More than Politics: Ability and Ideology in the British Appellate Committee

Matias Iaryczower and Gabriel Katz[*]

Abstract

We argue that a model of judicial behavior that accounts for differences in justices' ability and ideology provides a fruitful alternative for the empirical analysis of judicial decision-making around the world, and illustrate this by focusing on the case of the UK. We show that the model explains the decisions of the Lords of Appeal remarkably well, and improves the fit of a purely ideological model. We use our estimates to tackle previously unaddressed questions about the relative role of justices' preferences and ability in the Appellate Committee.

1 Introduction

The spatial voting model is at the core of a large number of empirical studies of voting behavior in legislatures around the world. In the last decade, it has also been used extensively to analyze the voting decisions of members of the US Supreme Court. Progress in this front inexorably led to the application of the spatial model to understand judicial decision-making in other latitudes. In the last few years, judges from Argentina to Canada and from Brazil to Australia have been classified as liberals and conservatives, extremists or moderates.[1]

This progress notwithstanding, the straight application of the ideological model to decision-making in the court is not universally accepted. The spatial voting model puts forth an unwavering representation of legal realism (Segal and Spaeth, 2002) in which judges decide cases exclusively based on their preferences, without attention to the context of the law. On the other side of the spectrum, the classical view of legal scholars is that "judging [is] more like finding than making".[2] While judges' ideology can shape their opinions to some extent, decisions in the court are primarily about how the facts of the case fit into the body of the law and established legal reasoning. These alternative perspectives to what judging is about meet head on when we think of desiderata for a member of the court. Ideally, Gerhardt (2004) argues,

> "[W]e would want to make sure that the [judge] has very sound legal skills; asks intelligent, probing questions; thinks clearly if not imaginatively about legal problems; identifies legal issues in a wide range of problems, is trained at problem-solving, and understands the special duties that she will be called upon to discharge." … "The ideal temperament for a justice is presumably to have the capacity to make decisions even-handedly [and] to be open-minded in listening to and considering the arguments in the cases that come before him."

The ideological model of judicial behavior is ill-equipped to capture several of these key aspects: should a judge who is better at "problem-solving" be placed to the right or the left of the political spectrum? Would we regard a judge who excels at "identifying legal issues in a wide range of problems" as a moderate or an extremist? There is simply no way to answer these questions in a sensible manner, because the ideological voting

---

[1]See for example Alarie and Green (2007), Desposato, Ingram and Lannes (2012), and Gonzalez Bertomeu, Dalla Pellegrina and Garoupa (2013).

[2]Rogat (1972), cited in Segal and Spaeth (2002), p. 87.

model precludes their examination. This, of course, is not necessarily a problem per se; understanding complex phenomena often requires substantial simplification. The relevant question is whether the simplifications we accept in a particular model entail a fundamental sacrifice in our understanding of the decision-making process in the court.

This paper considers this issue in the context of the British House of Lords. Besides being interesting in its own right, this case is also particularly relevant for our purposes in view of recent work by Hanretty (2013). In this paper, Hanretty shows that an item response theory (IRT) model does not fit the data better than a null model in which judges vote with the majority with a probability equal to the frequency with which they do this in the sample, and concludes from this that "policy-sensitive models of judicial behavior, whether attitudinal or strategic" are not useful to understand the voting behavior of the Law Lords.

We believe that the comparison of a model's fit against a mechanical (i.e., a-theoretical or non-behavioral) null hypothesis should not be the primal factor in deciding whether to discard the ideological voting model, or any model of judicial behavior for that matter. The frequency of cases in which a judge votes with the majority in a given sample can be effective to predict individual votes in the sample at hand, but tells us little of interest about the behavior of judges. Thus, if we dismiss the ideological model, we remain empty handed (see Achen (2002) for a similar point). We argue that a more productive approach is to incorporate the criticisms raised against a theory into an alternative behavioral model, and then contrast the empirical and substantive implications of the contending explanations. We pursue this approach here, comparing the ideological model with what we are labeling the *learning* model of judicial behavior, developed by Iaryczower and Shum (2012) in the context of the US Supreme Court.

The learning model introduces a tradeoff between judges' ideology and the information supporting one decision over the other in a given case on the grounds of legal reasoning.[3] Unlike in the ideological voting model, individual preferences do not fully determine judicial decisions, but establish an informational hurdle for judges to rule in a certain direction - e.g., liberal or conservative; in favor or against overturning the decision of the lower courts. A judge leaning in one direction will vote against his bias if the information based on the facts and on how the law applies to the case under consideration surpasses the threshold

---

[3]By ideology, we mean the system of ideas and manner of thinking characteristic of each individual (*New Oxford American Dictionary*, 2013). This system of ideas *can* be related to political ideology, but it can also be molded by personal experiences, judicial philosophies, and other factors that are unrelated to an alignment in terms of conservative and liberal values.

imposed by his preferences. The learning model therefore combines elements of both the "legal" and "attitudinal" models of judicial behavior (Segal and Spaeth, 2002).

We begin by evaluating the efficacy of the ideological and learning models to account for judges' decisions in the Appellate Committee of the House of Lords. Our estimates indicate that the purely ideological model of judicial behavior misses some relevant aspects of decision-making in the court, and that (a simple version of) the learning model outperforms (a simple version of) the spatial model according to a variety of goodness-of-fit measures commonly employed in the literature.

The learning model of judicial behavior also allows us to distinguish the impact of preferences and ability on justices' decisions. We can then quantify the discrepancies between the traits of real justices in the Appellate Committee and Gerhardt (2004)'s ideal justices.

First, justices can differ in their "capacity to make decisions even-handedly", as they might be biased to vote in a conservative or liberal direction or be more or less prone to reverse lower court rulings. In fact, we find that Law Lords tend to be quite moderate overall, in the sense that individual preferences only impose a low hurdle that the case-specific information must surpass for judges to allow or dismiss appeals. The median justice in the committee votes to uphold or overturn lower court decisions based essentially on the facts of the case alone. Furthermore, our results suggest that partisanship is only imperfectly correlated with judges' preferences. Rather than being simply shaped by political or ideological views, Lords' biases seem to reflect general attitudinal differences that cannot be easily ordered along a conservative-liberal dimension.

Second, justices might differ in their ability at "problem-solving" or, more generally, in their capacity to map the law to the specifics of the case under consideration. We find that there are indeed considerable differences in ability across justices. Moreover, these differences are correlated with their judicial experience, and vary substantially across areas of the law.

Finally, judges might differ from the ideal in terms of how open minded they can be "in listening to and considering the arguments in the cases that come before them". To capture how much information specific to the case can alter the preconceptions based on priors and bias, we compute the probability that each justice votes differently from what he would have voted in the absence of case-specific information. We show that the importance of case-specific information in the Appellate Committee has increased consistently between 1969 and 2002, leading to a court more prone to considering the facts and arguments of the appeal in its rulings. This finding is consistent with prior work underscoring the gradual

3

decline of political considerations and the heightened influence of professional merit in the judicial selection process (Bingham, 2009; Malleson, 2009).

The rest of the paper is organized as follows. Section 2 briefly reviews the relevant literature. Section 3 describes the learning model of judicial decision-making and relates it to the more familiar spatial voting model. Section 4 describes the data used in our empirical analysis, and Section 5 presents and discusses our main findings. Finally, Section 6 concludes. Additional results are reported in the Supplementary Materials Appendix accompanying this paper.

2   Relation with the Literature

Our paper builds on the contributions of several strands of literature. Our estimation of the ideological model of judicial behavior in the Appellate Committee follows a large body of empirical research - mostly centered on the US Supreme Court - aimed at estimating judges' preferences and assessing its role in the judicial decision-making process. One group of papers, beginning with Segal and Cover (1989) and Segal et al. (1995), used external sources of information (e.g., newspaper editorials) to construct measures of judges' policy preferences and contrasted them against observed voting patterns. More recently, the literature turned to recover justices' ideology directly from their voting decisions (Martin and Quinn (2002, 2007)) by estimating the spatial voting model commonly used in the analysis of legislatures (Poole and Rosenthal, 1985; Clinton, Jackman and Rivers, 2004).

Our specification of the learning model of judicial behavior follows Iaryczower and Shum (2012). This paper develops an empirical framework to estimate a model that allows for a common value component and incomplete information. The resulting model extends the purely ideological model by taking into consideration the quality of the case-specific information available to justices and their ability to interpret and incorporate this information into their opinions.[4]

In Britain, there is a vast literature on the political and legal aspects of the House of Lords' judicial role (Stevens, 1978; Paterson, 1982; Blom-Cooper, Dickson and Drewry, 2009; Paterson, 2013). Quantitative studies on the decisions of the Law Lords, however, have been relatively scarce. Robertson (1982) was one of the earliest papers to apply

---

[4]The model was also applied by Iaryczower, Lewis and Shum (2013) to explore the impact of alternative selection and retention methods on the performance of state supreme court justices. Iaryczower, Katz and Saiegh (2013) implements a variation of this approach to model strategic voting in the US Congress. See Feddersen and Pesendorfer (1997) and Duggan and Martinelli (2001) for theoretical foundations.

statistical methods in order to investigate the behavior of the Lords of Appeal in Ordinary. Using multidimensional scaling to examine non-unanimous votes between 1965 and 1978, Robertson concluded that judges' decisions were to a large extent discretionary and strongly influenced by their "legal" or "professional" ideology. Robertson (1998) further developed the idea. Similar results regarding the prevalence of substantive disagreements in Law Lords' judicial positions were reported by Arvind and Stirton (2012), who examined cases brought against state bodies following the Human Rights Act 1998 using Bayesian techniques. These studies generally adhere to the classical view of legal scholars that such differences are largely unrelated to justices' *politics*.[5] This is also the key finding of Hanretty (2013)'s paper, which constitutes the first attempt to analyze the decisions of the UK judiciary within the spatial voting framework. None of these studies, however, explicitly account for the interaction between idiosyncratic preferences and case-specific information in their empirical analyses of judicial decision-making.

3   Models of Decision-Making in the Court

In this section we describe the spatial voting model and the learning model of judicial behavior. While the spatial model is well known in the literature, it is useful to briefly review it in order to clarify both the interpretation of the results and the differences and similarities between the two models.

Common to both models is the decision-making environment. A panel of n justices, $i = 1, \ldots, n$, makes decisions on T independent cases. In each case $t = 1, \ldots, T$, justices face a decision between two alternatives, x_t^0 and x_t^1. Justice i's decision is coded as $v_{it} = 1$ if he votes in favor of x_t^1 and $v_{it} = 0$ if he votes in favor of x_t^0.[6] The court aggregates the decisions of the individual justices by simple majority rule; i.e. $v_t = 1$ if $\sum_i v_{it} \geq \frac{n+1}{2}$ and $v_t = 0$ otherwise. The two models, however, differ in their assumptions regarding the preferences and information of members of the court.

The spatial voting model assumes that judges are perfectly informed about the characteristics of the alternatives under consideration. Judges have euclidean preferences, and alternatives can be represented by points in an n-dimensional euclidean space, $X = R^n$.

---

[5]This view is contested by Griffith (1977), but his work remains an exception within the academic and popular literature.

[6]This coding is not uniquely determined. For example, we could code votes according to whether justices vote to overturn or uphold the decision of the lower courts, or according to whether judges vote for the "liberal" or "conservative" alternative.

Each justice i has an ideal point $z_i \in X$ and, for any two alternatives $x^0, x^1 \in X$, prefers x^0 to x^1 if and only if x^0 is closer to z_i than x^1.[7] When the model is taken to the data, it is standard to assume that the members of the court have quadratic preferences over a single-dimensional policy space. In addition, in each case t, judges receive an individual and alternative-specific additive shock to utility. Formally, when justice i votes for alternative x, he obtains a payoff $U_i(x) = -\frac{1}{2}(z_i - x)^2 + \varepsilon_{ix}$, where ε_{ix} is assigned a certain (e.g., normal) distribution. Then $v_{it} = 1$ if and only if $U_i(x_t^1) > U_i(x_t^0)$, or equivalently, letting $\eta_{it} \equiv \varepsilon_{i,x_t^1} - \varepsilon_{i,x_t^0} \sim N(0, \varsigma_t^2)$, if and only if $\eta_{it} > (x_t^1 - x_t^0)\left[\frac{x_t^1 + x_t^0}{2} - z_i\right]$. Thus, letting $\lambda_t \equiv \frac{x_t^1 - x_t^0}{\varsigma_t}$ and $\kappa_t \equiv (\frac{x_t^1 + x_t^0}{2})$, the probability of $v_{it} = 1$ is $\Phi(\lambda_t[z_i - \kappa_t])$.

We can then estimate ideal points z_i and midpoints κ_t relative to each other from the likelihood function

$$\Pr(\mathbf{v_t}) \equiv \prod_{i=1}^{n} \Phi(\lambda_t[z_i - \kappa_t])^{v_{it}}[1 - \Phi(\lambda_t[z_i - \kappa_t])]^{1-v_{it}} \tag{1}$$

Note that (1) is a simple two-parameter item response model. As such, one could directly use (1) for estimation without "buying into" the spatial voting model itself. It should be clear, though, that if no commitment is made to the underlying theoretical model, the resulting estimates cannot be interpreted as providing ideal points and midpoints in the space of alternatives.

A criticism raised against the application of the spatial model to judicial settings is that a model based solely on ideology misses important elements of the decision-making process in the court. In this view, judges do not necessarily decide "on the basis of their political positions, but rather on the basis of shared views of what the law requires" (Hanretty, 2013). In fact, incorporating this element does not preclude judges' preferences or ideology to come into play. As Iaryczower and Shum (2012) argue, uncertainty in the meaning of the law allows justices to differ in their opinions about a case. With anything less than complete certainty, opinions can vary across justices because of idiosyncratic thresholds of proof brought by individual (e.g., ideological) biases or due to discrepancies in their ability to evaluate the available information in different contexts (associated, for instance, to differences in judges' training or expertise).

This logic is captured in the learning model of judicial behavior, which introduces a tradeoff between justices' preferences and information. Judges are assumed to be imper-

---

[7]In this standard formulation, the ideal point is treated as constant over time. Martin and Quinn (2002, 2007) allow judges' ideal points to change in each period.

fectly informed about the correct ruling in each case, according to the law. This imperfect information can be due to the effect of time constraints, the complexity of the law, or the intricacy of the case. We represent this ideal ruling by a latent or unobservable variable $\omega_t \in \{0, 1\}$ indicating whether the case-specific information favors the alternative x_t^1 (if $\omega_t = 1$) or x_t^0 (if $\omega_t = 0$). Justices' prior belief that $\omega = 1$ is denoted by ρ_t. In addition, before ruling in each case t, each justice i observes a private signal $s_{it} = \omega_t + \sigma_{it}\varepsilon_t$, where $\varepsilon_t \sim \mathcal{N}(0, 1)$. This private signal represents the case-specific information that is salient for each judge before making his decision. The scale parameter $\theta_{it} = 1/\sigma_{it}$ captures the informativeness of i's signal, with higher values indicating that i is more certain of the directionality of the outcome.

Judges' payoffs depend on the realization of ω_t. In particular, given $\pi_{it} \in (0, 1)$, justice i has a payoff of $-\pi_{it}$ if his decision is $v_{it} = 1$ (in favor of x_t^1) when $\omega_t = 0$, and of $-(1 - \pi_{it})$ if he incorrectly votes $v_{it} = 0$ (in favor of x_t^0) when $\omega_t = 1$. The payoffs of $v_{it} = \omega_t = 0$ and $v_{it} = \omega_t = 1$ are normalized to zero. Thus, given information s_{it}, individual i votes $v_{it} = 1$ if and only if $\Pr(\omega_t = 1|s_{it}) \geq \pi_{it}$.

Let s_{it}^* denote the value of s_{it} that solves the equation $\Pr(\omega_t = 1|s_{it}) = \pi_{it}$ or, equivalently,

$$\frac{\phi(\theta_{it}[s_{it} - 1])}{\phi(\theta_{it}s_{it})} = \frac{\pi_{it}}{1 - \pi_{it}}\frac{1 - \rho_t}{\rho_t} \tag{2}$$

At s_{it}^*, judge i is indifferent between voting in one or other direction. Because the likelihood ratio $L(s) \equiv \Pr(s|\omega_t = 1)/\Pr(s|\omega_t = 0)$ is increasing in s, higher signals make the state $\omega = 1$ more likely.[8] As a result, the judge follows a cutoff strategy: i votes $v_{it} = 1$ if $s_{it} \geq s_{it}^*$ and $v_{it} = 0$ otherwise. The cutpoint s_{it}^* therefore completely characterizes individual behavior, and we can write the likelihood of justices' votes in case t as

$$\Pr(\mathbf{v_t}) \equiv \sum_{\omega_t} \Pr(\omega_t) \prod_{i=1}^{n}[1 - \Phi(\theta_{it}[s_{it}^* - \omega_t])]^{v_{it}}\Phi(\theta_{it}[s_{it}^* - \omega_t])^{1 - v_{it}} \tag{3}$$

Based on (3), we can estimate (θ_{it}, s_{it}^*) for all justices and recover the parameters π_{it} reflecting their preferences or bias. We also consider a strategic version of the learning model in which justices care about the decision of the court - and thus, about the votes of the other members of the committee hearing the appeal. Estimation of the model with outcome-oriented judges can be carried out as for the benchmark specification. In fact,

---

[8]It should be emphasized that the model does not assume that judges are more likely to make correct than incorrect decisions, but only that the signals they receive are informative, in the sense that higher signals are more likely when the state is 1 than when the state is 0.

because the equilibrium of the outcome-oriented model is still in cutoff strategies, the estimates of the voting strategies s_{it}^* and of the precision of judges' information θ_{it} are unchanged with respect to the baseline specification. The fit of the model is also unaffected. The difference is entirely in the estimate of the preference parameters π_{it}.

Although in principle the ideological and learning models might seem completely unrelated, this is not necessarily the case. To make this connection clear, consider again the case in which judges have quadratic utility functions over a one-dimensional policy space. Suppose that in addition to the ideological dimension, justices obtain a payoff $g(\omega_t, x_t)$, which takes the value ς_t if their decision is appropriate on a legal basis (i.e., if the judge chooses x_t^1 when $\omega_t = 1$ and x_t^0 when $\omega_t = 0$) and is zero otherwise. Then, judge i's payoff when choosing x_t is $U_i(x_t, \omega_t) = -\frac{1}{2}(z_i - x_t)^2 + g(\omega_t, x_t)$, and he will prefer x_t^1 to x_t^0 if and only if $2\omega_t - 1 \geq \lambda_t[\kappa_t - z_i]$. This expression is similar to the one derived before for the spatial voting model, with λ_t and κ_t defined analogously; i.e., $\lambda_t \equiv \frac{x_t^1 - x_t^0}{\varsigma_t}$ and $\kappa_t \equiv (\frac{x_t^1 + x_t^0}{2})$. The difference is that ω is an unobserved common value component. The common value induces a correlation in individuals' voting behavior, the extent of which is mediated by justices' ability and preference parameters. The fact that it is unobserved means that *judges* (and not only the econometrician) are uncertain about the realization of ω_t, and will vote in favor of x_t^1 only if

$$Pr^i(\omega_t = 1 | s_{it}) \geq \frac{1}{2} + \frac{\lambda_t}{2}[\kappa_t - z_i] \equiv \pi_{it} \tag{4}$$

Expression (4) illustrates that justices' decisions in the learning model come from a tug of war between their preferences and their case-specific information. If $\frac{\lambda_t}{2}[\kappa_t - z_i]$ is large, justices need a lot of information to overcome their bias π_{it}. This happens when the judge is not close to being indifferent between the two alternatives ($|(x_t^1 + x_t^0)/2 - z_i|$ is large), when the two alternatives are very different to one another ($x_t^1 - x_t^0$ is large), or when the payoff of getting the decision right under the law is low (low ς_t). In fact, $\lim_{\varsigma_t \to 0} \pi_{it} = \pm\infty$ depending on whether $\frac{x_t^1 + x_t^0}{2} > z_i$ or $\frac{x_t^1 + x_t^0}{2} < z_i$, so i will vote based solely on his preferences as the importance of deciding in accordance with the law vanishes. If instead individual preferences are less determinant, legal reasoning takes precedence. Note that $\lim_{\varsigma_t \to \infty} \pi_{it} = \frac{1}{2}$, so i becomes more likely to base his vote on case-specific information only as the gain from voting in the "correct" direction becomes more important vis-à-vis his personal preferences.

8

4 Research design

4.1 Data

We obtained judges' voting records from the *High Courts Judicial Database* (HCJD, Stacia et al. 2011). This data set includes all the cases heard by the Appellate Committee of the House of Lords (or the Judicial Committee of the Privy Council) between 1969 and 2002, divided in seven issue areas: *Civil Liberties*, *Commercial*, *Criminal*, *Family*, *Public Law*, *Torts*, and *Other*. The vast majority (79%) of these decisions have been classified as "liberal" ($v_t = 1$) or "conservative" ($v_t = 0$), and we use this classification to code the individual votes of the committee members sitting on the panel hearing each appeal. For robustness, we also consider an alternative operationalization of the dependent variable, coding decisions based on whether judges voted to overturn ($v_t = 1$) or uphold ($v_t = 0$) the ruling of the lower tribunals. This definition, which highlights the "quality control" function of the Appellate Committee aimed at ensuring that lower courts do not make wrong or inconsistent decisions (Dickson, 2007; Drewry and Blom-Cooper, 2009), assumes that judges differ in their proneness to lower-court reversal rather than on their ideological leanings, and avoids potential errors in the coding of liberal and conservative outcomes.[9]

We examine the votes of all judges who sat on the Appellate Committee, but exclude from our analysis those cases heard by less than three justices and the decisions made by a small number of Lords who heard very few cases.[10] In total, the sample using the liberal-conservative coding of outcomes comprises 1,206 unanimous and non-unanimous cases heard by 54 judges; the corresponding number of appeals using lower-court deference as dependent variable is 1,467. Table 1 below presents the distribution of cases by area.[11]

In Section 5.2, we estimate the learning model of judicial behavior allowing preference and information parameters to be a function of the characteristics of the cases and the

---

[9] For instance, pro-government rulings in public law cases are generally labeled as liberal in the *High Courts Judicial Database*. However, as noted by a reviewer, this is arguably not how most decisions against deportation or asylum claims would be currently interpreted among scholars and pundits in the UK. The criteria followed to classify outcomes across different issue areas can be found in Appendix C of the HCJD Codebook.

[10]Latent class regressions such as the one used to estimate the learning model are not identified with less than 3 votes per case (Huang and Bandeen Roche, 2004). Cases heard by less than 3 Law Lords were extremely rare, as panels of 5 judges became the norm - with additional members occasionally empaneled (Blom-Cooper, Dickson and Drewry, 2009). We excluded the decisions made by Lords Avonside, Devlin, Emslie, Gardiner, Havers, MacDermott, Parker of Waddington, Phillips of Worth Matravers, Taylor of Gosforth, and Wheatley.

[11]Family law cases are not classified as liberal or conservative in the HCJD, but they are included in the sample using lower-court deference as outcome.

Table 1: Decisions by Issue Area

Area	Outcome: Liberal/Conservative					Outcome: Overturn/Uphold lower courts				
	Liberal	Conservative	Unanimous	Minimal Winning	Total	Overturn	Uphold	Unanimous	Minimal Winning	Total
Civil Liberties	36 (40.4%)	53 (59.6%)	75 (84.3%)	5 (5.6%)	89	42 (47.8%)	46 (52.2%)	75 (85.2%)	5 (5.7%)	88
Commercial	67 (40.9%)	97 (59.1%)	141 (86.0%)	13 (7.9%)	164	177 (49.7%)	179 (50.3%)	299 (83.9%)	28 (7.9%)	356
Criminal	114 (33.7%)	224 (66.3%)	296 (87.6%)	23 (6.8%)	338	139 (41.0%)	200 (59.0%)	296 (87.3%)	23 (6.8%)	339
Family						26 (49.1%)	27 (50.9%)	49 (92.4%)	3 (5.7%)	53
Public Law	214 (61.3%)	135 (38.7%)	282 (80.8%)	34 (9.7%)	349	158 (44.5%)	197 (55.5%)	288 (81.1%)	34 (9.6%)	355
Torts	122 (49.6%)	124 (50.4%)	210 (85.4%)	13 (5.3%)	246	116 (46.4%)	134 (53.6%)	213 (85.2%)	13 (5.2%)	250
Other	7 (35.0%)	13 (65.0%)	18 (90.0%)	2 (10.0%)	20	4 (15.4%)	22 (84.6%)	24 (92.3%)	2 (7.7%)	26
Total	560 (46.4%)	646 (53.6%)	1,022 (84.7%)	90 (7.5%)	1,206	662 (45.1%)	805 (54.9%)	1,244 (84.8%)	108 (7.4%)	1,467

Note: Percentages calculated in terms of the row-totals. "Minimal winning" are cases decided by a difference of one vote.

individuals. Biographical information on each committee member was obtained from the *Oxford Dictionary of National Biography*, the volumes edited by Carmichael and Dickson (1999) and Blom-Cooper, Dickson and Drewry (2009), and newspaper articles and obituaries from the *Guardian* and the *Daily Telegraph*. Based on these sources, we define a justice as *Conservative* (*Liberal*) if he had contested a seat for the Conservative/Unionist (Labour/Liberal) party or held a government post under a Conservative (Labour) administration by the time of his appointment, with no (partisan) political experience as the reference category. The vector of judge-specific covariates also includes prior judicial experience, whether the judge was appointed to the court during a Conservative or Labour government, and a dummy for *English* members of the committee.

Among the case-specific covariates we include the substantive issue considered in the case, the type of Appellant and Respondent – distinguishing between the *State* and *Private Parties* (natural persons, businesses, non-profit organizations), whether the case came directly from the Court of Appeal or from other tribunals, the fraction of justices appointed by a Conservative administration, the proportion of Conservative/Unionist and Labour/Liberal judges in the committee as well as in the panel that heard the case, and the identities of the *Senior Law Lord* and of the *Lord Chancellor*.

Tables S.1 and S.2 in the Supplementary Materials Appendix provide descriptive statistics for these variables.

4.2 Estimation approach

Following the contributions of Poole and Rosenthal (1985), Heckman and Snyder (1997) and Clinton, Jackman and Rivers (2004), estimation of the spatial voting model is standard practice. Here we fit a hierarchical item response model with the parametrization suggested by Bafumi et al. (2005).

Estimation of the learning model proceeds in two steps, following a Bayesian version of the approach of Iaryczower and Shum (2012).[12] In the first stage, we fit a finite mixture model for the case-specific decisions *via* Markov chain Monte Carlo (MCMC) simulations (Gelman et al., 2004), obtaining posterior summaries for justices' prior beliefs ρ and the coefficients of the variables affecting the reduced-form voting probabilities. In the second step, we recover the equilibrium strategies' cutpoints and structural parameters $\{\theta_i, s_i^*, \pi_i\}_{i=1}^n$ from these reduced-form probabilities.

First step

Let $\rho_t \equiv \Pr(\omega_t = 1)$ and denote the conditional voting probabilities $Pr(v_{it} = 1|\omega_t = 1)$ and $Pr(v_{it} = 1|\omega_t = 0)$ by $\gamma_{i,1}$ and $\gamma_{i,0}$, respectively. Given this notation, the likelihood of judges' votes in case t is given by:

$$\Pr(\mathbf{v}_t) = \rho \prod_{i=1}^n \left[\gamma_{i,1}^{v_{it}}(1 - \gamma_{i,1})^{1-v_{it}}\right] + (1 - \rho) \prod_{i=1}^n \left[\gamma_{i,0}^{v_{it}}(1 - \gamma_{i,0})^{1-v_{it}}\right] \qquad (5)$$

---

[12]Some of the advantages of the Bayesian framework in this setting are that it allows for a detailed description of the parameters and auxiliary quantities of interest via examination of their posterior distributions, and that it helps account for the uncertainty in the recovered structural parameters while avoiding asymptotic approximations - a convenient feature given that the total number of Lords included in our analysis is rather small.

Conditional on the state ω_t, the individual votes v_{it} are independent across justices. Thus, the vector of votes $\mathbf{v}_t = (v_{1t}, v_{2t}, \ldots, v_{nt})^\top$ follows a finite mixture distribution with mixing probability ρ. In order to account for heterogeneity in the voting probabilities and common prior beliefs, we allow ρ to vary as functions of case-specific variables X_t and model $\gamma_{i,j}$, $j = 0, 1$ as functions of individual characteristics Z_i:

$$\rho(X_t) = \frac{\exp(X_t'\alpha)}{1 + \exp(X_t'\alpha)}, \tag{6}$$

$$\gamma_{i,j}(Z_i) = \frac{\exp(\delta_{i,j} + Z_i'\beta_j)}{1 + \exp(\delta_{i,j} + Z_i'\beta_j)} \tag{7}$$

where α and β are "fixed-effects" regression coefficients, and $\delta_{i,j} \sim N(0, \sigma_j^2)$ are judge-specific random intercepts. Following Garrett and Zeger (2000), we assigned α a multivariate $N(0, (9/4)I)$ distribution that yields relatively flat priors for ρ centered around $1/2$, and used conjugate Gaussian and Inverse-Gamma priors for β and σ_j^2, $j = 0, 1$, respectively.[13] This hierarchical specification allows "borrowing strength" across individuals with different number of observations and enables us to estimate judge-specific vote probabilities even though the composition of the panels varies across cases.

The MCMC sampler alternates between three main steps: (i) updating α given the state variable ω using a random-walk Metropolis step, and calculating ρ; (ii) sampling ω_t from its full conditional distribution given the other model parameters; (iii) updating δ and β conditional on ρ and ω using random-walk Metropolis steps, and computing γ. Each cycle of the algorithm is completed by updating the variances of the judge-specific random intercepts from their full conditional Inverse-Gamma distributions.[14] This leads to an iterative scheme whereby, starting from an arbitrary set of initial values and under mild regularity conditions (Gelman et al., 2004), we obtain samples of $(\rho^m, \gamma^m, \omega^m)$ from their posterior density at each iteration $m = 1, \ldots, M$. Three parallel chains with dispersed initial values were run for 75,000 cycles each after an initial burn-in period, with convergence

---

[13]We also estimated an alternative specification assuming a bivariate $N(0, \Sigma)$ distribution for δ_i and a conjugate Inverse-Wishart (IW) prior for the variance-covariance matrix. The main results reported in the manuscript do not change, although in this case a relatively informative IW distribution is required to avoid weak identification.

[14]Additional technical details about the sampling algorithm are presented in the Supplementary Materials Appendix (Section S.2).

assessed based on Gelman and Rubin (1992)'s potential scale reduction factor, $\hat{R}$. Routine sensitivity checks were performed to assess the robustness of the estimates to the prior distributions. The average overlap between the prior and posterior distributions for the parameters governing the latent class membership probabilities was quite small, indicating that the model is well identified and relatively insensitive to prior assumptions (Garrett and Zeger, 2000).

In order to deal with potential "label switching", a well-known problem for MCMC estimation of latent class models, we implemented the decision-theoretic post-processing algorithm proposed by Stephens (2000). This provides a more theoretically sound and methodologically general identification procedure than approaches imposing prior constraints on the parameter space (e.g., Stephens, 2000, Section 5). At each cycle $m = 1, \ldots, M$, the relabeling algorithm considers the two possible permutations of the model parameters and selects the one that minimizes the Kullback-Leibler distance between the posterior probabilities $P(\omega_t^m = j|\rho^m, \gamma^m)$, $j = 0, 1$, and their posterior means $\frac{\sum_{m=1}^{M} P(\omega_t^m = j|\rho^m, \gamma^m)}{M}$. More precisely, we choose the permutation ν_m of the sampled parameter values $\Psi^m = (\alpha^m, \beta^m, \delta^m, (\sigma^2)^m)$ that minimizes

$$\sum_{t=1}^{T} \sum_{k} P(\omega_t^m = k|\nu_m(\Psi^m)) log \left[\frac{P(\omega_t^m = k|\nu_m(\Psi^m))}{(M^{-1}) \sum_{m=1}^{M} P(\omega_t^m = k|\nu_m(\Psi^m))} \right] \qquad (8)$$

It must be noted, though, that visual inspection of the trace plots shows little evidence of label switching in our analyses.

Second step

Since we assumed that $s_{it} \sim N(\omega_t, 1/\theta_i^2)$, $\gamma_{i,1} \equiv 1 - \Phi(\theta_i[s_i^* - 1])$ and $\gamma_{i,0} \equiv (1 - \Phi(\theta_i s_i^*))$. Solving these equations for θ_i and s_i^* given $\gamma_{i,1}^m$ and $\gamma_{i,0}^m$, $m = 1, \ldots, M$, yields:

$$\theta_i^m = \Phi^{-1}(1 - \gamma_{i,0}^m) - \Phi^{-1}(1 - \gamma_{i,1}^m) \quad \text{and} \quad s_i^{*m} = \frac{\Phi^{-1}(1 - \gamma_{i,0}^m)}{\Phi^{-1}(1 - \gamma_{i,0}^m) + \Phi^{-1}(\gamma_{i,1}^m)} \qquad (9)$$

To obtain π_i^m, we simply plug θ_i^m and s_i^{*m} into the equilibrium voting condition (2).

5   Results

In this section we present our main results. We begin in Section 5.1 by evaluating the power of the ideological and learning models to explain the voting decisions of the Law Lords.

Since it is standard practice to estimate the spatial voting model without including individual or case-specific covariates, we use this "unconditional" specification for the comparison between the two models. This also seeks to attenuate the influence of the differences in functional forms and identification strategies used by the competing models, which can pose some difficulties for in-sample comparisons (Clinton and Jackman, 2009).

We show that the learning model fits the data remarkably well according to a variety of criteria commonly used in the literature, outperforming the ideological model of judicial behavior. Based on these results, we describe the substantive findings of the learning model in Section 5.2, specifying the preferences and information parameters as a function of characteristics of the judges and/or the cases they consider. This allows us to directly assess the correlation between factors such as justices' partisanship or judicial experience and their preference and ability parameters.

5.1   Comparing models of judicial behavior in the House of Lords

We evaluate the explanatory power of the ideological and learning models based on various goodness-of-fit indicators commonly used in the literature. The different measures have their own advantages and drawbacks, but together provide a fairly comprehensive picture of each model's ability to account for the observed data patterns.

Table 2 presents a detailed set of model comparisons for the analysis using the "liberal"-"conservative" coding of outcomes. The first two rows present the most widely used goodness-of-fit indicators for binary response models: the percentage of correctly classified decisions and the proportionate reduction in error (PRE).[15] We also include the expected percent of correctly predicted decisions (ePCP) proposed by Herron (1999). This measure is designed to overcome a potential deficiency of the classification success rates and PRE, which can overstate the accuracy of the results due to their rather "coarse" treatment of fitted probabilities. More recently, Bafumi et al. (2005) used the excess error rate, namely, the proportion of error beyond what would be expected given a model's predicted values, as another way to test how well it fits a set of observations. Table 2 reports the absolute value of the realized error rates averaged across judges and across cases.[16] We also report the Akaike (AIC) and Schwartz (BIC) information criteria for each model.

---

[15]The PRE measures the relative improvement in classification success rates vis-à-vis a naive baseline model in which each justice chooses the modal outcome.

[16]The interested reader is referred to Herron (1999) and Bafumi et al. (2005) for a description of these indicators and the formulas used to calculate them.

The first column of the table presents the goodness-of-fit indicators for the learning model in the full data set comprising all unanimous and non-unanimous votes. This is the appropriate sample for estimation and evaluation of the learning model, because the relative frequency of unanimous decisions provides information about justices' biases and ability. Specifically, since justices disposing of a case are learning about an unobserved common value component, we expect the size of the majority to be increasing in the precision of justices' private information. The large proportion of unanimous decisions observed in the Appellate Committee can thus be naturally interpreted as a consequence of the correlation in voting behavior induced by the common value component in our learning model.

In the context of the spatial voting model, however, unanimous decisions provide no information about judges' ideological position, and are therefore typically not included in the estimation. For comparison purposes, we re-estimate the learning model using only the cases in which there was at least one dissenting opinion. Columns 2 and 3 compare the goodness-of-fit of the learning and spatial models in this restricted sample. We want to emphasize, though, that this is done solely for comparison purposes. Unless we knew a priori that cases decided by unanimity are uninteresting and fundamentally different from the rest, fitting a model to this restricted sample comprising only about 15% of the appeals would essentially constitute a form of data sub-setting that could generate selection effects and distort conclusions about the judicial function of the House of Lords.

Two main findings emerge from Table 2. First, the learning model of judicial behavior explains the data remarkably well when both unanimous and non-unanimous decisions are considered, correctly predicting 95% of the individual votes. The proportionate reduction in error and the expected percent of correctly predicted decisions also exceed 90%.

Second, the fit of the learning model is consistently better than that of the ideological model. For instance, even in the restricted sample comprising only divided decisions, the classification success rate and the ePCP for the learning model are roughly 10 percentage points higher than for the spatial voting model. The absolute excess error rates averaged across justices and cases are 6 to 11 p. points higher under the ideological model, with more than 70% of the Lords having smaller realized error rates in the learning model of judicial behavior (see Figure S.1 in the Supplementary Materials Appendix). It is also worth noting that the average excess error rate under a null model in which committee members vote with the majority is more than twice as large as under the learning model, and almost 3 times larger in the full sample comprising all the committee's decisions. The learning model also fares better when we take into account the relative "complexity" of the

Table 2: Measures of Complexity and Fit for the Learning and Spatial Voting Models
Liberal (conservative) outcomes coded as 1(0)

| | All decisions | Non-unanimous | |
	Learning Model	Learning Model	Spatial Model
Correctly predicted	0.95	0.70	0.66
PRE	0.90	0.40	0.31
ePCP	0.92	0.61	0.50
Excess error rate: by judge	0.02	0.12	0.18
Excess error rate: by case	0.07	0.11	0.22
AIC	6,247.54	1,616.29	2,110.05
BIC	14,313.68	2,513.42	4,145.48
Number of cases	1,206	184	184

Note: In the spatial voting model, unanimous decisions provide no information about the position of the Lords. Hence, we only report goodness-of-fit statistics for this model for the sample in which there is at least one dissenting opinion in the panel. To identify the IRT model, we normalized all the parameters as in Bafumi et al. (2005) and constrained the ideal point of Baron Wilberforce - who ran for Parliament as Conservative candidate - to lie on the conservative end of the policy dimension. Similar results were obtained using other identification restrictions - e.g., constraining Fraser and Morris of Borth-y-Gest to lie on opposite sides of the ideological space, or Griffith to be more liberal than Bridge (Hanretty, 2013).

two competing approaches. As seen in the bottom part of Table 2, both the Akaike and Schwartz information criteria are markedly lower for the learning model in the restricted sample, providing strong evidence against the ideological model.[17]

A better approach to guard against overfitting is to compare the out-of-sample performance of the competing models. As is well known, however, an important limitation of the ideological model is its inability to generate out-of-sample predictions (Clinton and

---

[17]As a "rule of thumb", AIC/BIC differences larger than 10 provide overwhelming evidence in favor of the model with the lower value. See Ntzoufras (2011) and the references therein for an overview of model comparisons based on information criteria.

Jackman, 2009). In contrast, we can evaluate the out-of-sample predictive performance of the learning model *via* leave-one-out cross-validation, contrasting judges' decisions in each case t with the votes predicted from a training sample excluding t. To simplify the computations, we approximate the learning model's cross-validatory predictive performance by weighted resampling from the full posterior density (Gelfand, 1996). This yields an out-of-sample classification success rate of about 95% in the full sample. The proportion of correctly predicted decisions in the restricted sample is about 68%. For comparison, a null model in which each judge chooses the modal outcome (of the training sample) correctly classifies only slightly more than 50% of the votes; a null model in which committee members vote with the majority in each case cannot form out-of-sample predictions.

An important distinction between the ideological and learning models is that while the former can recover the liberal-conservative orientation of each vote from the parameter estimates (subject to identification constraints), the latter requires it to be defined a priori through the coding of the dependent variable. It must be noted that, for about a third of the individual decisions, the liberal-conservative classification taken from the *High Courts Judicial Database* does not coincide with the labeling obtained from the IRT model, the largest proportion of them involving public law appeals. In fact, the posterior mean of the discrimination parameter is negative for roughly 10% of the non-unanimous appeals, suggesting possible miscoding of their outcome (see Figure S.3 in the Supplementary Materials Appendix). Almost two-thirds of these cases belong to the public law area, where the coding of liberal and conservative outcomes is arguably more controversial (e.g., footnote 9). Nevertheless, the 90% credible intervals of virtually all the discrimination parameters in the IRT model overlap zero, regardless of the identification constraints adopted to deal with rotational invariance.[18] Thus, the inconsistency should be taken with a grain of salt.

Recall from Section 3 that the discrimination parameter is given by $\lambda_t \equiv \frac{x_t^1 - x_t^0}{\varsigma_t}$, where x_t^1 and x_t^0 are the positions of the alternatives in case t and ς_t is the variance of judges' private preference shocks. Thus, the fact that $\lambda_t \approx 0$ implies that in order to fit the data, the spatial voting model needs preference shocks to be large relative to the distance between alternatives. This suggests that the stable ideological component of judges' preferences (as captured by judges' ideal points) does a poor job at explaining voting behavior.

---

[18]Similarly, when using judge- or case-specific predictors to resolve the aliasing (Bafumi et al., 2005), the coefficient of the chosen separator variable is always statistically indistinguishable from 0.

Alternative definition of the outcome variable. In the previous specification, we assumed that the dimension of conflict in preferences among judges can be summarized by a conservative/liberal policy space. In a more decidedly "legal" model, however, we can think that disagreement among justices is primarily about the different juridical arguments and how they apply to specific situations. In particular, in view of the Appellate Committee's role as the final appeal court in the UK until 2009, a potentially relevant distinction between judges might be in the extent of the deference they give to the decisions of the lower tribunals. We thus re-estimated the ideological and learning models coding a vote in favor of allowing appeals as $v_{it} = 1$, and $v_{it} = 0$ otherwise.[19]

Figure 1 contrasts the performance of the learning and spatial voting models in this scenario, plotting their receiver operating characteristic (ROC) curves. The ROC curves for the learning model clearly dominate the curve for the ideological model, indicating again the better fit of the former. In the restricted sample considering only non-unanimous decisions, the area under the curve - a measure commonly used to compare the performance of binary classifiers - is 0.75 for the ideological model and 0.82 for the learning model, a difference that is statistically significant at the 0.01 level based on both DeLong and bootstrap tests.[20] The proportion of correctly classified votes to overturn lower court rulings ("true positive rate") is more than 3 percentage points higher for the learning model, and the proportion of decisions dismissing appeals that was incorrectly classified ("false positive rate") almost 4 p.p. lower. Once again, using the full sample to estimate the ideological model leads to a markedly improved fit: the "true positive rate" goes up from 0.65 to 0.85, the "false positive rate" is reduced by more than half (0.13 *versus* 0.33), and the ROC area exceeds 0.95. The superiority of the learning model is further demonstrated in Section S.3 of the Supplementary Materials Appendix, which presents a more detailed set of comparisons based on the same indicators used in Table 2.

---

[19]For this exercise, we assumed constant discrimination across cases when fitting the IRT model.

[20]In the same direction, Venkatraman's test comparing the two actual ROC curves (rather than their areas) yields a p-value $< 5.434e - 13$. See Robin et al. (2011) and the references therein for an overview of different statistical tests comparing ROC curves.
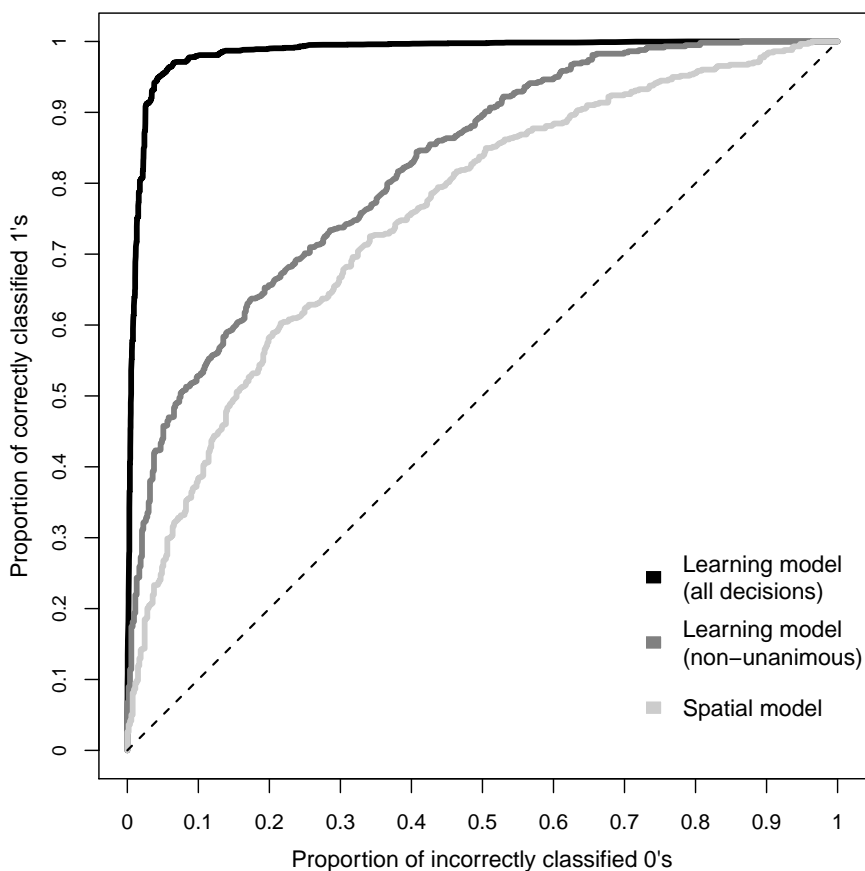
Figure 1: Model comparison based on ROC curves, using votes in favor of (against) overturning lower court decisions as the dependent variable. Solid lines represent the ROC curves for the learning and spatial voting models; the 45-degree line corresponds to a random prediction model.

In sum, the results in this section support two fundamental conclusions. First, that the learning model of judicial behavior fits the data extremely well, especially when all the information contained in the committee's decisions is taken into consideration. Second, that - regardless of the particular coding of the dependent variable - allowing for differences in justices' ability provides a better account of the judicial work of the House of Lords than a priori assuming that it is entirely driven by ideological conflict or idiosyncratic preferences. In fact, as we show below, information about the cases was frequently powerful enough to overturn the biases of the Law Lords during the period analyzed.

5.2 Ability and Ideology in the Appellate Committee

In this section, we describe the substantive findings of the learning model of judicial behavior in the Appellate Committee. We focus here on proneness to revert lower court rulings as the outcome of interest.[21] We thus fit the model to the full sample of 1,467 unanimous and unanimous decisions coded as $v_t = 1$ when they allow the appeal, and $v_t = 0$ otherwise.[22]

A crucial feature of the hybrid learning model is that it can disentangle the impact of judges' ability and biases on the court's decisions. This, in turn, allows us to address a variety of interesting questions that were previously beyond the scope of empirical models of judicial decision-making in the House of Lords. First, to what extent are differences in judges' observed voting behavior attributable to idiosyncratic biases? Are political factors associated with Lords' biases in favor or against overturning the judgments of lower tribunals? Second, what do we learn about differences in ability or quality of information among justices in the court? In particular, what is the impact of judicial experience on Lords' ability to process information effectively? Third, we are interested in how bias and ability interact with one another. Do individual preferences trump differences in ability and information? Or, on the contrary, does case-specific information lead judges to vote against their initial leanings? These are the questions we answer here.

In this setting, the parameter π_i quantifies the barrier that idiosyncratic preferences put on the information supporting the decision to allow or dismiss appeals. More specifically, $\pi_i - 1/2$ represents a (possibly negative) hurdle to overturn the decision of the lower courts. A completely unbiased judge ($\pi_i = 1/2$) will vote to allow the appeal only when the law and the facts of the case support a decision to reverse the ruling of the court below; i.e., depending on whether $\Pr(\omega_t|s_{it}) \geq 1/2$ or $\Pr(\omega_t|s_{it}) < 1/2$. A judge who is a priori more (less) inclined to defer to lower tribunals has a positive (negative) hurdle, and only votes to allow an appeal if the case-specific information surpasses the hurdle imposed by his preferences; i.e., if $\Pr(\omega_t|s_{it}) > \pi_i$.

---

[21] This avoids controversies regarding the liberal/conservative labeling of cases. Using lower court deference as outcome also mitigates a potential concern raised by a reviewer regarding panel composition. The criterion followed in the HCJD to classify outcomes is essentially based on who "wins" the appeal. For instance, decisions favoring the government are usually coded as liberal (conservative) in public law (criminal) cases. Since the government wins the vast majority of the cases, if panel selection is - at least partly - based on substantive expertise, judges with experience in these areas could appear as liberal/conservative even when they might not lean in that direction.

[22] We note, however, that the main findings regarding the role of information in the court and the differences in judges' ability hold for the liberal-conservative classification of outcomes as well. The interpretation of judges' biases, of course, varies according to the definition of the dependent variable. This additional set of results is available from the authors upon request.

Figure 2 summarizes the relevant findings about judges' biases. The posterior means of π range between a lower bound of 0.33 for Lord Diplock and an upper bound of 0.75 for Lord Steyn. That is, Lord Diplock would be inclined to allow an appeal even when he thinks that affirming the lower court ruling is more than twice as likely to be the correct choice. On the other hand, Lord Steyn would only do the same if he thinks that the probability that dismissing the appeal is correct under the law is below 25%. The bias of the median judge (Lord Brightman) is 0.52, indicating that he votes to allow or dismiss appeals depending essentially on the facts of the case alone. In fact, 95% of the estimated $\hat{\pi}_i$ lie between 0.33 and 0.67, and most of the biases are statistically indistinguishable from 0.5, implying that the informational hurdle to vote for or against overturning the decision of the lower courts is rather moderate among committee members.

The relative ordering of justices' preference parameters in Figure 2 seems to be related to the party which appointed them. The average bias for a Labour/Liberal nominee is 0.47, against 0.54 for a judge appointed during a Conservative administration, indicating that the former are typically be more inclined to allow appeals than the latter. The likelihood that a Labour/Liberal appointee is more predisposed to overturn a lower court decision than a Conservative nominee is 0.87 on average across all cases heard by the Appellate Committee.
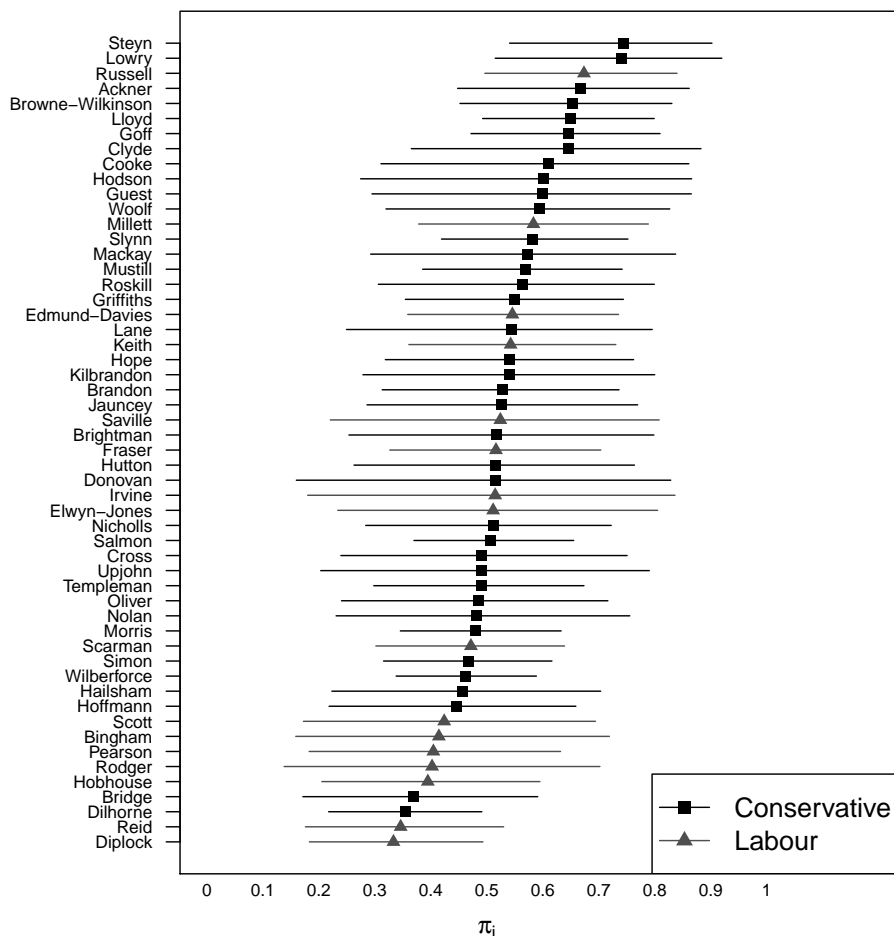
Figure 2: Individual biases. The figure provides posterior summaries of π for each judge. Solid squares (circles) give the posterior means for Conservative (Liberal/Labour) nominees, with horizontal lines corresponding to the 90% credible intervals.

Political or ideological considerations, however, are not perfect predictors of justices' biases. On the one hand, there is a considerable overlap in the posterior distribution of individual preferences among Conservative and Labour nominees. Thus, the posterior means of π_i for Lords Dilhorne and Bridge, two conservative nominees, are below 0.4. Both are more prone to reverse lower court decisions than Elwyn-Jones, the median Labour appointee. Moreover, individual biases tend to be quite moderate also for justices that could be assumed to be more decidedly "partisan", such as those who served as Lord Chancellors during Conservative (Hailsham, Mackay of Clashfern) or Labour (Elwyn-Jones, Irvine of Lairg) administrations. There is also no systematic relationship between judges'

own partisan affiliation and their predisposition to allow or dismiss appeals. Lord Reid, one of the justices less a priori inclined to uphold lower tribunals ($\hat{\pi}_{Reid} = 0.35$), ran for Parliament as a Conservative candidate, but so did Lord Clyde, who requires considerably more evidence to overturn ($\hat{\pi}_{Clyde} = 0.64$).[23] The relationship between justices' preference parameters and their politics is further attenuated when we consider a strategic version of the model in which judges care about the outcome of the case, rather than about their own decision exclusively (Section S.5 of the Supplementary Materials Appendix).

Voting behavior in the learning model, though, is driven not only by individuals' preferences, but also by differences in their ability to extract relevant information from the case facts. This is captured by the parameter θ_i.[24] Of course, it might turn out that in practice there are only small differences in ability between judges, in which case voting behavior would be mostly dictated by justices' preferences. However, this is not what we observe among the Lords of Appeal (see Figure 3). The value of θ_i ranges from 2.19 to 4.24 and is quite precisely estimated, with much less overlap between judges than for the bias parameter. This indicates important differences in ability among committee members.

To better understand the implications of these differences in ability, recall from equation 9 that the estimate of θ_i is a function of the difference between the conditional probabilities of voting to dismiss the appeal when legal considerations indicate that the lower court decision should be upheld ($\omega_t = 0$) and when they indicate that the ruling should be overturned ($\omega_t = 1$). That is, θ_i increases in the probability of correctly allowing the appeal ($\gamma_{i,1}$) and decreases in the probability of incorrectly overturning the lower tribunal ($\gamma_{i,0}$). Hence, Lord Morris, the judge with the lowest $\hat{\theta}_i$, has an 11% probability of incorrectly allowing appeals and an 18% probability of incorrectly upholding lower court decisions. The corresponding probabilities for Lord Bridge, the justice with the highest estimated ability parameter, are substantially lower: 2.3% and 1.5%, respectively. On average, the estimated probability that a Law Lord reaches an incorrect decision is 4.8%, with an asymmetric pattern of mistakes: justices are more likely to vote to uphold lower court rulings when they should overturn them (5.7%) than to erroneously allow appeals (4.1%).

---

[23] A similar conclusion holds if we consider experts' opinions of judges' ideological stance. For instance, the biases of Ackner and Hobhouse, both seen as "conservative" by reliable commentators (Hanretty, 2013), lie at opposite sides of the distribution of the preference parameter.

[24] To make this abundantly clear, we think of differences in θ as the differences we might see between how an expert referee evaluates a manuscript and how an inexperienced graduate student does the same job (keeping bias fixed for comparison). While the expert referee knows what to look for in a paper, an inexperienced referee might get distracted with minor details and choose to publish a paper that should not be published, or reject a paper that should.
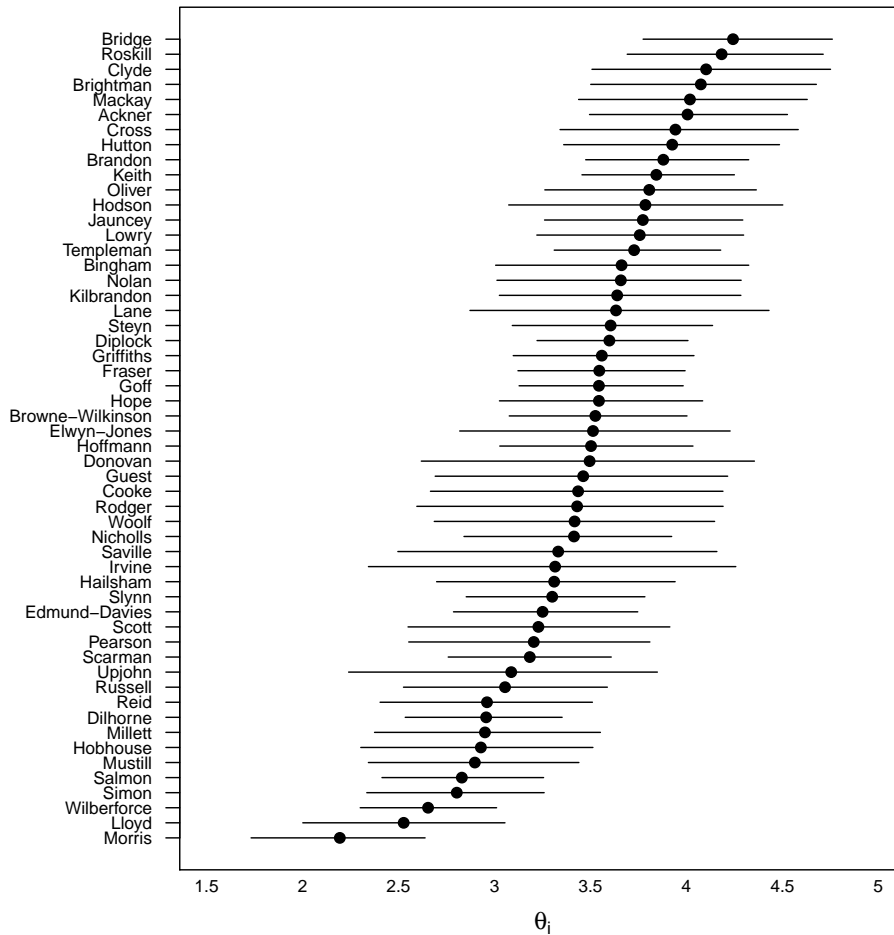
Figure 3: Precision of judges' private information. The figure provides posterior summaries for θ for each justice. Solid circles correspond to posterior means and horizontal lines to the 90% credible intervals.

We find no statistically significant differences in ability between judges with and without *political* experience, or between Conservative and Liberal/Labour nominees. However, Lords' ability estimates are correlated with their *judicial* experience. Keeping everything else constant, each additional year on the bench is associated with a 0.4% increase in $\hat{\theta}_i$. A committee member with no experience in the High Court is more than twice as likely to incorrectly overturn lower tribunal decisions than one with 25 years of judicial experience, and almost 31% more likely to erroneously dismiss appeals.

Naturally, both justices' bias and ability can differ across issue areas, be it because individual preferences weigh more heavily in some areas than in others or due to differences

24

in justices' expertise. In order to assess the variability of these structural parameters, we re-estimated the model separately for each area of the law. Figure 4 reveals that Law Lords are indeed significantly less predisposed to overturn decisions involving civil rights and liberties than other judgments. The average judge is inclined to allow appeals in this area only if he assigns a probability of at least 87% to the event that the standing decision is incorrect under the law. In family law cases, in contrast, the average justice is willing to dismiss appeals even if, based on the available information, there is a probability of only 12% that the lower court's decision is incorrect. The relative ordering of judges' biases to overturn or uphold lower court rulings also varies across areas of the law, as shown in Figure S.4 of the Supplementary Materials Appendix.
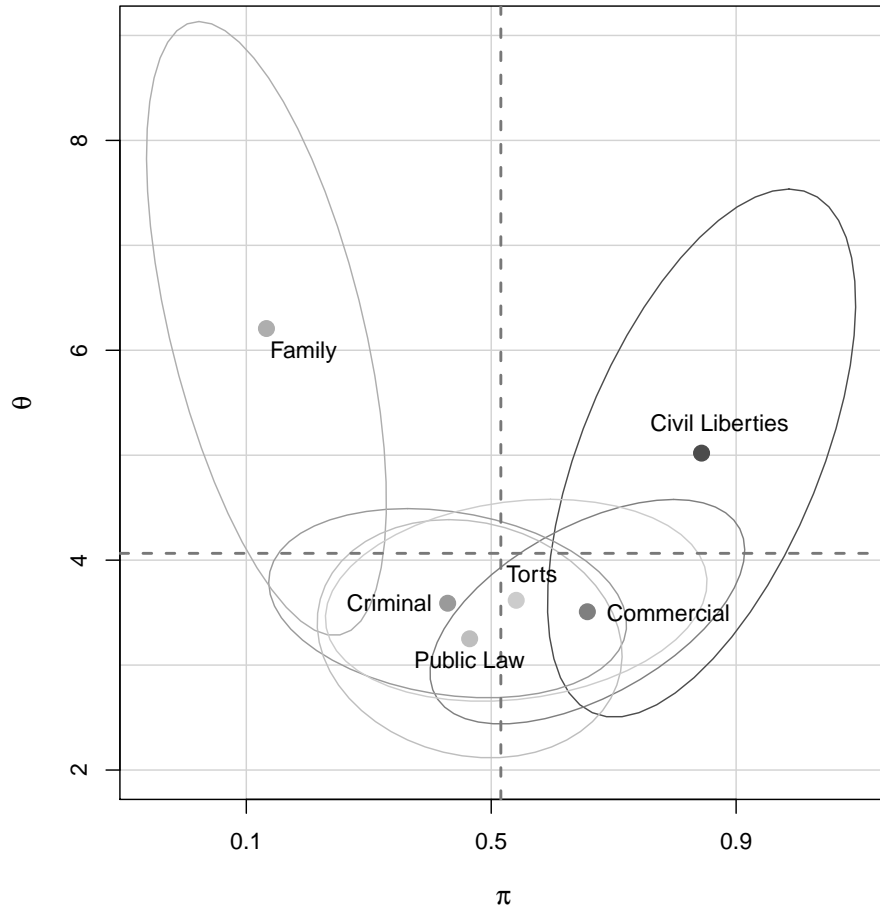
Figure 4: Relationship between bias and ability across issue areas. The figure plots point estimates and approximate 90% credible regions for judges' biases and ability in civil liberties, commercial, criminal, family, public law and tort cases. Dashed lines correspond to the average bias and information precision across all the areas.

While we also observe some differences in ability across areas, these are not statistically significant at the usual levels. We do, however, find important variations in the impact of judicial experience on θ (Figure 5). Holding everything else constant, a one standard deviation increase in judicial experience - roughly 5 years - is associated with an 8%-15% increase in the precision of judges' information in appeals involving commercial law, torts and civil liberties. In the other areas of the law, the correlation between years on the bench and ability is not statistically significant.
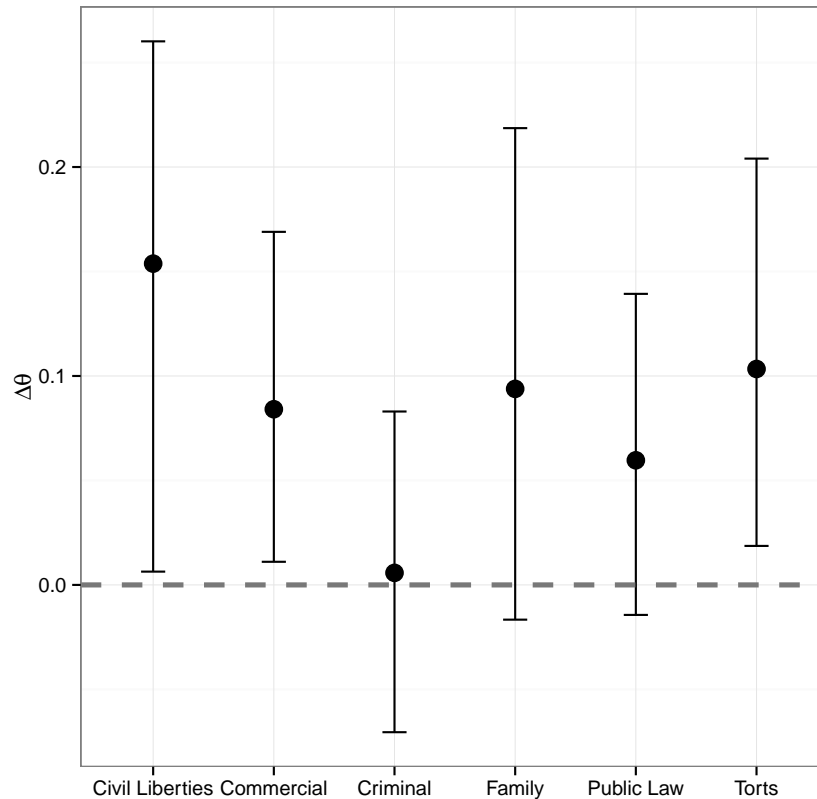
Figure 5: Relationship between judicial experience and ability across issue areas. The figure plots the expected percentage change in θ associated with a one-standard deviation increase in judges' years of experience on the bench. Circles represent posterior means, while vertical lines give the 90% credible intervals.

In addition to their ability and judicial preferences, judges' voting decisions are also shaped by their *prior beliefs*. For given characteristics of the appeal and the decision-making environment X_t, $\rho(X_t)$ measures the initial probability that a judge gives to the appropriate decision being overturning the lower court ruling *before* observing additional information pertinent to the case.

The impact of prior beliefs on judges' voting behavior depends on how precise is the case-specific information. The noisier and more ambiguous this information, the larger the scope for judicial discretion and the influence of judges' biases and initial beliefs on their decisions. To quantify the relative weight of information and predispositions in the Appellate Committee, we use the "FLEX score" proposed by Iaryczower and Shum (2012). For given characteristics of the case, the FLEX score measures the probability that, after

observing a private signal, each Lord votes differently from what he would have decided based on his prior belief alone, i.e., that $v_{it} = 0$ ($v_{it} = 1$) even though $\rho_t \geq \pi_i$ ($\rho_t < \pi_i$). It is then given by $\rho_t \Phi(\theta_i[s_i^* - 1]) + (1 - \rho_t)\Phi(\theta_i s_i^*)$ if $\rho_t \geq \pi_i$, and $\rho_t[1 - \Phi(\theta_i[s_i^* - 1])] + (1 - \rho_t)[1 - \Phi(\theta_i s_i^*)]$ if $\rho_t < \pi_i$.[25]

The average FLEX score in the sample is almost 0.45, demonstrating the sizable value of information in the Appellate Committee: about half of the time the Law Lords voted against their initial consideration of the case, based on bias and prior alone. The left panel of Figure 6 shows that the likelihood that the median justice "changes his mind" after observing his private information varies across areas of the law: FLEX scores are lowest on average in criminal (0.39) appeals, and highest in commercial (0.48) cases. This is consistent with Robertson (1998), who asserts that commercial law operates as an "hermetically sealed system" capable of generating routine and very precise answers based on analogy to precedents. Moreover, an overwhelming proportion of the Lords of Appeal was drawn from the Commercial Bar. It is therefore reasonable for the value of information to be higher in this domain than in criminal cases, an area in which Law Lords have traditionally had little expertise (Drewry and Blom-Cooper, 2009), or in civil rights matters, where judges' political views are arguably more salient and differences in their preferences more prevalent (Malleson, 2009).

In the right panel we turn to considering the evolution of FLEX scores over time, on average across all issue areas. Bingham (2009) and Malleson (2009), among others, assert that the appointment process to the Appellate Committee became increasingly apolitical in the second half of the 20th century, with a greater emphasis being placed on the legal proficiency of the candidates rather than on their political or ideological views. Accordingly, we would expect to see justices becoming more responsive to their private signals over time.

The evidence in Figure 6 is consistent with this view: the median FLEX score in the committee rose by 13% between 1969 and 2002, with a particularly marked increase in commercial and family law cases (see Figure S.5 in the Supplementary Materials Appendix). It is also worth noting that FLEX scores in criminal appeals increased by almost 10% since the 1970s, in line with Blom-Cooper and Drewry (2009)'s claim that the work of the Law

---

[25]The FLEX score is inherently a counterfactual quantity that can be computed directly from the parameter estimates. For this counterfactual exercise to be convincing, the structural model has to be invariant to the change in the environment under consideration. This is verified here because, by definition, FLEX takes the characteristics of the case as *given*. This type of exercise is fundamentally different from counterfactuals in Rubin's causal model, where the key assumption is the "stable unit treatment value assumption", requiring the potential outcome for any particular unit to be unaffected by the assignment of treatments to other units. No such assumption is needed in our context.
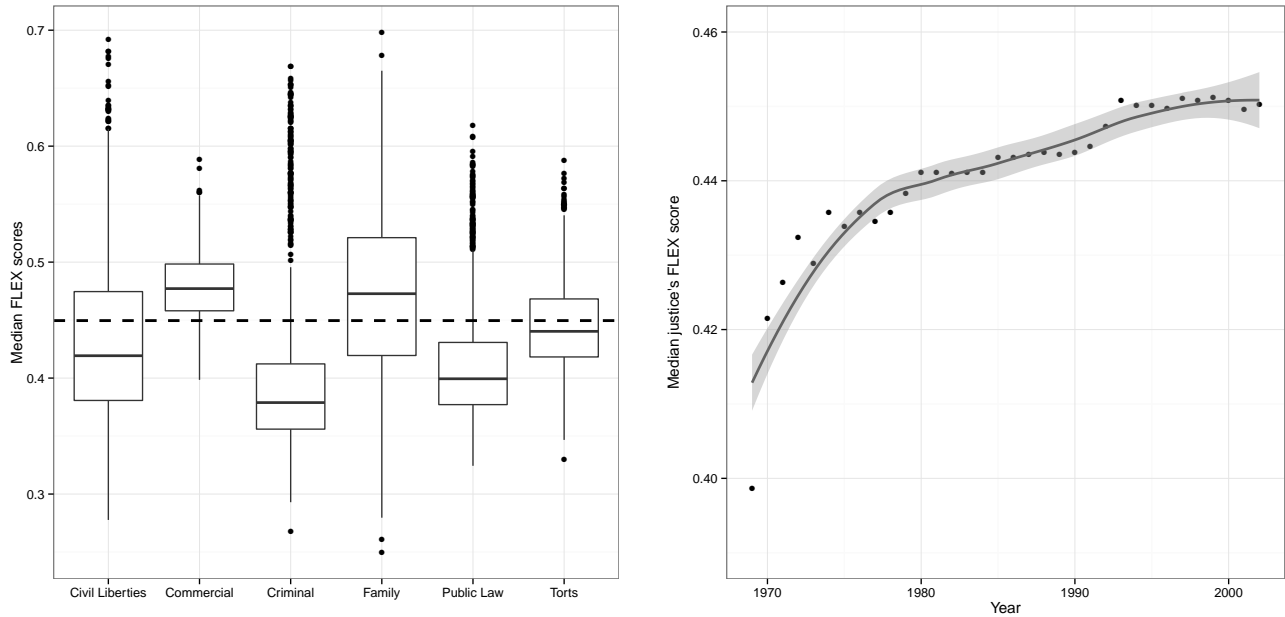
Figure 6: Judges' FLEX scores. The left panel plots the distribution of the median justice's FLEX score in each issue area; the dashed horizontal line gives the overall median score across all areas. The right panel tracks the evolution of FLEX scores over time. Circles represent the scores of the median committee member in each year, the solid lines gives the smoothed temporal trend, and the shaded area corresponds to the 90% credible intervals.

Lords in this area - traditionally viewed as unsatisfactory - showed a distinct improvement in the later part of the century.

Nonetheless, we must note that the median FLEX score was around 0.4 even in 1969. By then, of course, the influence of politics in the selection of the senior judges in the UK had diminished dramatically in comparison to the pre-1945 years, when - in the words of Laski (1925) - judicial office was often "a reward for political service" (p. 535). While our data does not allow us to test the long-term impact of this depoliticization process on the work of the Law Lords, our results attest to the sizable weight of information in the court throughout the period under study.

Agenda Setting. The House of Lords had ample discretion to determine its own judicial workload - particularly in its later years - and typically dealt with a rather small, "hand-picked" number of appeals (Drewry and Blom-Cooper, 2009; Paterson, 2013). The fact that the selection of cases in the sample was not random means that we should pause before

extending the findings to other cases not included in our data set. It should be noted, though, that to the extent that differences in the pool of cases can be captured by the observable covariates X_t entering into the prior function $\rho(X_t)$, variations in the prior will measure this effect. While we would not argue that the covariates in our specification can account for all possible case selection effects, we can use this fact to evaluate the presence of a basic kind of agenda setting.

To do this, we allowed the prior to be a function of the committee's composition, of the political experience of the judges sitting on each panel, and of the identity of the Senior Law Lord and the Lord Chancellor. This allows a basic test of whether the type of cases heard varied with the political color of the majority or the party in control of appointments. Figure 7 shows that this turns out not to be the case: none of these covariates has a statistically significant influence on Lords' common prior beliefs about the "right" judgment in the appeals.[26] The same result holds true for each of the issue areas considered separately (see Figure S.6 in the Supplementary Materials Appendix). Because the proportion of cases that came to the court after being granted leave to appeal directly by the Lords - as opposed to *via* the Court of Appeal - varied considerably over time (Paterson, 2013), we also re-estimated the model including a time trend among the predictors of ρ and replicated the analysis for different sub-periods. The conclusions remained essentially unchanged.

---

[26]In fact, the only variable systematically correlated with ρ is the identity of the litigants. Other things equal, judges are a priori significantly more likely to believe that lower court decisions should be overturned (affirmed) when the State is the appellant (respondent).
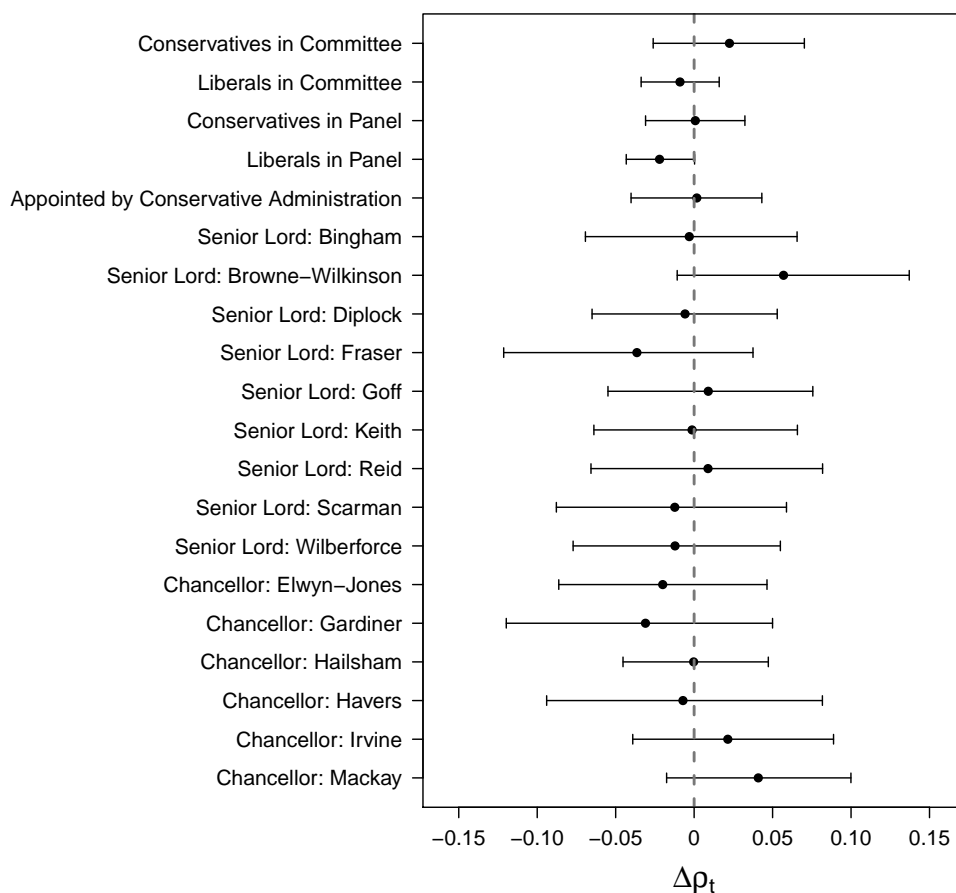
Figure 7: Influence of the committee's composition on justices' common prior beliefs. The figure plots the expected percentage change in ρ associated with changes in the predictors. Circles represent posterior means, while horizontal lines give the 90% credible intervals.

This suggests that even though the House of Lords increasingly controlled its own docket, the choice of cases heard by the Appellate Committee was not contingent on the judges' politics or driven by changes in the positions of power. This marks a clear difference with previous findings for the US Supreme Court (Iaryczower and Shum, 2012).

6   Conclusion

The ideological model of judicial behavior has been used extensively and productively to analyze the voting decisions of members of the US Supreme Court. In recent years, its scope has widened to cover courts around the world. Whether the spatial voting model is

the most useful to understand decision-making in the courts, however, is less obvious. In this framework, judges purely create law, apparently unconstrained by existing legislation. This outlook clashes bluntly with the legal view of judging, in which decisions in the court are primarily about finding how the facts of the case fit into the body of the law and established legal reasoning.

In this paper, we compare the ideological model with an alternative model of judicial decision-making that incorporates features of both the legal and the attitudinal models and accounts for differences in ability and ideology among justices. We show that this alternative model of judicial behavior explains the decisions of the Lords of Appeal remarkably well, and improves the fit of the ideological model.

The estimates of the learning model allow us to tackle a number of interesting questions about the role of preferences and information in the Appellate Committee. First, we show that even after controlling for differences in ability, the Law Lords are quite moderate overall, in the sense that their biases do not impose overwhelming informational hurdles to either dismiss or allow appeals. In fact, we show that the Law Lords are generally open to change their initial stance based on the facts of the case and on how the law applies to it. This occurs more often in commercial law cases and least so in criminal appeals. Although judges' biases are seemingly related to the party which appointed them, there is substantial overlap in the distribution of individual preferences between Conservative and Labour nominees, and Lords' own political or ideological views have little influence on their propensity to affirm or dismiss appeals. We also find that committee members are quite heterogeneous in their ability to map the law to the specifics of the case under consideration, and that judicial experience is correlated with judges' ability.

Altogether, our results indicate that both differences in preferences *and* ability are useful to understand the judicial function of the House of Lords. Law Lords are not either merely finding how the facts of the case fit into the body of the law, nor freely creating law to match their individual biases. Instead, their decisions are shaped by an evolving balance between information and preferences, which reflects the power of the facts in each case to override ideological considerations. Although we have focused our analysis on the Appellate Committee, we believe that this tradeoff between preferences and information is a defining characteristic of judicial decision-making.

References

Achen, Christopher H. 2002. "Toward a new political methodology: Microfoundations and ART." *Annual Review of Political Science* 5(1):423–450.

Alarie, Benjamin R. and Andrew Green. 2007. "The Reasonable Justice: An Empirical Analysis of Frank Iacobucci's Career on the Supreme Court of Canada." *University of Toronto Law Journal* 57(2):195–226.

Arvind, Thiruvallore T. and Lindsay Stirton. 2012. "Lawyers and the Legal Model: Judicial Ideology, Judicial Professionalism and Institutional Strategy Among the Law Lords." Paper Presented at the 16th Annual Conference of the International Society for New Institutional Economics, University of Southern California, Los Angeles, CA, 14–16 June.

Bafumi, Joseph, Andrew Gelman, David K. Park and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." *Political Analysis* 43(13):171–187.

Bingham, Tom. 2009. The Law Lords: Who Has Served. In *The Judicial House of Lords 1876-2009*, ed. Brice Dickson Louis J. Blom-Cooper and Gavin Drewry. Oxford: Oxford University Press pp. 112–121.

Blom-Cooper, Louis J., Brice Dickson and Gavin Drewry. 2009. *The Judicial House of Lords 1876-2009*. Oxford: Oxford University Press.

Blom-Cooper, Louis J. and Gavin Drewry. 2009. "Towards a System of Administrative Law: The Reid and Wilberforce Era, 1945-1982." In Louis Blom-Cooper, Brice Dickson and Gavin Drewry (eds.), *The Judicial House of Lords 1876-2009*, pp. 209–231. Oxford: Oxford University Press.

Carmichael, Paul and Brice Dickson. 1999. *The House of Lords. Its Parliamentary and Judicial Roles*. Oxford: Hart Publishing.

Clinton, Joshua D. and Simon Jackman. 2009. "To Simulate or NOMINATE?" *Legislative Studies Quarterly* 34(4):593–621.

Clinton, Joshua D., Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2):355–370.

Desposato, Scott, Matt Ingram and Osmar Lannes. 2012. "The New Politics of the Brazilian Judiciary." Unpublished Manuscript, University of California, San Diego.

Dickson, Bruce. 2007. "The processing of appeals in the House of Lords." *Law Quarterly Review* 123:571–601.

Drewry, Gavin and Louis J. Blom-Cooper. 2009. The House of Lords and the English Court of Appeal. In *The Judicial House of Lords 1876-2009*, ed. Brice Dickson Louis J. Blom-Cooper and Gavin Drewry. Oxford: Oxford University Press pp. 48–63.

Duggan, John and Cesar Martinelli. 2001. "A Bayesian Model of Voting in Juries." *Games and Economic Behavior* 37(2):259–294.

Feddersen, Timothy and Wolfgang Pesendorfer. 1997. "Voting Behavior and Information Aggregation in Elections With Private Information." *Econometrica* 65(5):1029–1058.

Garrett, Elizabeth S. and Scott L. Zeger. 2000. "Latent Class Model Diagnosis." *Biometrics* 56(4):1055 –1067.

Gelfland, Alan E. 1996. Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice*, ed. Sylvia Richardson Walter Gilks and David Spiegelhalter. Boca Raton: Chapman & Hall pp. 145–162.

Gelman, Andrew and Donald B. Rubin. 1992. "Inferences from Iterative Simulation Using Multiple Sequences." *Statistical Science* 7(4):457–511.

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin. 2004. *Bayesian Data Analysis*. Boca Raton: Chapman & Hall.

Gerhardt, Michael J. 2004. "Merit vs. Ideology." *Cardozo Law Review* 26:353–376.

Gonzalez Bertomeu, Juan F., Lucia Dalla Pellegrina and Nuno M. Garoupa. 2013. "Estimating Judicial Ideal Points in Latin America: The Case of Argentina." Illinois Program in Law, Behavior and Social Science Paper No. LBSS, 14–17.

Griffith, John A. 1977. *The politics of the judiciary*. Manchester: Manchester University Press.

Hanretty, Chris. 2013. "The Decisions and Ideal Points of British Law Lords." *British Journal of Political Science* 43(3):703–716.

Heckman, James and James Jr. Snyder. 1997. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *The RAND Journal of Economics* 28:S142–S189.

Herron, Michael C. 1999. "Postestimation uncertainty in limited dependent variable models." *Political Analysis* 8(1):83–98.

Huang, Guan-Hua and Karen Bandeen Roche. 2004. "Building an identifiable latent class model with covariate effects on underlying and measured variables." *Psychometrika* 69(1):5–32.

Iaryczower, Matias, Gabriel Katz and Sebastian Saiegh. 2013. "Voting in the Bicameral Congress: Large Majorities as a Signal of Quality." *Journal of Law, Economics and Organization* 29(5):957–991.

Iaryczower, Matias, Garrett Lewis and Matthew Shum. 2013. "To Elect or to Appoint? Bias, Information, and Responsiveness of Bureaucrats and Politicians." *Journal of Public Economics* 97:230–244.

Iaryczower, Matias and Matthew Shum. 2012. "The Value of Information in the Court. Get it Right, Keep it Tight." *American Economic Review* 102(1):202–237.

Laski, Harold J. 1925. "Technique of judicial appointment." *Michigan Law Review* 24(6):529–543.

Malleson, Kate. 2009. Appointments to the House of Lords: Who Goes Upstairs. In *The Judicial House of Lords 1876-2009*, ed. Brice Dickson Louis J. Blom-Cooper and Gavin Drewry. Oxford: Oxford University Press pp. 112–121.

Martin, Andrew and Kevin Quinn. 2002. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953-1999." *Political Analysis* 10(2):134–153.

Martin, Andrew and Kevin Quinn. 2007. "Assessing Preference Change on the US Supreme Court." *Journal of Law, Economics and Organization* 23(2):365–385.

Ntzoufras, Ioannis. 2011. *Bayesian modeling using WinBUGS*. Vol. 698. Hoboken: John Wiley & Sons.

Paterson, Alan. 1982. *The Law Lords*. Toronto: University of Toronto Press.

Paterson, Alan. 2013. *Final Judgment. The Last Law Lords and the Supreme Court.* Oxford: Hart.

Poole, Keith and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29(2):357–384.

Robertson, David. 1982. "Judicial ideology in the House of Lords: A jurimetric analysis." *British Journal of Political Science* 12(1):1–25.

Robertson, David. 1998. *Judicial discretion in the House of Lords*. Oxford: Clarendon Press.

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez and Markus Müller. 2011. "pROC: an open-source package for R and S+ to analyze and compare ROC curves." *BMC Bioinformatics* 12(77):77.

Segal, Jeffrey A and Albert D Cover. 1989. "Ideological values and the votes of US Supreme Court justices." *American Political Science Review* 83(2):557–565.

Segal, Jeffrey A. and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model revisited*. Cambridge: Cambridge University Press.

Segal, Jeffrey A., Lee Epstein, Charles M. Cameron and Harold J. Spaeth. 1995. "Ideological values and the votes of US Supreme Court justices revisited." *The Journal of Politics* 57(3):812–823.

Stacia, L. Haynie, Reginald S. Sheehan, Donald R. Songer and C. Neal Tate. 2011. "High Courts Judicial Database. Version 1.2.". Available from http://sitemason.vanderbilt.edu/site/d5YnT2/data_sets, accessed 1 October 2013.

Stephens, Matthew. 2000. "Dealing with Label Switching in Mixture Models." *Journal of the Royal Statistical Society, Series B (Methodological)* 62(4):795–809.

Stevens, Robert. 1978. *Law and Politics: The House of Lords as a Judicial Body, 1800-1976.* Chapel Hill: University of North Carolina Press.

More than Politics: Ability and Ideology in the British Appellate Committee

Supplementary Materials Appendix.

Matias Iaryczower   Gabriel Katz[*]

Contents

[*]Department of Politics, Princeton University, miaryc@princeton.edu and Department of Politics, University of Exeter, G.Katz@exeter.ac.uk

S.1 Descriptive statistics

Table S.1: Descriptive statistics for the judge-specific covariates

Covariate	Mean	Std. Dev.	Range
Political Experience: Conservative	0.17	0.38	0 - 1
Political Experience: Liberal	0.07	0.26	0 - 1
Appointed by Conservative Administration	0.69	0.47	0 - 1
Judicial Experience (years)	12.69	5.72	0 - 25
English	0.67	0.48	0 - 1

Table S.2: Descriptive statistics for the case-specific covariates

Covariate	Mean	Std. Dev.	Range
Appellant: State	0.26	0.44	0 - 1
Respondent: State	0.40	0.49	0 - 1
% of cases received from the Court of Appeal	0.96	0.20	0 - 1
% conservatives in the committee	0.16	0.17	0 - 1
% of liberals in the committee	0.02	0.05	0 - 1
% of conservatives in the panel	0.19	0.20	0 - 1
% of liberals in the panel	0.03	0.08	0 - 1
% of committee members appointed by a Conservative Administration	0.67	0.21	0 - 1

Note: Descriptive statistics based on the sample coding votes as liberal ($v_{i,t} = 1$) or conservative ($v_{i,t} = 0$). Summary statistics for the case-specific covariates are similar for the sample using lower-court deference as outcome.

S.2 Additional details of the MCMC algorithm

We describe the MCMC algorithm for estimating the learning model applied in Section 5.2 of the paper. Starting with initial values for the parameters $\Theta = (\omega, \alpha, \beta, \delta, \sigma^2)$, the algorithm iterates through the following steps until convergence (as determined by MCMC diagnostics):

1. Updating α from its full conditional distribution

$$p(\alpha|\Theta_{-\alpha}) \propto \prod_{t=1}^{T} Pr(\omega_t = 1|\alpha)^{I(\omega_t=1)} p(\alpha)$$

$$= \prod_{t:\omega_t=1} \left(\frac{\exp(X_t'\alpha)}{1 + \exp(X_t'\alpha)} \right) \mathbf{N}(\alpha, 0, (9/4)I)$$

   where $\rho = \frac{\exp(X_t'\alpha)}{1+\exp(X_t'\alpha)}$ and the prior $p(\alpha)$ is a multivariate normal distribution evaluated at α. This conditional distribution does not have a closed form. However, α can be updated through a random-walk Metropolis step with a multivariate Student-$t_3(s_\alpha A)$ proposal, using the empirical covariance matrix of α from an extended burn-in period to tune A and improve mixing (Haario et al. 2005) and adjusting the scaling parameter s_α to achieve an acceptance rate of $\approx 25\%$ (Robert and Casella, 2010).

2. Updating ω_t from its full conditional multinomial or "categorical" distribution, $\omega_t|\Theta_{-\omega_t} \sim Cat(p_{t,1}, p_{t,2})$, with

$$p_{t,1} = \frac{\rho \prod_{i=1}^{N} \gamma_{i,1}^{v_{it}} (1 - \gamma_{i,1})^{(1-v_{it})}}{\rho \prod_{i=1}^{N} \gamma_{i,1}^{v_{it}} (1 - \gamma_{i,1})^{(1-v_{it})} + (1 - \rho) \prod_{i=1}^{N} \gamma_{i,0}^{v_{it}} (1 - \gamma_{i,0})^{(1-v_{it})}}$$

3. Updating β from the posterior distributions

$$p(\beta_j|\Theta_{-\beta_j}) \propto \prod_{t:t=j} \prod_{i=1}^{n} \gamma_{i,1}^{v_{it}} (1 - \gamma_{i,1})^{(1-v_{it})} N(\beta_j, 0, (9/4)I) \qquad j = 0, 1$$

   using random walk Metropolis steps with a multivariate Student-$t_3(s_\beta B_j)$ proposal, basing the scale matrix on the empirical covariance from an extended burn-in period and adjusting s_β to achieve an appropriate acceptance rate.

4. Updating $\delta_{i,j}$ from

$$p(\delta_{i,j}|\Theta_{-\delta_{i,j}}) \propto \prod_{t:\omega_t=j} \gamma_{i,1}^{v_{it}}(1-\gamma_{i,1})^{(1-v_{it})} N(\delta_{i,j},0,\sigma_j^2) \qquad i=1,\ldots,n, j=0,1$$

using a Student-t_3 proposal density. Conditional on ω_t, the acceptance ratio is a function of β_j and σ_j^2 only.

5. Updating σ_j^2 from the posterior Inverse-Gamma$\left(0.1 + n_j/2, 0.1 + \dfrac{\sum_i \delta_{i,j}^2}{2}\right)$ distribution, where n_j is the number of judges deciding cases for which $\omega_t = j$, $j = 0, 1$.

4

S.3 Additional goodness-of-fit measures
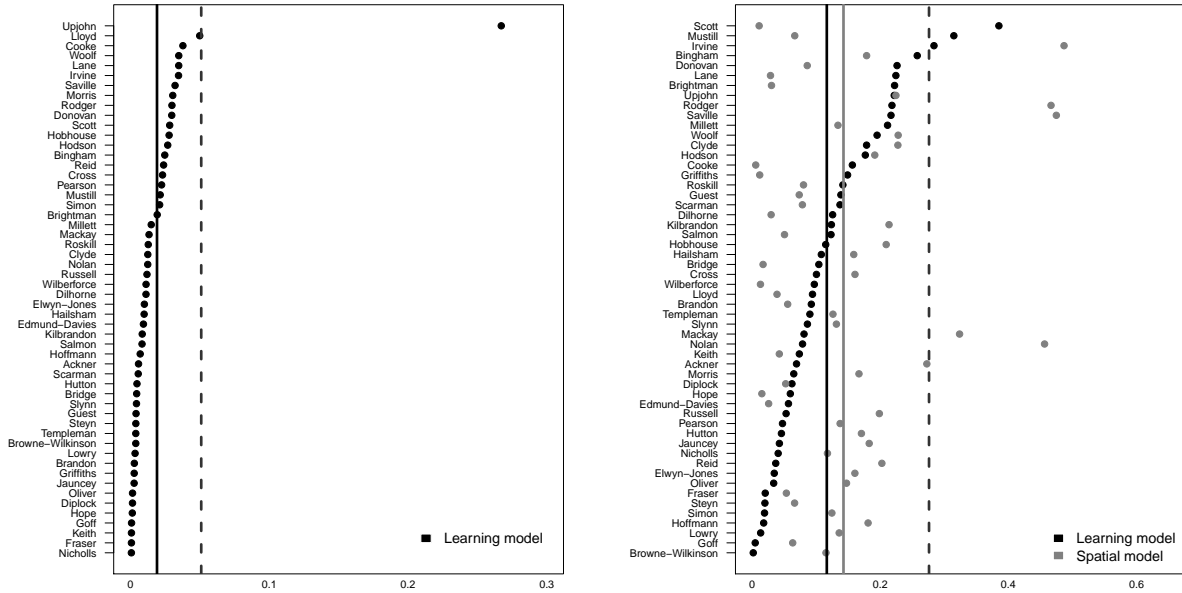
Figure S.1: Excess error rates in the analysis of liberal/conservative decisions. The left panel plots the realized error rates of the learning model fitted to the full sample. The right panel compares the error rates for the learning and ideological models fitted to the sample including only non-unanimous decisions. Filled circles correspond to the absolute excess error rate per justice averaged over 1,000 posterior draws. Solid vertical lines represent the (absolute) average excess errors across all judges under the learning and ideological models. Dashed vertical lines correspond to the average excess error for a model assuming that each judge votes with the majority in every appeal.
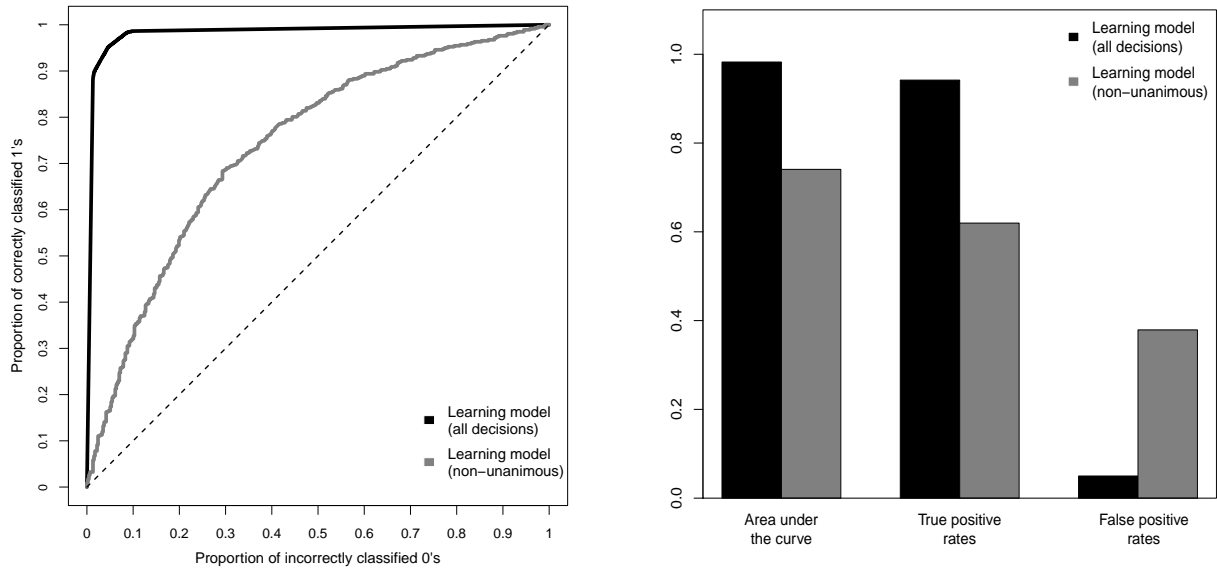
Figure S.2: Out-of-sample performance of the learning model. The left panel plots ROC curves for the learning model obtained from the out-of-sample predictive densities approximated using leave-one-out cross-validation; the 45-degree line corresponds to a random prediction model. The right panel compares the proportion of correctly predicted votes for the learning model in the full sample comprising unanimous and non-unanimous decisions and in the restricted sample including only cases with at least one dissenting opinion.
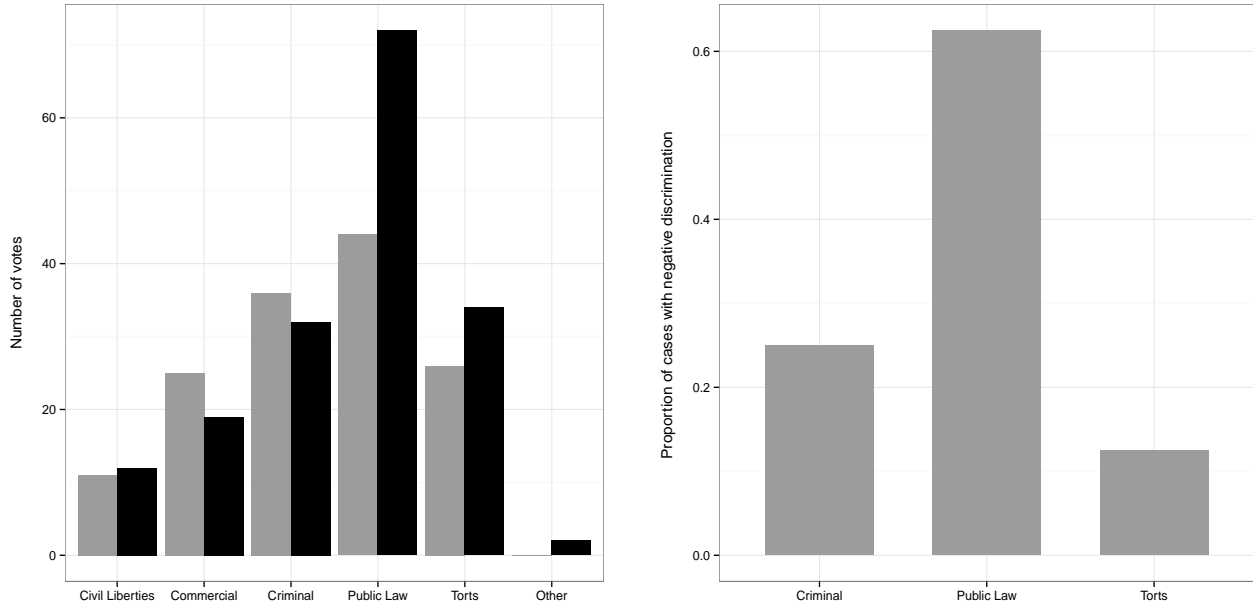
Figure S.3: Comparing the hand-coded liberal-conservative classification with the labeling recovered from the ideological model. The left panel plots the number of votes coded as conservative (gray bars) and liberal (black bars) in the *High Courts Judicial Database* that switch labels in the ideological model, for the different issue areas. These switches reflect both the sign of the discrimination parameter of the IRT model and the estimated location of cases and judges in the policy space. The right panel plots the proportion of cases for which the posterior mean of the discrimination parameter is < 0, divided by issue area. The IRT model is identified by imposing constraints on judges' ideal points (see Table 2 in the paper). Results are similar for various alternative identification procedures.

Table S.3: Measures of Complexity and Fit for the Learning and Spatial Voting Models
Overturning (upholding) lower court rulings coded as 1(0)

| | All decisions | Non-unanimous | |
	Learning Model	Learning Model	Spatial Model
Correctly predicted	0.95	0.70	0.68
PRE	0.90	0.36	0.34
ePCP	0.92	0.60	0.53
Excess error rate: by judge	0.02	0.11	0.17
Excess error rate: by case	0.07	0.10	0.18
AIC	7,572.29	1,970.86	1,966.63
BIC	17,668.18	3,098.99	3,355.49
Number of cases	1,467	223	223

Note: In the spatial voting model, unanimous decisions provide no information about the position of the Lords. Hence, we only report goodness-of-fit statistics for this model for the sample in which there is at least one dissenting opinion in every panel. The parameters of the IRT model were normalized following **?**, with discrimination assumed constant across cases.

S.4 Additional estimation results

Table S.4: Posterior Summaries for the Coefficients of Voting Probabilities Overturning (upholding) lower court rulings coded as 1(0)

Covariate	$\gamma_{i,0}$	$\gamma_{i,1}$
Political Experience: Conservative	0.13 (-0.64, 0.89)	-0.18 (-0.89, 0.55)
Political Experience: Liberal	0.09 (-0.90, 1.05)	-0.48 (-1.44, 0.56)
Appointed by Conservative Administration	-0.46 (-1.00, 0.11)	0.10 (-0.40, 0.65)
Prior Judicial Experience	-0.04 (-0.09, 0.02)	0.00 (-0.05, 0.05)
English	0.47 (-0.05, 1.03)	-0.14 (-0.67, 0.38)

Note: The table reports the posterior means of the regression coefficients in equation 7 (see Section 4.2 of the paper). 90% credible intervals in parentheses.

Table S.5: Posterior Summaries for the Coefficients the Common Prior ρ
Overturning (upholding) lower court rulings coded as 1(0)

Covariate		Mean	Std. Dev.	90% credible intervals
Appellant: State		0.48	(0.16)	(0.23, 0.75)
Respondent: State		-0.50	(0.15)	(-0.75, -0.25)
Court of Appeal		-0.30	(0.27)	(-0.74, 0.15)
Conservatives in the committee		0.54	(0.72)	(-0.63, 1.69)
Liberals in the committee		-0.77	(1.20)	(-2.90, 1.34)
Conservatives in the panel		0.01	(0.39)	(-0.63, 0.65)
Liberals in the panel		-1.19	(0.72)	(-2.36, 0.01)
Appointed by Conservative Admin.		0.03	(0.52)	(-0.83, 0.88)
Senior Law Lord				
	Bingham	-0.01	(0.19)	(-0.32, 0.31)
	Browne-Wilkinson	0.24	(0.21)	(-0.07, 0.60)
	Diplock	-0.02	(0.17)	(-0.31, 0.26)
	Fraser	-0.16	(0.23)	(-0.55, 0.18)
	Goff	0.04	(0.20)	(-0.27, 0.36)
	Keith	0.00	(0.20)	(-0.32, 0.32)
	Reid	0.04	(0.21)	(-0.30, 0.37)
	Scarman	-0.05	(0.22)	(-0.40, 0.28)
	Wilberforce	-0.05	(0.19)	(-0.37, 0.26)

Covariate	Mean	Std. Dev.	90% credible intervals
Lord Chancellor			
Elwyn-Jones	-0.09	(0.21)	(-0.44, 0.24)
Gardiner	-0.14	(0.26)	(-0.61, 0.26)
Hailsham	-0.01	(0.16)	(-0.27, 0.25)
Havers	-0.04	(0.27)	(-0.47, 0.40)
Irvine	0.09	(0.20)	(-0.24, 0.43)
Mackay	0.17	(0.19)	(-0.14, 0.50)
Issue Area			
Civil Liberties	0.08	(0.19)	(-0.23, 0.40)
Commercial	0.13	(0.17)	(-0.13, 0.42)
Criminal	-0.02	(0.16)	(-0.28, 0.25)
Family	0.12	0.22	(-0.23, 0.51)
Public Law	0.00	0.16	(-0.26, 0.27)
Torts	0.03	0.17	(-0.23, 0.32)
Other	-0.37	0.31	(-0.91, 0.05)

Note: The table reports the posterior means of the regression coefficients in equation 6 (see Section 4.2 of the paper).
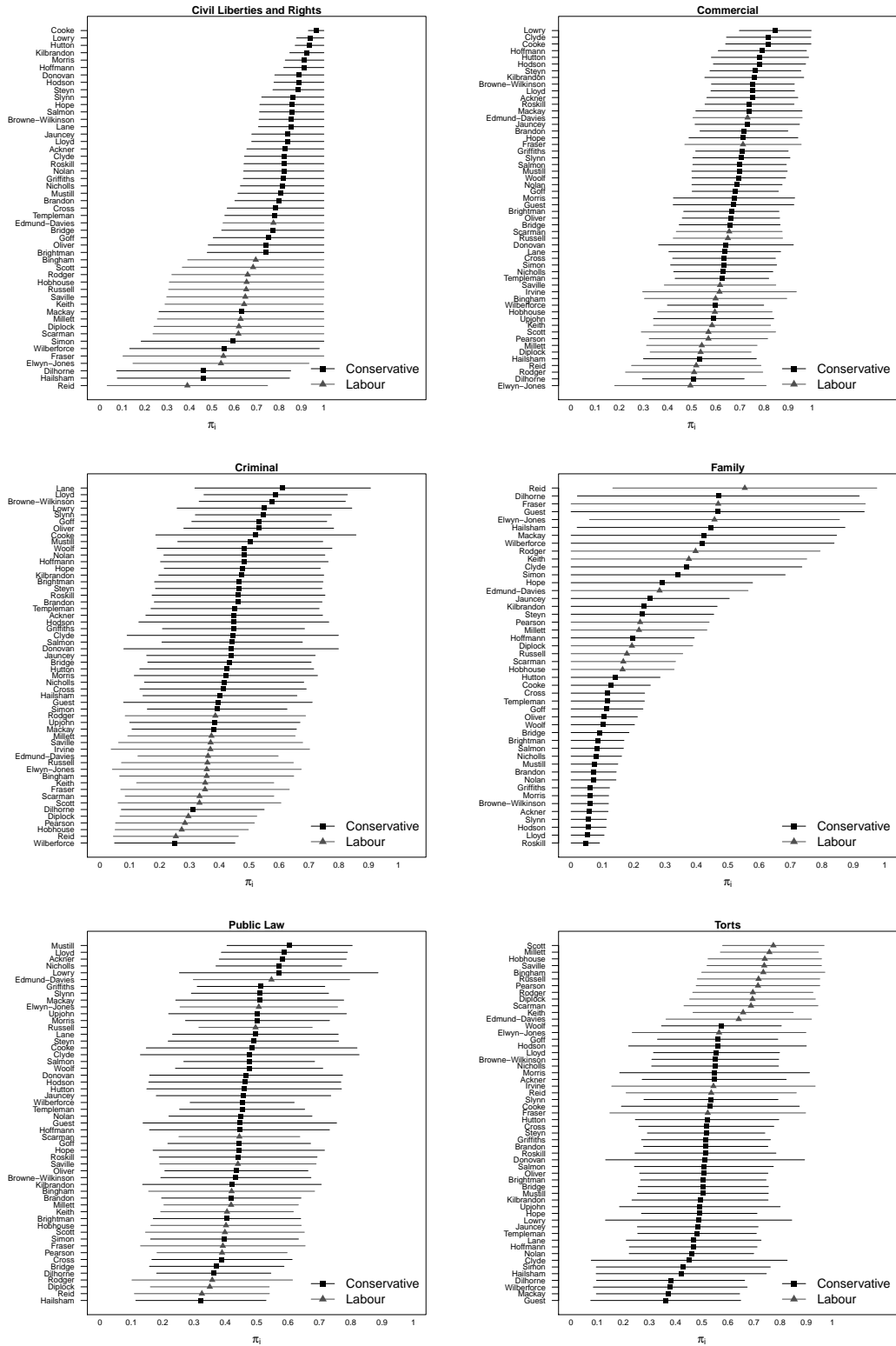
Figure S.4: Individual biases, by issue area. The figure provides posterior summaries of π_i for different areas of the law. Filled markers represent the posterior means, while horizontal lines give the 90% credible intervals.
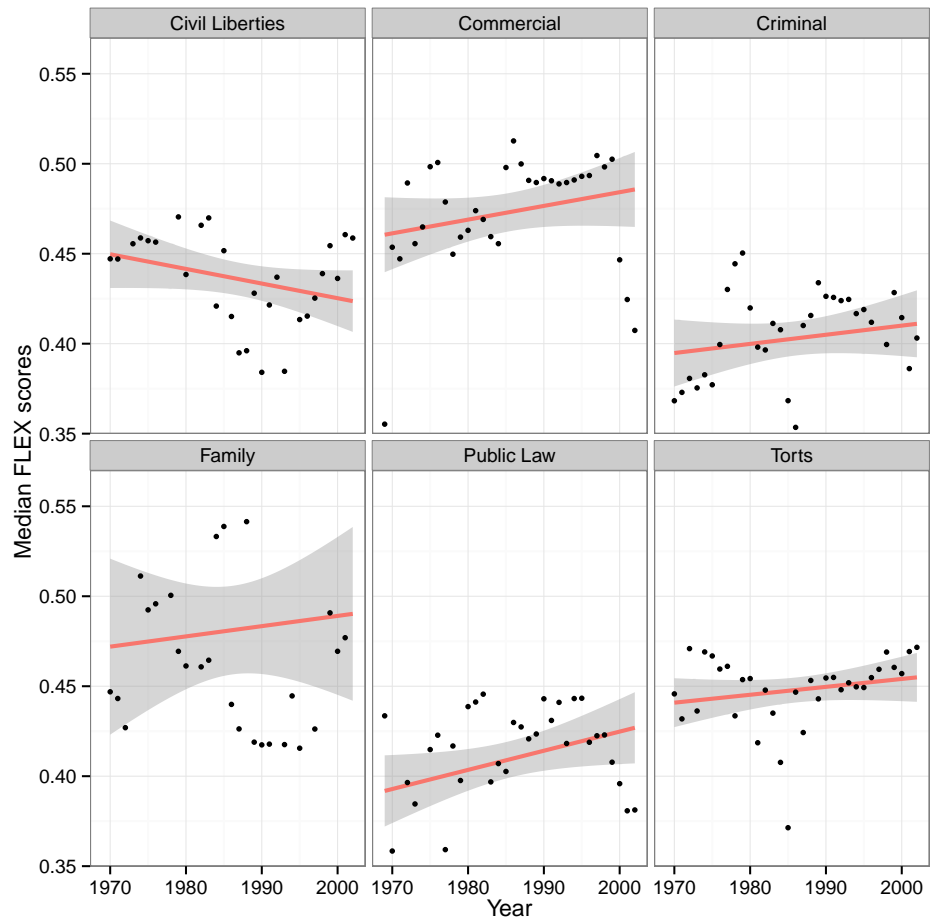
12

Figure S.5: Evolution of FLEX scores, by year and issue area. Circles represent the median FLEX score for each year, solid lines the smoothed temporal trends, and shaded areas the 90% credible intervals.
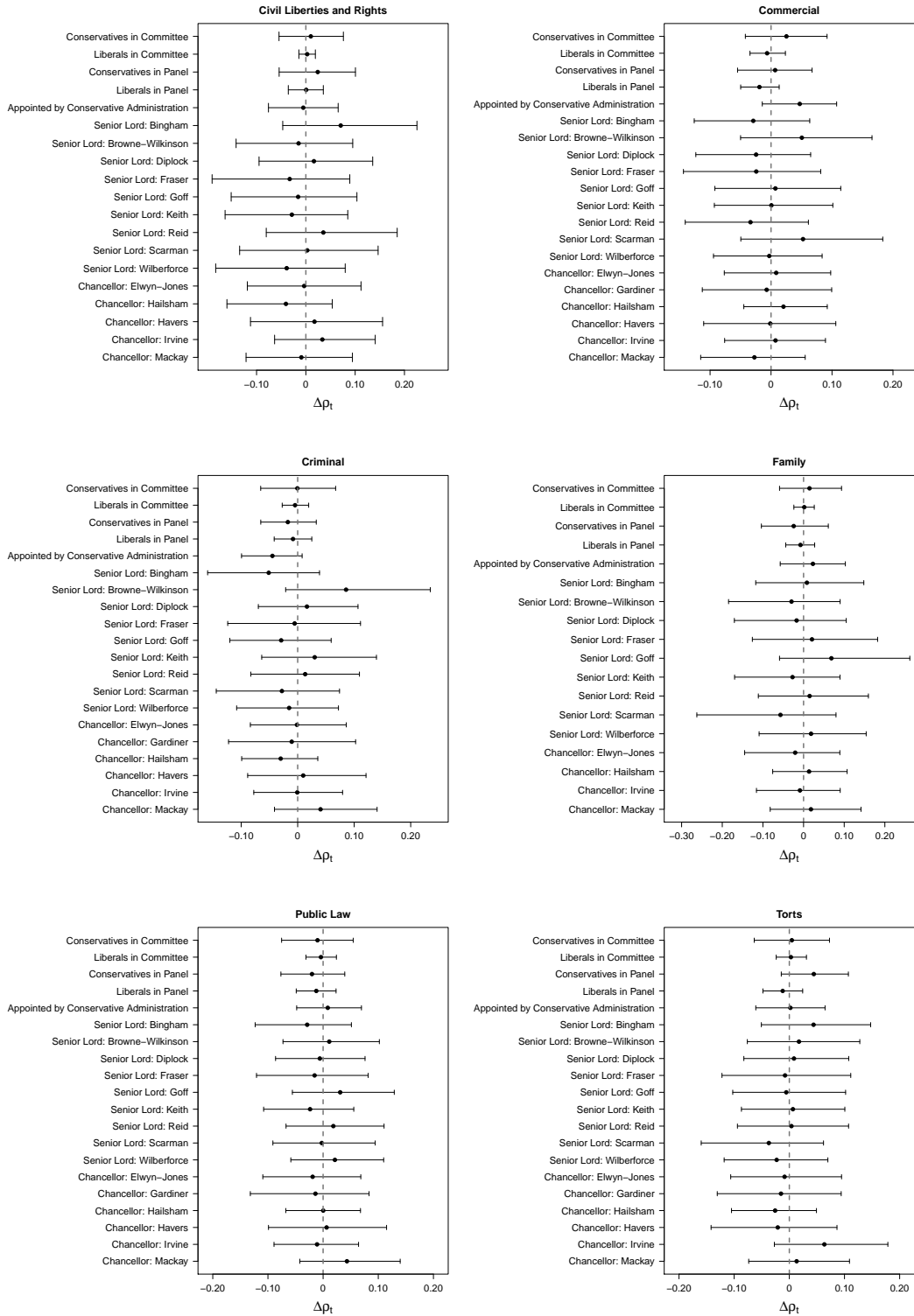
Figure S.6: Influence of the committee's composition on judges' common prior beliefs, discriminated by issue area. The figure plots the expected change in ρ associated with changes in each covariate, for all the issue areas. Solid circles correspond to point estimates (posterior means); horizontal lines give the 90% credible intervals.

S.5 Accounting for interdependence among judges

In the text we assumed that each judge cares only about his own vote - i.e., Lord i's goal is to decide in consonance with his own best understanding of how the law applies to the particulars of the case. This seems to be a plausible view of the functioning of the Appellate Committee, at least according to **?**.[1]

However, if judges care about the outcome of the case, new strategic considerations come into play. Since any vote outcome in which i is not pivotal to the decision of the court is not relevant to his payoffs, i will effectively choose the direction of his vote *as if* he was in fact pivotal.[2] In this case, judge i's relevant information in case t is not only his private information s_{it}, but also the equilibrium information contained in the event that he is pivotal for the panel's decision, given the equilibrium strategy profile followed by the remaining panel members. The relevant equilibrium condition in this case is

$$\frac{\sum_{C \in \mathcal{C}_{R-1}} \left(\prod_{j \in C}[1 - \Phi(\theta_{jt}[s_{jt}^* - 1])]\right) \left(\prod_{j \neq i, j \notin C} \Phi(\theta_{jt}[s_{jt}^* - 1])\right)}{\sum_{C \in \mathcal{C}_{R-1}} \left(\prod_{j \in C}[1 - \Phi(\theta_{jt} s_{jt}^*)]\right) \left(\prod_{j \neq i, j \notin C} \Phi(\theta_{jt} s_{jt}^*)\right)} \frac{\phi(\theta_{it}[s_{it}^* - 1])}{\phi(\theta_{it} s_{it}^*)} = \frac{\pi_{it}}{1 - \pi_{it}} \frac{1 - \rho_t}{\rho_t}$$

where $\mathcal{C}_{R-1}^i$ is the the set of coalitions $C \subset n \setminus i$ with $R - 1$ members, $R = (n + 1)/2$.

As noted in the text, the estimates of the voting strategies s_{it}^* and of the precision of judges' information θ_{it} are unchanged with respect to the baseline specification. The only difference with the baseline model lies in the estimate of the preference parameters π_{it}. The relative ordering of justices' preference parameters in this strategic version of the model are reported in Figure S.7.

---

[1]Robertson writes that "Law Lords make their law as individuals, being satisfied as long as the argument they construct, or consent to, satisfies their own sense of legal correctness" (**?**, p. 16).

[2]This assumption is not correct when i is not pivotal, but precisely because of this such mistakes have no cost for the outcome-oriented justices.

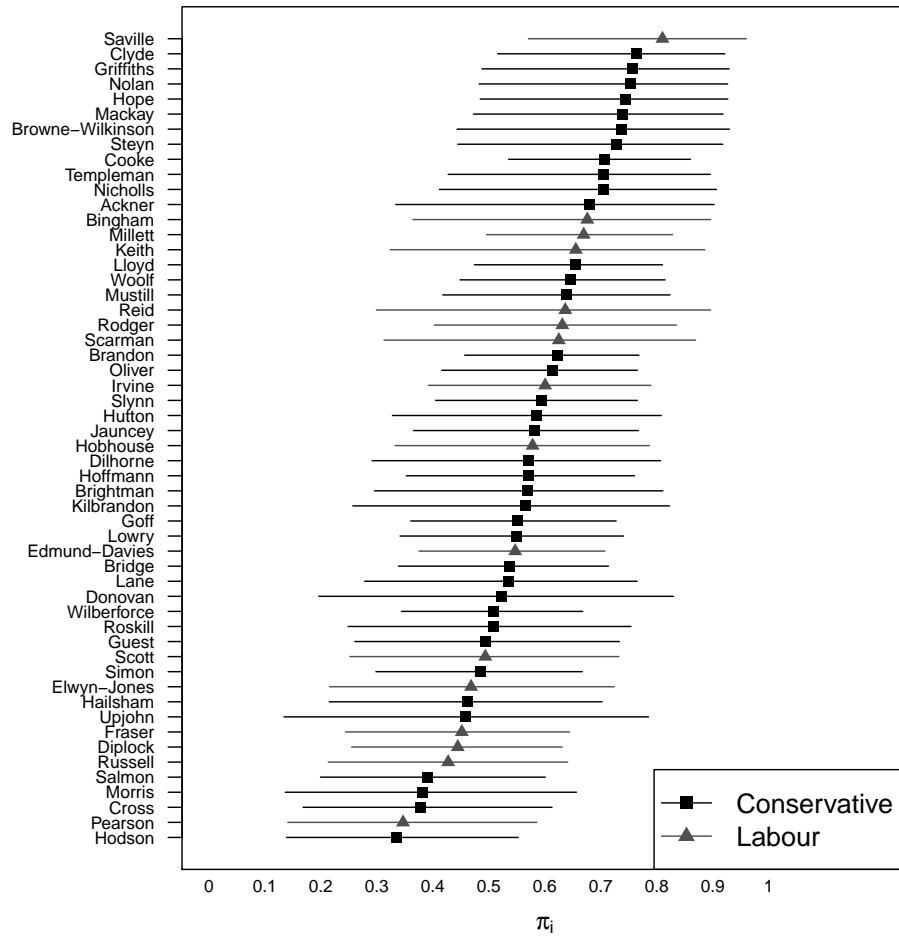**Figure S.7: Judges' Bias Parameters in the Strategic Voting Model.** The figure provides posterior summaries for π_i assuming interdependence between the decisions of the judges hearing each case. Circles represent posterior means, and horizontal lines the 90% credible intervals.