exist a \( \pi \), implicitly defined by \( \left(\mu/16\right)(\bar{\lambda}(\bar{\pi}) - \bar{\lambda}(\bar{\pi}))^2 = \Psi \), such that for \( \pi \geq \bar{\pi} \) the legislator praises either \( \{L\} \) and \( \{H\} \), and for \( \pi < \bar{\pi} \) \( \hat{\lambda} \) prevails. At either \( \pi = \bar{\pi} \) or \( \pi = \bar{\pi} \), the quantity in (A2) is weakly (strictly) greater than that in (A3) for all \( \mu \ (\mu \in (0, 1)) \) being
\[
\frac{\mu(2-\mu)}{16}(A-A)^2 - \Psi \geq \frac{\mu^2}{8}(A-A)^2 - 2\mu \Psi \leftrightarrow \frac{\mu(2-3\mu)}{16}(A-A)^2 \geq (1-2\mu)\Psi = \frac{\mu(2-5\mu+2\mu^2)}{16}(A-A)^2.
\]

Thus, \( \bar{\pi} \geq \pi \). From \( t=1 \) on and for \( \pi \geq \bar{\pi} \), the loss of social welfare is \((\mu^2/8)(A-A)^2\) and increases with both \( \mu \) and \( \pi \). No group finds it worthy to pay a \( \Psi > 1/16 \). The following comparative statics complete the proof:

\[
(1/16)\mu(2-\mu)\frac{\partial}{\partial \pi}(\bar{A}(\pi)-A(\pi))^2 \ d\pi - d\Psi = 0 \rightarrow \frac{d\pi}{d\Psi} > 0;
\]

\[
(1/16)\mu(2-\mu)\frac{\partial}{\partial \pi}(\bar{A}(\pi)-A(\pi))^2 \ d\pi + (1/8)(1-\mu)(\bar{A}(\pi)-A(\pi))^2 \ d\mu = 0 \rightarrow \frac{d\pi}{d\mu} < 0;
\]

\[
(1/8)\mu^2 \frac{\partial}{\partial \pi}(\bar{A}(\pi)-A(\pi))^2 \ d\bar{\pi} - 2\mu d\Psi = 0 \rightarrow \frac{d\bar{\pi}}{d\Psi} > 0;
\]

\[
(1/8)\mu^2 \frac{\partial}{\partial \pi}(\bar{A}(\pi)-A(\pi))^2 \ d\bar{\pi} + (1/4)\mu(\bar{A}(\pi)-A(\pi))^2 \ d\mu = 0 \rightarrow \frac{d\bar{\pi}}{d\mu} < 0.
\]

\[\blacksquare\]

**Proof of Lemma 3**

The set of all possible coalition is \( \Omega = \{\emptyset, \{L\}, \{H\}, \{E\}, \{L,H\}, \{L,E\}, \{E,H\}, \{L,E,H\}\} \) and \( \bar{A} = (1/3)(\bar{A} + A + A') \). Therefore, for every possible \( l \), (6) rewrites as:

\[
\hat{A}_i(\{\emptyset\}) = \bar{A}, \quad \hat{A}_i(\{L\}) = (1-\mu)\bar{A} + \mu A; \quad \hat{A}_i(\{H\}) = (1-\mu)\bar{A} + \mu A; \quad \hat{A}_i(\{E\}) = (1-\mu)\bar{A} + \mu A';
\]

\[
\hat{A}_i(\{L\}, \{H\}) = \frac{(1-\mu)\bar{A} + \mu A}{1+\mu}; \quad \hat{A}_i(\{L\}, \{E\}) = \frac{(1-\mu)\bar{A} + \mu (A' + \bar{A})}{1+\mu}; \quad \hat{A}_i(\{L\}, \{E,H\}) = \frac{(1-\mu)\bar{A} + \mu (A + A')}{1+\mu}; \quad \hat{A}_i(\{L,E\}) = \frac{(1-\mu)\bar{A} + \mu (A + A' + A')}{1+2\mu}.
\]

The equilibrium willingness to pay for each of the seven possible coalitions is:

\[
W_i(\hat{A}_i(\{L\}), \bar{A}) = (1/6)\mu(2-\mu)(\bar{A} - \bar{A})^2 - \Psi = (1/6)\mu(2-\mu)\left[(A' + A - 2\bar{A})/3\right]^2 - \Psi; \quad (A4)
\]

\[
W_i(\hat{A}_i(\{H\}), \bar{A}) = (1/6)\mu(2-\mu)(\bar{A} - \bar{A})^2 - \Psi = (1/6)\mu(2-\mu)\left[(\bar{A} + A' - 2\bar{A})/3\right]^2 - \Psi; \quad (A5)
\]
\[ W_i(\hat{A}_i([E]), \bar{A}) = \frac{1}{6} \mu(2-\mu)(\hat{A} - A')^2 - \Psi; \quad W_i(\hat{A}_i([L,H]), \bar{A}) = \frac{\mu}{3(1+\mu)^2} \left( \frac{2\bar{A} - \bar{A} - \bar{A}}{3} \right)^2 - \Psi; \]

\[ W_i(\hat{A}_i([L,E]), \bar{A}) = \frac{\mu}{3(1+\mu)^2} \left( \frac{2\bar{A} - \bar{A} - \bar{A}}{3} \right)^2 - \Psi; \quad W_i(\hat{A}_i([E,H]), \bar{A}) = \frac{\mu}{3(1+\mu)^2} \left( \frac{2\bar{A} - \bar{A} - \bar{A}}{3} \right)^2 - \Psi; \]

\[ W_i(\hat{A}_i([L,E,H]), \bar{A}) = \frac{2\mu}{3(1+2\mu)} - \frac{3\mu^2}{3(1+2\mu)} \right) \left( \frac{3\bar{A} - \bar{A} - \bar{A}}{2} \right) - \Psi = -\Psi. \]

Finally the legislator’s utility for each coalition \( l \) is:

\[ U(\hat{A}_i([\phi]), \bar{A}) = 0; \quad U(\hat{A}_i([E]), \bar{A}) = \frac{\mu^2}{2} (\bar{A} - \bar{A})^2 - 3\mu\Psi; \]

\[ U(\hat{A}_i([L]), \bar{A}) = \frac{\mu^2}{2} (\bar{A} - \bar{A})^2 - 3\mu\Psi = \frac{\mu^2}{2} \left( \frac{A' + \bar{A} - 2\bar{A}}{3} \right)^2 - 3\mu\Psi; \quad (A6) \]

\[ U(\hat{A}_i([H]), \bar{A}) = \frac{\mu^2}{2} (\bar{A} - \bar{A})^2 - 3\mu\Psi = \frac{\mu^2}{2} \left( \frac{\bar{A} + A' - 2\bar{A}}{3} \right)^2 - 3\mu\Psi; \quad (A7) \]

\[ U(\hat{A}_i([L,H]), \bar{A}) = \frac{\mu^2}{18(1+\mu)} \left( 2\bar{A} - A' - \bar{A} \right)^2 - 3\mu\Psi; \quad U(\hat{A}_i([L,E]), \bar{A}) = \frac{\mu^2}{18(1+\mu)} \left( 2\bar{A} - A' - \bar{A} \right)^2 - 3\mu\Psi; \]

\[ U(\hat{A}_i([L,E,H]), \bar{A}) = \frac{\mu^2}{18(1+\mu)} \left( 2\bar{A} - A' - \bar{A} \right)^2 - 3\mu\Psi; \quad U(\hat{A}_i([L,H]), \bar{A}) = \frac{\mu^2}{18(1+\mu)} \left( 3\bar{A} - A' - \bar{A} \right)^2 - 3\mu\Psi. \]

Because, for every \( \pi, \bar{A} \leq A' \leq \bar{A} \) then \( \min(|\bar{A} - \bar{A}|, |\bar{A} - \bar{A}|) > |\bar{A} - A'| \) which implies that coalition \( \{E\} \) will never be chosen. Besides, coalition \( \{L,E,H\} \) will never pay a \( \Psi > 0 \).

Therefore, it remains to compare coalitions \( \{L\}, \{H\}, \{L,H\}, \{L,E\} \) and \( \{E,H\} \). The legislator prefers \( \{L\} \) to \( \{H\} \) when \( |A' + \bar{A} - 2\bar{A}| = 2\bar{A} - A' - \bar{A} > \bar{A} + A' - 2\bar{A} \) which is true for \( \lambda \leq 1 \), while the opposite is true for \( \lambda > 1 \). Notice also that \( 2\bar{A} - A' - \bar{A} = 2\bar{A} - A' - \bar{A} \), \( 2\bar{A} - A' - \bar{A} = A' + \bar{A} - 2\bar{A} \), \( 2\bar{A} - A' - \bar{A} = 2\bar{A} - A' - \bar{A} \) for \( \lambda > 1 \) and \( 2\bar{A} - A' - \bar{A} = \bar{A} + A' - 2\bar{A} \) for \( \lambda \leq 1 \). Thus, for \( \lambda \leq 1 \), \( \{E,H\} \) is preferred to \( \{L,H\} \) (being \( 2\bar{A} - A' - \bar{A} > \bar{A} + A' - 2\bar{A} \) always) and to \( \{L,E\} \) (because \( 2\bar{A} - A' - \bar{A} \geq \bar{A} + A' - 2\bar{A} \) for \( \lambda \leq 1 \)) but the legislator
weakly (strictly) prefers \( \{L\} \) for every \( \mu \geq 0 \) \( (\mu > 0) \). For \( \lambda > 1 \), \( \{L,E\} \) is preferred to \( \{E,H\} \) (because \( \bar{A} + A' - 2 \bar{A} > 2 \bar{A} - A' - A \) for \( \lambda > 1 \)) which, in turn, leads to a higher utility for the legislator than \( \{L,H\} \) (being \( 2 \bar{A} - A' - A > 2 \bar{A} - \bar{A} - \bar{A} \) always). Again, the legislator picks an extremist coalition because she weakly (strictly) prefers \( \{H\} \) for every \( \mu \geq 0 \) \( (\mu > 0) \). The legislator’s utility in praising \( \{L\} \) or \( \{H\} \) and the two coalitions’ bribing schedules are strictly increasing with both \( \mu \) and \( \pi \) being

\[
\frac{\partial (A' + A - 2 \bar{A})^2}{\partial \pi} = 2 \left( A' + A - 2 \bar{A} \right) \left[ \Gamma(\lambda/\pi)(1/\pi^2) - 2 \Gamma(\lambda/\pi) \pi \right] = \frac{\partial (A' + A - 2 \bar{A})^2}{\partial \pi} > 0.
\]

The quantity in \((A4)\) is weakly (strictly) greater than that in \((A6)\) for all \( \mu \) \((\mu \in (0,1))\)

\[
\frac{\mu(2 - \mu)}{6} \left( \frac{A' + A - 2 \bar{A}}{3} \right)^2 - \Psi \geq \frac{\mu^2}{2} \left( \frac{A' + A - 2 \bar{A}}{3} \right)^2 - 3 \mu \Psi \leftrightarrow
\]

\[
\frac{\mu(2 - 4 \mu)}{6} \left( \frac{A' + A - 2 \bar{A}}{3} \right)^2 \geq (1 - 3 \mu) \Psi = \frac{\mu(2 - 7 \mu + 3 \mu^2)}{16} \left( \frac{A' + A - 2 \bar{A}}{3} \right)^2.
\]

A similar relation exist between \((A5)\) and \((A7)\). For \( \lambda \leq 1 \) \((\lambda > 1)\), provided that \( \Psi > 0 \) is sufficiently small, it exists a \( \pi \) defined by \((1/54)\mu(2 - \mu) \left( A' + A(\bar{\pi}_i) - 2 \bar{A}(\bar{\pi}_i) \right)^2 = \Psi \), \((1/54)\mu(2 - \mu) \left( \bar{A}(\bar{\pi}_h) + A' - 2 \bar{A}(\bar{\pi}_h) \right)^2 = \Psi \) such that, for \( \pi \geq \bar{\pi}_i \), \( \{L\} \) \((\pi \geq \bar{\pi}_h \), \( \{H\} \) pays the fixed fee and for \( \pi < \bar{\pi}_i \) \((\pi < \bar{\pi}_h \) it remains dormant. Also, it does exist a \( \pi \) implicitly defined by \((1/54)\mu(2 - \mu) \left( A' + A(\bar{\pi}_i) - 2 \bar{A}(\bar{\pi}_i) \right)^2 = \Psi \), \((1/54)\mu(2 - \mu) \left( \bar{A}(\bar{\pi}_h) + A' - 2 \bar{A}(\bar{\pi}_h) \right)^2 = \Psi \), such that for \( \pi \geq \bar{\pi}_i \) \((\pi \geq \bar{\pi}_h \) the legislator praises \( \{L\} \) \((\{H\}) \) and, for \( \pi < \bar{\pi}_i \) \((\pi < \bar{\pi}_h \) \), \( \bar{A} \) prevails. For \( \pi \geq \bar{\pi}_i \) \( \geq \bar{\pi}_i \) and from \( t = 1 \) on, the loss of welfare is \( \mu^2 / 18 (A' + A - 2 \bar{A})^2 \) which increases with \( \pi \). No group finds it worthy to pay a \( \Psi > 1/27 \). The comparative statics on \( \bar{\pi}_i \) and \( \bar{\pi}_i \) are in all similar to those developed in the proof of lemma 3 and, consequently, they are omitted.
Proof of Proposition 2

For \( \lambda > 1 \) the legislator picks \( \{H\} \) and the expected losses of welfare under the two institutions are equal for \( \mu = \hat{\mu}_H \) with

\[
\hat{\mu}_H = \left[ \frac{1}{2(1+2K)} \right] \left[ 2 \left( 3\bar{A}^2 + 3(A')^2 + (\bar{A} + A' + \bar{A})^2 \right) \right] \geq 0. \tag{A8}
\]

First notice that the difference between the numerator and denominator is such that:

\[
6(\bar{A})^2 + 6(A')^2 - 2(\bar{A} + A' + \bar{A})^2 = 6(\bar{A})^2 - 3(\bar{A} + A' + \bar{A})^2 - 2\bar{A}(\bar{A} + A' + \bar{A}) + (A')^2 = 3(\bar{A} - A')^2 > 0,
\]

so that (A8) can be written as

\[
\hat{\mu}_H = \left[ \frac{1}{2(1+2K)} \right] \left[ 1 + 3(\bar{A} - A')^2 \left( \bar{A} + A' - 2\bar{A} \right)^2 \right].
\]

As a result, the right hand side of (A8) is decreasing with \( \pi \) because

\[
\frac{\partial \hat{\mu}}{\partial \pi} < 0 \iff \frac{\partial}{\partial \pi} \left( \frac{\bar{A} - A'}{\bar{A} + A' - 2\bar{A}} \right)^2 < 0 \iff \frac{\partial}{\partial \pi} \left( \frac{\bar{A}}{\bar{A} + A' - 2\bar{A}} \right) - \left( \frac{\partial}{\partial \pi} - 2 \frac{\partial}{\partial \pi} \right) (\bar{A} - A') < 0 \iff \frac{\lambda}{1 + \lambda} \left( \frac{\lambda + \lambda}{\pi + \lambda} \right) - \frac{\lambda}{(1 + \lambda)^2} \left( \frac{\lambda}{1 + \lambda} \right) < 0 \iff \pi + \lambda < 1 + \lambda \pi \iff \lambda > 1.
\]

To see why \( \hat{\mu}_H = 1 \) for \( K = 0 \), notice that the de l'Hôpital’s (1696) theorem applies to

\[
(\bar{A} - A')^2 (\bar{A} + A' - 2\bar{A})^2 \quad \text{and} \quad \lim_{x \to 1} (\bar{A} - A')^2 (A' + A - 2\bar{A})^2 = \lim_{x \to 1} (\partial \bar{A})/(\partial \pi) (\partial \bar{A}/\partial \pi - 2 \partial A/(\partial \pi))^2 = 1/3.
\]

So a sufficient condition such that statute law always prevails is that \( 1 + 2K \leq 1 \iff K = 0 \).

A similar analysis applies, by symmetry, to the \( \lambda \leq 1 \) case and is, therefore, omitted.

The expected losses of welfare are equal under the two institutions for \( \mu = \hat{\mu}_L \) with

\[
\hat{\mu}_L = \left[ \frac{1}{2(1+2K)} \right] \left[ 1 + 3(A' - A)^2 \left( A' + A - 2A \right)^2 \right] \geq 0.
\]

Accounting for Tastes versus Efficiency

The willingness to pay for the three coalitions are:

\[
W_i \left( \hat{\lambda}_L (\{L\}), \tilde{\lambda}^e \right) = (1/16) \mu (2 - \mu) (1 - T)^2 (\bar{A} - A)^2 - \Psi = W_i \left( \hat{\lambda}_L (\{H\}), \tilde{\lambda}^e \right),
\]
Straightforward algebra shows that the legislator is, again, indifferent between \( \{L\} \) and \( \{H\} \) and both her utility in praising one of these two coalitions and the coalitions’ bribing schedules are strictly increasing with both \( \mu \) and \( \pi \). For \( \Psi > 0 \) sufficiently small, it does exist a \( \pi^e \) implicitly defined by \((1/16)\mu(2-\mu)(1-T)^2(\bar{A}(\pi^e) - A(\pi^e))^2 = \Psi\), such that for \( \pi \geq \pi^e \) both groups become organized and for \( \pi < \pi^e \) they remain dormant and it does exist a level of \( \pi^e \) implicitly defined by \((1/16)\mu(1-T)^2(\bar{A}(\pi^e) - A(\pi^e))^2 = \Psi\), such that for \( \pi \geq \pi^e \) the legislator praises either \( \{L\} \) or \( \{H\} \) and, for \( \pi < \pi^e \), \( \bar{A} \) is chosen. From \( t=1 \) on and for \( \pi \geq \pi^e \geq \pi^e \), the loss equal \( \mu^2/(1-T)^2(\bar{A} - A)\) which rises with \( \pi \). No coalition pays a \( \Psi > (1-T)^2/16 \). The usual comparative statics apply.

**Proof of Lemma 4**

A type \( \{H\} \) judge, faced with a type \( \{L\} \) ’s precedent, distinguish to the first best when

\[
-\frac{\bar{K}}{\mu} \geq -(1/2)\left[ (A - A)^2 + A(1 - A) \right] \leftrightarrow 2\frac{\bar{K}}{\mu} \leq (A - A)^2 + A(1 - A). \tag{A9}
\]

Similarly, a type \( \{L\} \) judge faced with the precedent implemented by a type \( \{H\} \) judge, will distinguish to the first best whenever

\[
-\frac{\bar{K}}{\mu} \geq -(1/2)\left[ (A - A)^2 + A(1 - A) \right] \leftrightarrow 2\frac{\bar{K}}{\mu} \leq (A - A)^2 + A(1 - A). \tag{A10}
\]

The right hand side of \((A9)\) and \((A10)\) equal \( A'(1 - A') < 1 \) for \( \pi = 1 \) and 1 for \( \pi \rightarrow \infty \).

Next, notice that the right hand side of \((A9)\) is increasing with \( \pi \) whenever:

\[
\frac{\partial}{\partial \pi}\left[ (A - A)^2 + A(1 - A) \right] = 2\left( A - A \right)(\frac{\partial A}{\partial \pi} - \frac{\partial A}{\partial \pi}) + (1 - 2A)\frac{\partial A}{\partial \pi} = 2\left( A - A \right)\frac{\partial A}{\partial \pi} - (2A - 1)\frac{\partial A}{\partial \pi}
\]

is positive. This is always true if \( 2A > 1 \) or \( \lambda > 1/\pi \) but could be false for \( \lambda < 1 \).
Indeed, \( \partial \bar{A}/\partial \pi > - \partial A/\partial \pi \) when \( (\pi + \lambda)^2 > (1 + \lambda \pi)^2 \) or \( \pi + \lambda > 1 + \lambda \pi \leftrightarrow \lambda < 1 \). For \( \lambda < 1 \), a sufficient condition such that the right hand side of \((A9)\) is increasing with \( \pi \) is that \( 2\bar{A} - 2\bar{A} > 1 - 2\bar{A} \) which surely fails for \( \pi = 1 \) (being \( 2\bar{A} < 1 \) for \( \lambda < 1 \)) but it is true for every \( \pi > \hat{\pi} \) with \( 4\bar{A}(\hat{\pi}) - 2\bar{A}(\hat{\pi}) = 1 \). The right hand side of \((A10)\) increases with \( \pi \) if
\[
\frac{\partial}{\partial \pi} \left[ (\bar{A} - \bar{A})^2 + A(1 - \bar{A}) \right] = 2(\bar{A} - \bar{A}) \left( \frac{\partial \bar{A}}{\partial \pi} - \frac{\partial A}{\partial \pi} \right) + (1 - 2\bar{A}) \frac{\partial \bar{A}}{\partial \pi} = (2\bar{A} + 1) \frac{\partial \bar{A}}{\partial \pi} - 2(\bar{A} - \bar{A}) \frac{\partial A}{\partial \pi}
\]
is positive, which is always true if \( 1 > 2\bar{A} \) or \( \lambda < \pi \) but can be false for \( \lambda > 1 \). In the latter case, \( \partial \bar{A}/\partial \pi < - \partial A/\partial \pi \) so that a sufficient condition to have the right hand side of \((A10)\) increasing with \( \pi \) is that \( 2\bar{A} - 2\bar{A} > 2\bar{A} + 1 \) which fails for \( \pi = 1 \) (being \( 2\bar{A} > 1 \) for \( \lambda > 1 \)) but is the case for every \( \pi > \hat{\pi} \) with \( 2\bar{A}(\hat{\pi}) - 4\bar{A}(\hat{\pi}) = -1 \). So, for \( \hat{K} > \tilde{K} = A'(1 - A')/2 \), the fixed distinguish cost is too high at \( \pi = 1 \) and a fortiori for every \( \pi \leq \hat{\pi} = \min\{\hat{\pi}, \hat{\pi}\} \).

If, instead, \( \pi > \hat{\pi} \) it exists a \( \hat{\pi} \) such that every judge finds it worthy to introduce the second dimension for \( \pi \geq \hat{\pi} > \hat{\pi} \), and for every \( \hat{K} \in [\hat{K}, \tilde{K}] \) with \( \tilde{K} = 1/2 \).

\( \square \)

**Proof of Proposition 3**

First notice that the loss of welfare under case law for \( \pi < \hat{\pi} \) is always given by
\[
\bar{A}(\hat{\pi}) = (1/2)(\bar{A} - E(\hat{A}))^2 = (1/2)(A - E(\hat{A}))^2 = (1/2)(\bar{A} - \bar{A})^2 = (1/8)(\bar{A} - \bar{A})^2 \geq (1/8)(\bar{A} - \bar{A})^2, \forall \mu \in [0, 1].
\]

Therefore, whether or not \( \{L\} \) and \( \{H\} \) are organized statute law weakly prevails. For \( \pi > \hat{\pi} \) the first best is achieved, and case law is strictly better.

\( \square \)

**Proof of Lemma 5 and Proposition 4**

Whenever \( \hat{K} < 1/2 \) it always exist a \( \hat{\pi} \) implicitly defined by \((\bar{A}(\hat{\pi}) - A(\hat{\pi}))^2 = 2\hat{K} \) such that for \( \pi < \hat{\pi} \) a first period judge selects \( \hat{A}_{\pi} = \hat{A} \) and she is never reversed. In this case, the losses of welfare under case law are \((1/8)(\bar{A} - \bar{A})^2 \). As \( \pi \) grows over \( \hat{\pi} \) then the first
period judge can either fix her own preferred threshold—which will be overruled in the second period—or a compromise that is not overruled by a judge of type $\neg i$ but could be eventually be overruled by a judge of type $i$. This compromise is given by $\hat{A}_j = \bar{A} - \sqrt{2K}$ ($\hat{A}_j = \bar{A} + \sqrt{2K}$)—which is increasing in $\pi$ and defined by $(\bar{A}_j - \hat{A}_j)^2 = 2K$ ($\hat{A}_j - \bar{A}_j)^2 = 2K$)—if the first period judge is of type $H$ ($L$). A second period judge of the same type does not overrule it if $(\bar{A}_j - \hat{A}_j)^2 > (\hat{A}_j - \bar{A}_j)^2 \iff E(\hat{A}_j) > \hat{A}_j$ ($E(\hat{A}_j) < \hat{A}_j$). So, a first period type $H$ judge will prefer the compromise for $\hat{A}_j \leq \pi \leq \hat{\pi}$ (with $\hat{E}(\hat{A}_j) - \bar{A}_j = (1/4)(\pi - \bar{A}_j)^2 = 2K$) and $\delta(\bar{A}_j - \hat{A}_j)^2 > (1 + 2\delta)(\bar{A}_j - \hat{A}_j)^2$ which is true for $4\delta(\bar{A}_j - \hat{A}_j)^2 > (1 + 2\delta)(\bar{A}_j - \hat{A}_j)^2 \iff \delta > 1/2$. Instead, for $\pi > \hat{\pi}$—which can be the case only if $K < 1/8$—she always prefers the compromise and is overruled by a judge of the same type being $\delta(\bar{A}_j - \hat{A}_j)^2 > \delta(\bar{A}_j - \hat{A}_j)^2$. The analysis is in all similar when the first period judge is of type $L$. The loss of social welfare is $(1/8)(\bar{A}_j - \hat{A}_j)^2$ for $\pi < \hat{\pi}$, $(1/2)[(1/2)(\bar{A}_j - \hat{A}_j)^2 - \sqrt{2K}]$ for $\hat{\pi} \leq \pi \leq \hat{\pi}$ and it equals $(1/8)(\bar{A}_j - \hat{A}_j)^2 + (1/8)(\bar{A}_j - \hat{A}_j)^2 + (1/2)[(1/2)(\bar{A}_j - \hat{A}_j)^2 - \sqrt{2K}]$ for $\pi > \hat{\pi}$. The rest of the proposition is proved by inspection. ■

**Proof of Proposition 5**

Lower courts of type $H$ and $L$ will make use of discretion whenever:

$$E\left(\hat{A}_j(\hat{A}_j)\right) > E\left(\hat{A}_j(\hat{A}_j)\right) + \theta \iff (\hat{A}_j - \bar{A})^2 \geq (\hat{A}_j - \alpha - \bar{A})^2 + 2\theta \iff \hat{A}_j \leq \left(2\theta + \alpha^2\right)/2\alpha + \bar{A};$$

$$E\left(\hat{A}_j(\hat{A}_j)\right) > E\left(\hat{A}_j(\hat{A}_j)\right) + \theta \iff (\hat{A}_j - \bar{A})^2 \geq (\hat{A}_j + \alpha - \bar{A})^2 + 2\theta \iff \hat{A}_j \leq \left(-2\theta + \alpha^2\right)/2\alpha + \bar{A}.$$

As a result, of type $H$ ($L$) will use discretion only if $\hat{A}_j > E(\hat{A}_j)$ ($\hat{A}_j < E(\hat{A}_j)$) when $\theta \geq \bar{d}(\pi) = (\alpha/2)[(\bar{A}(\pi) - \bar{A}(\pi)) - \alpha]$ with $\bar{d}(\pi) \geq 0$ being $\alpha \leq (1/2)(\bar{A}_j - \bar{A})$. Furthermore, being $\alpha^2 \leq V(\hat{A}_j)$, the variance of case law is lower under the pure common law system if:
\[ V(\hat{\lambda})[(1+2K)+(\rho^2+2\rho+2\rho K)]^{-1}+(\alpha^2/2)<V(\hat{\lambda})|(1+2K)^{-1} \leftrightarrow \]

\[ \rho^2+2\rho+2\rho K > 1+2K >(1+2K)+\alpha^2(2V(\hat{\lambda})-\alpha^2)^{-1} \leftrightarrow \rho > \bar{\rho}(K). \]

The pure common law system is introduced whenever:

\[ V(\hat{\lambda})[(1+2K)+(\rho^2+2\rho+2\rho K)]^{-1}+(\alpha^2/4)+\Theta < V(\hat{\lambda})|(1+2K)^{-1} \leftrightarrow \]

\[ 0<\Theta \equiv V(\hat{\lambda})|(1+2K)^{-1} - V(\hat{\lambda})[(1+2K)+(\rho^2+2\rho+2\rho K)]^{-1}+(\alpha^2/4), \]

With \( \Theta > 0 \) whenever \( \rho > \bar{\rho}(K). \)
Appendix 2: Tables

<table>
<thead>
<tr>
<th>Table 1: Variables – Description and Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genetic_Distance</strong> =</td>
</tr>
<tr>
<td><strong>Av_Elf</strong> =</td>
</tr>
<tr>
<td><strong>Constraint_Executive</strong> =</td>
</tr>
<tr>
<td><strong>Time_Independence</strong> =</td>
</tr>
<tr>
<td><strong>Soil_Diversity</strong> =</td>
</tr>
<tr>
<td><strong>Pronoun_Drop</strong> =</td>
</tr>
<tr>
<td><strong>Second_Person</strong> =</td>
</tr>
<tr>
<td><strong>Latitude</strong> =</td>
</tr>
<tr>
<td><strong>Altitude_Diff</strong> =</td>
</tr>
<tr>
<td><strong>Malaria</strong> =</td>
</tr>
<tr>
<td><strong>Economic_Power</strong> =</td>
</tr>
</tbody>
</table>

* In the G-Econ dataset variables are reported for each grid-cell within countries. Each grid-cell corresponds to an area of 1 degree latitude times 1 degree longitude, which is approximately 100km by 100km.

** Dominant soil type (of a list of 27 types) are listed for each grid-cell.
### Table 2: Variables – Means and Standard Deviations

<table>
<thead>
<tr>
<th></th>
<th>Case Sample [94]</th>
<th>Common Sample [130]</th>
<th>Case Sample [94]</th>
<th>Common Sample [130]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic_Distance</td>
<td>0.504</td>
<td>0.442</td>
<td>0.288</td>
<td>0.286</td>
</tr>
<tr>
<td></td>
<td>[0.288]</td>
<td>[0.286]</td>
<td>[0.094]</td>
<td>[0.094]</td>
</tr>
<tr>
<td>Av_Elf</td>
<td>0.358</td>
<td>0.320</td>
<td>0.295</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td>[0.295]</td>
<td>[0.286]</td>
<td>[0.148]</td>
<td>[0.148]</td>
</tr>
<tr>
<td>Constraint_Executive</td>
<td>0.420</td>
<td>0.488</td>
<td>0.177</td>
<td>0.187</td>
</tr>
<tr>
<td></td>
<td>[0.215]</td>
<td>[0.215]</td>
<td>[0.090]</td>
<td>[0.090]</td>
</tr>
<tr>
<td>Veg_Div</td>
<td>0.787</td>
<td>0.783</td>
<td>1.203</td>
<td>1.345</td>
</tr>
<tr>
<td></td>
<td>[0.215]</td>
<td>[0.215]</td>
<td>[0.134]</td>
<td>[0.134]</td>
</tr>
<tr>
<td>Pronoun_Drop</td>
<td>0.596</td>
<td>0.662</td>
<td>0.475</td>
<td>0.475</td>
</tr>
<tr>
<td></td>
<td>[0.475]</td>
<td>[0.475]</td>
<td>[0.345]</td>
<td>[0.345]</td>
</tr>
<tr>
<td>Second_Person</td>
<td>0.723</td>
<td>0.692</td>
<td>1.573</td>
<td>1.496</td>
</tr>
<tr>
<td></td>
<td>[0.463]</td>
<td>[0.463]</td>
<td>[1.709]</td>
<td>[1.709]</td>
</tr>
</tbody>
</table>

### Table 3: Statute Versus Case Law – Exogenous Cultural Heterogeneity and Democracy

#### PANEL A

<table>
<thead>
<tr>
<th></th>
<th>Reform toward Case Law</th>
<th>Reform toward Statute Law</th>
<th>Reform toward Case Law</th>
<th>Reform toward Statute Law</th>
<th>Reform toward Case Law</th>
<th>Reform toward Statute Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic_Distance</td>
<td>0.153</td>
<td>– 0.185</td>
<td>0.186</td>
<td>– 0.153</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.052]**</td>
<td>[0.079]**</td>
<td>[0.078]**</td>
<td>[0.075]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av_Elf</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0007</td>
<td>0.0006</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0002]**</td>
<td>[0.0004]**</td>
<td>[0.0003]**</td>
<td>[0.0004]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time_Independence</td>
<td>0.0109</td>
<td>– 0.197</td>
<td>0.051</td>
<td>– 0.167</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.087]</td>
<td>[0.090]**</td>
<td>[0.072]</td>
<td>[0.083]**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### PANEL B

<table>
<thead>
<tr>
<th></th>
<th>Reform toward Case Law</th>
<th>Reform toward Statute Law</th>
<th>Reform toward Case Law</th>
<th>Reform toward Statute Law</th>
<th>Reform toward Case Law</th>
<th>Reform toward Statute Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic_Distance</td>
<td>0.222</td>
<td>– 0.263</td>
<td>0.195</td>
<td>– 0.246</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.090]**</td>
<td>[0.113]**</td>
<td>[0.089]**</td>
<td>[0.108]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Av_Elf</td>
<td>0.145</td>
<td>– 0.165</td>
<td>0.105</td>
<td>– 0.098</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.112]</td>
<td>[0.096]*</td>
<td>[0.100]</td>
<td>[0.088]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constraint_Executive</td>
<td>– 0.253</td>
<td>– 0.371</td>
<td>– 0.157</td>
<td>– 0.377</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.110]**</td>
<td>[0.114]**</td>
<td>[0.087]*</td>
<td>[0.122]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time_Independence</td>
<td>0.145</td>
<td>0.165</td>
<td>0.227</td>
<td>0.240</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.096]*</td>
<td>[0.100]</td>
<td>[0.100]</td>
<td>[0.088]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Notes: 1. Robust standard errors (z distribution) in parentheses; 2. *** denotes significant at the 1% confidence level; **, 5%; *, 10%.
### Table 4: Statute Versus Case Law – Endogenous Cultural Heterogeneity and Democracy

<table>
<thead>
<tr>
<th></th>
<th>Case Law</th>
<th>Case Law</th>
<th>Case Law</th>
<th>Case Law</th>
<th>Case Law</th>
<th>Case Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic_</td>
<td>0.064</td>
<td>– 0.312</td>
<td>0.148</td>
<td>0.099</td>
<td>0.102</td>
<td>0.123</td>
</tr>
<tr>
<td>Distance</td>
<td>[0.175]</td>
<td>[0.408]</td>
<td>[0.015]***</td>
<td>[0.0001]**</td>
<td>[0.002]**</td>
<td>[0.003]**</td>
</tr>
<tr>
<td>Av_Elf</td>
<td>0.844</td>
<td>0.188</td>
<td>0.093</td>
<td>0.040</td>
<td>0.161</td>
<td>0.087</td>
</tr>
<tr>
<td>Constraint_</td>
<td>– 0.313</td>
<td>– 1.511</td>
<td>– 0.167</td>
<td>– 0.149</td>
<td>– 0.185</td>
<td>– 0.155</td>
</tr>
<tr>
<td>Executive</td>
<td>[0.186]*</td>
<td>[0.518]***</td>
<td>[0.017]***</td>
<td>[0.0002]**</td>
<td>[0.002]**</td>
<td>[0.010]**</td>
</tr>
<tr>
<td>Latitude</td>
<td>– 0.169</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altitude_Diff</td>
<td>[0.0065]</td>
<td>[0.066]</td>
<td>[0.059]***</td>
<td>[0.066]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaria</td>
<td>– 0.010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economic_</td>
<td>0.034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.0000]**</td>
<td></td>
</tr>
</tbody>
</table>

#### First Stage for Av_Elf

|                        | 0.282    | 0.293    | 0.385    | 0.238    | 0.305    |
| Diversity              | [0.141]**| [0.145]**| [0.146]**| [0.103]**| [0.148]**|
| Pronoun_               | – 0.031  | – 0.069  | 0.059    | – 0.046  |
| Drop                   | [0.061]  | [0.066]  | [0.055]  | [0.062]  |
| Second_                | – 0.077  | – 0.093  | – 0.170  | – 0.078  |
| Person                 | [0.065]  | [0.066]  | [0.059]***| [0.066]  |
| R² in First Stage      | 0.44     | 0.45     | 0.51     | 0.57     | 0.45     |

#### First Stage for Constraint_Executive

|                        | 0.136    | 0.131    | 0.155    | 0.116    |
| Diversity              | [0.140]  | [0.160]  | [0.137]  | [0.139]  |
| Pronoun_               | – 0.092  | – 0.096  | – 0.128  | – 0.072  |
| Drop                   | [0.052]* | [0.051]* | [0.054]* | [0.052]**| [0.054]**|
| Second_                | – 0.247  | – 0.249  | – 0.217  | – 0.247  |
| Person                 | [0.062]*** | [0.058]***| [0.058]***| [0.058]***| [0.061]***|
| R² in First Stage      | 0.52     | 0.52     | 0.53     | 0.53     |

Notes: 1. Robust standard errors (clustered at the country level in the upper panel of columns 4 to 6) in parentheses; 2. Control variables used in the second stage are also included in the first stage but not reported to save space; 3. *** denotes significant at the 1% confidence level; **, 5%; *, 10%.

### Table 5: Building the Convergence Measure

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Appelate</th>
<th>Equity</th>
<th>Adversarial</th>
<th>Oral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor Loadings</td>
<td>1</td>
<td>0.746</td>
<td>0.548</td>
<td>0.691</td>
<td>0.109</td>
</tr>
<tr>
<td>(fixed)</td>
<td>[0.403]*</td>
<td>[0.262]**</td>
<td>[0.310]**</td>
<td>[0.085]</td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio Test</td>
<td>38.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4 degrees of freedom)</td>
<td>(0.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Units</td>
<td>980 (Level 1)</td>
<td>196 (Level 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor Loadings</td>
<td>1</td>
<td>0.758</td>
<td>0.543</td>
<td>0.701</td>
<td></td>
</tr>
<tr>
<td>(fixed)</td>
<td>[0.409]*</td>
<td>[0.254]**</td>
<td>[0.317]**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio Test</td>
<td>3.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3 degrees of freedom)</td>
<td>(0.36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Units</td>
<td>784 (Level 1)</td>
<td>196 (Level 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1. Standard errors clustered at the country level in parentheses; 2. *** denotes significant at the 1% confidence level; **, 5%; *, 10%. 3. Likelihood ratio test for the null hypothesis that all the factor loadings do not differ from 1 (p-value in parenthesis).
### Table 6: Civil Versus Common Law – Exogenous Cultural Heterogeneity and Democracy

**Panel A**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reform toward</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Law</td>
<td>0.160</td>
<td>0.430</td>
<td>0.122</td>
</tr>
<tr>
<td>Civil Law</td>
<td>[0.161]</td>
<td>[0.212]**</td>
<td>[0.196]</td>
</tr>
<tr>
<td>Reform toward</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Law</td>
<td>0.217</td>
<td>0.159</td>
<td>0.183</td>
</tr>
<tr>
<td>Civil Law</td>
<td>[0.227]</td>
<td>[0.226]</td>
<td>[0.256]</td>
</tr>
</tbody>
</table>

**Estimation**

Multinomial Logit

**Other Controls**

Population

**Notes:** 1. Robust standard errors (z distribution) in parentheses; 2. *** denotes significant at the 1% confidence level; **, 5%; *, 10%.

### Table 7: Civil Versus Common Law – Endogenous Heterogeneity and Democracy

**Panel B**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reform toward</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Law</td>
<td>0.027</td>
<td>0.303</td>
<td>– 0.086</td>
<td>0.360</td>
</tr>
<tr>
<td>Civil Law</td>
<td>[0.196]</td>
<td>[0.214]</td>
<td>[0.242]</td>
<td>[0.263]</td>
</tr>
<tr>
<td>Reform toward</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Law</td>
<td>0.025</td>
<td>0.244</td>
<td>0.059</td>
<td>0.135</td>
</tr>
<tr>
<td>Civil Law</td>
<td>[0.309]</td>
<td>[0.291]</td>
<td>[0.350]</td>
<td>[0.321]</td>
</tr>
<tr>
<td>Constraint, Executive</td>
<td>– 0.478</td>
<td>0.551</td>
<td>– 0.466</td>
<td>0.567</td>
</tr>
<tr>
<td></td>
<td>[0.222]**</td>
<td>[0.286]*</td>
<td>[0.220]**</td>
<td>[0.285]**</td>
</tr>
</tbody>
</table>

**Estimation**

Multinomial Logit

**Other Controls**

Time_Independence, GDP, Population

**Notes:** 1. Robust standard errors (z distribution) in parentheses; 2. *** denotes significant at the 1% confidence level; **, 5%; *, 10%.