RAND Journal of Economics

Careerist judges and the appeals process

Gilat Levy*

I analyze how careerist judges formulate their decisions using information they uncover during deliberations as well as relevant information from previous decisions. I assume that judges have reputation concerns and try to signal to an evaluator that they can interpret the law correctly. If an appeal is brought, the appellate court's decision reveals whether the judge interpreted the law properly and allows the evaluator to assess the judge's ability. The monitoring possibilities for the evaluator are therefore endogenous, because the probability of an appeal depends on the judge's decision. I find that judges with career concerns tend to be "creative," i.e., to inefficiently contradict previous decisions.

1. Introduction

■ Judging by the surge in recent articles counting the number of times a judge's opinion or article is cited (or web-searched), reputation, influence, prestige, and career concerns are essential features of the judicial world.[1] This is, of course, nothing new; but as ways of measuring features such as prestige or influence have became more sophisticated, there has been a renewed interest in judicial reputation. As Posner (2000, p. 392) writes,

> An even more audacious use of citations as a judicial management tool is to grade appellate judges . . . the ranking is a rough guide to quality, or influence, or reputation—it is not altogether clear which is being measured.

Judges may care about how others perceive their quality (and accordingly rank them) for two reasons. First, they may have a human concern about their prestige and influence. In this sense, the judicial world is similar to the academic world. The second, more interesting, reason is that the perception of quality can influence one's career. Although judges who have life-tenured positions need not fear losing their jobs, promotion to a better position in the judicial system may depend on whether others consider them able. It is a common tradition that appellate judges are trial judges who got promoted and Supreme Court Justices are judges from lower-echelon appellate courts. For a judge, these higher-echelon positions can increase both her pay and her possibilities of influencing other judges. Thus, trial judges may desire to become appellate judges, and judges of intermediate appellate courts may aspire to become judges in the courts of last resort.

---

* London School of Economics; g.levy1@lse.ac.uk.

I thank Oriana Bandiera, Michael Birenhak, Patrick Bolton, Leonardo Felli, Gene Grossman, Alessandro Lizzeri, Zvika Neeman, Wolfgang Pesendorfer, Ronny Razin, Steven Shavell, and seminar participants in Tel Aviv University, University of Pennsylvania, New York University, and Southampton University for helpful remarks. I also thank the Editor Jennifer F. Reinganum and two anonymous referees for their insightful comments. Financial support from the ESRC, grant no. R0000223323, is greatly acknowledged.

[1] A few examples are Klein and Morrisroe (1999); Ayres and Vars (2000); Shapiro (2000); and Bhattacharya and Smyth (2001).

Empirical research finds that the perceived quality of judges plays a notable part in their promotion. This is, for example, the result in Salzberger and Fenn (1999), who analyze the promotion probability from the court of appeal to the House of Lords in England. In several common law countries (such as Canada), judges are appointed by an independent committee composed not only of legal professionals but also of selected members of the public, whose main concern may be competence. Even when the judicial appointments procedure is rife with politics and special interests, as is the procedure for appointing U.S. federal court judges, judicial quality can still play a big role. For example, Landes and Posner (1975) suggest that politicians have an interest in keeping an independent judiciary and therefore will not condition promotion on views.[2] Politicians may therefore focus on ability as the main desired trait for a judge.[3]

In this article I formalize the effect of reputation-seeking behavior on judicial decision making. In particular, I assume that a judge is interested in creating a reputation for high judicial ability, which is the ability to interpret the law correctly. Traditionally, political and legal scholars assume either that judges try to make the right decision, i.e., to interpret the law correctly (the "legal" model) or that judges have ideological preferences and follow them when adjudicating a case (the "political" model). But some have taken reputation motives more seriously; Landes and Posner (1976) conjecture that judges follow precedents to avoid the disutility of being reversed, whereas both Miceli and Coşgel (1994) and Whitman (2000) assume that judges suffer a utility loss when being overturned by others and gain utility when being cited. As opposed to these articles, I derive these motivations as well as aversions endogenously, from fundamental preferences.

In other contexts, several articles model careerist decision makers, such as managers or experts, who use their decisions to signal their abilities to the market (see, for example, Holmström, 1999; Scharfstein and Stein, 1990; Levy, 2004; and Zwiebel, 1995). However, judicial decision making has distinctive features. The assumption in the career-concerns literature is that the market eventually realizes whether managers have made the correct decision. In the judicial system, it may never be found out whether the judge's decision is correct or not. This makes the task of assessing and monitoring the ability of the judge difficult. Nevertheless, it may be possible to extract more information about the case, and consequently about the judge's ability, if an appeal is brought and the case is adjudicated once more. Thus, monitoring the ability of the judge is *endogenous*, because whether an appeal is brought depends on her particular decision. The judge herself can then influence the flow of information about the case and, incidentally, about her type.

Given the above discussion, I focus the analysis on judicial systems that allow for appellate review; the judge may be subject to review by an appeals court if the losing litigant believes that her decision is likely to be reversed.[4] An important element of the analysis is that the judge takes into consideration how her decision affects the probability that the case will be brought before a higher court. Secondly, I assume an availability of previous decisions in similar cases, i.e., nonbinding precedent, that can assist the judge in her current decision. As in Daughety and Reinganum (1999), judges in my model may learn some information from decisions of other courts, termed "persuasive influence."[5]

These features are incorporated in the following Bayesian signalling model: a judge receives some private information regarding the application of the law in a particular case. The accuracy of this information depends on her ability; the more able the judge, the more "accurate" her

[2] This accords with the study of Salzberger and Fenn, who empirically show that British politicians (when in power) do not punish candidates for the judiciary by not promoting them even if these candidates had issued decisions at odds with the politicians' ideological views.

[3] In times of a divided government in the United States, confirmation gridlock may occur, i.e., the Senate will not approve judicial candidates suggested by the President if they are perceived as "too ideological." Thus, competence may be the only common ground for different politicians to agree upon when justifying promotion. See Maltese (2003).

[4] The court system is often modelled as a hierarchy, for example in Spitzer and Talley (2000), Daughety and Reinganum (2000), Iossa and Palumbo (2004), and and Shavell (1995, forthcoming).

[5] Daughety and Reinganum (1999) assume that judges are interested in making the right decision, whereas in my model judges are careerist. Miceli and Coşgel (1994), Whitman (2000), and Rasmusen (1994) also analyze the use of nonbinding previous decisions by judges and focus on the evolution of the law. My work focuses on the efficiency of judicial decisions.

interpretation of the law. She then delivers her decision, based both on her private trial information and on past decisions. The losing party decides whether to bring an appeal, after weighing the costs of an appeal and its probability of success (the probability of reversal). If an appeal is brought, an appeals court delivers a reversal or affirmation decision. Social welfare depends positively on the probability of attaining the correct decision (through efficient aggregation of information) and negatively on the costs of appeals. A careerist judge, however, is motivated not by social efficiency but by accumulating a reputation for being able. The first task of the article is to characterize the careerist judge's decisions in equilibrium.

I show that in equilibrium, a careerist judge tends to be "creative," that is, she tends to contradict previous decisions more than an efficient judge would do. Contradicting previous decisions becomes a signal of the judge's ability, since able judges have accurate information of their own about the correct interpretation of the law and take less account of previous decisions. Since this signal increases reputation, other types of judges, and in particular less-able types, tend to use it excessively and inefficiently.[6]

However, another equilibrium feature is "reversal aversion," which arises endogenously, because reversal signals that the judge's decision was mistaken and reduces her reputation. Thus, the least-able types realize that if they contradict previous decisions they may be "caught" by the appeals court. Therefore, less-able judges cannot fully mimic the behavior of the more-able judges. This allows for an informative equilibrium even when monitoring is endogenous and the judge cares only for reputation.

The second goal of the article is to assess which institutional features can mitigate the distortive behavior of the careerist judge. In particular, I consider the effect of different judicial appointment procedures on reputation concerns and as a result on the efficiency of judicial decisions. The procedure of judges' selection is heavily debated in many countries. For example, a newly proposed reform in Britain calls for judges to be selected by an independent panel of experts, most of whom would be drawn from outside the legal profession. This is in contrast to the tradition so far, in which judges are appointed by the British government after consultation with existing judges and senior lawyers.[7]

The model provides a natural way to assess whether judges should be selected for promotion by senior judges or by nonlegal professionals. In particular, I can differentiate the two selection systems according to the information held by the evaluator: Supreme Court Justices may find it easier to identify the correct interpretation of the law, as opposed to nonprofessionals who must rely on information revealed from appeals. This difference may affect the incentives of the judge and consequently her equilibrium behavior. I therefore compare the equilibrium of the model when the judge is selected by a committee of experts from outside the legal profession, who may know what is the right decision only when an appeal is brought (*endogenous monitoring*) to the equilibrium of the model when the judge is selected by senior judges or legal professionals, who, at some relatively small cost, can learn the correct interpretation of the law independently of appeals (*exogenous monitoring*).

I find that judges behave more efficiently when monitoring is endogenous. It may therefore be welfare improving if judges are promoted by a committee of well-intentioned selected members of the public who are nonetheless outsiders to the legal profession. The intuition is that an endogenous monitoring system "punishes" judges who contradict previous decisions by a higher likelihood of an appeal, and therefore a higher likelihood of being proved wrong by the higher court. This mitigates the incentive of the less-able judges to mimic able judges who contradict previous decisions, and reduces the distortion. This cannot happen with exogenous monitoring because then judges would be proved right or wrong disregarding their decision.

[6] The results of the model thus differ from those of Daughety and Reinganum (1999), who predict that judges may engage in inefficient herding, that is, they excessively follow previous decisions. The reason is that in their analysis, judges are interested in making the right decision (or what the Supreme Court perceives as the right decision), whereas in my model judges are careerist and engage in active signalling.

[7] The reform is suggested by the Commission for Judicial Appointments, which advises the government on legal matters.

The rest of the article is organized as follows. In Section 2 I present the model, a Bayesian signalling game with incomplete and asymmetric information about the state of the world (i.e., the correct interpretation of the law). Section 3 states the main results; I first analyze a benchmark model in which the judge behaves efficiently and then investigate the equilibrium behavior of the careerist judge. A comparison between the two types of judges follows. In Section 4 I analyze the effect of different judicial appointment procedures on the equilibrium outcomes and also explore the effect of binding precedent. Section 5 concludes by discussing some extensions of the model, such as possible concerns of judges to prove, on top of their competence, that they have the "right" preferences. All proofs not in the text are relegated to the Appendix.

2. The model

■ **Players and actions.** The model describes a two-tier hierarchy of a judicial process, formed of a lower-court judge and a higher-court judge. The lower court and the higher court may represent either two appellate courts (intermediate and a court of last resort) or a trial court and an appellate court.[8]

The lower-court judge J must make a dichotomous decision, $d \in \{y, n\}$, namely whether to accept the plaintiff's argument ($d = y$) or to reject it ($d = n$). Many judicial decisions are dichotomous in nature. Also, each decision may be viewed as a collection of binary decisions, i.e., whether some evidence is valid or not. Thus, the model could be applied to any of these "minidecisions."[9] In any case, the behavior of the careerist judge would be exactly the same if we assume that her decision is a continuous one, for example if she has to issue a holding that determines the likelihood of the plaintiff's argument on a scale from 0 to 1. In equilibrium, due to career concerns, she would use only two decisions. For clarity of exposition we can then assume from the start a simple binary decision-making process.

Given the lower court's decision, the losing litigant L can advance his case to the higher court by bringing an appeal. Thus, L's action is to appeal or not. An appeal is costly; the cost is a random variable c, which is distributed uniformly on the interval [0, 1]. Each side has to bear c. For simplicity, I assume that the costs are not known prior to the lower-court judge's decision. The costs are realized by the litigants only after the decision is made. This assumption implies that the judge views appeals as uncertain. If c is known in advance, it adds another parameter to the model, but the nature of the results would still be maintained.

If L brings an appeal, the higher court H adjudicates the case. H must decide whether to affirm (A) the decision of J or to reverse it (R), i.e., $d^h \in \{A, R\}$. Define the final decision, D, as $D = d$ if $d^h = A$ or if no appeal took place, and $D = d'$ for $d' \neq d$ if $d^h = R$. The legal process ends if no appeal is brought or after the higher court's ruling.

There is one additional player in the game, the evaluator, E, who represents those that the judge would like to impress. The evaluator observes the judicial process and forms beliefs on the ability of the lower-court judge, as explained below. I analyze a one-shot game—the adjudication of one legal case. I discuss some dynamic issues in Section 5.

□ **The information structure.** The underlying state of the world, w, is initially unknown. It can be either y or n, with the interpretation that the correct decision is $d = y$ in state y and $d = n$ in state n. The state of the world w represents the underlying "truth" about the case or the correct interpretation of the law for the particular case.

In reality, there is not necessarily an "objective" correct decision or interpretation of the law. The assumption here is just an approximation for the idea that all players would agree, if they have the same information on the specific case, on what is the best holding. Actually, the results

[8] Daughety and Reinganum (2000) analyze a judicial system in which a trial court is engaged in fact finding whereas the appellate court interprets the law. They reject the idea that a trial court is Bayesian and offer an axiomatic decision-making approach.

[9] On the binary nature of judicial decisions, see Kornhauser (1992).

hold even if players have different views about the correct decision, as long as there is common knowledge about the preferences of the players.[10]

Earlier decisions by other courts in similar cases can provide information about w ("persuasive influence"). Assume, without loss of generality, that the body of previous decisions indicates that the correct decision is y and is accurate with probability $q \in (.5, 1)$. The prior belief of the players about the state of the world is therefore that $\Pr(w = y) = q$.

The modelling of previous decisions as imperfect information about the current case has several interpretations. First, the current case may be only partially similar to previous cases. The parameter q can measure then the degree of similarity between cases. Second, norms, conventions, and other conditions may have changed, and q may reflect the degree of relevance of past decisions to the current case.

For simplicity, I assume that the information about earlier decisions is common knowledge. The players may differ, however, in the amount of additional private information they possess about w.

The judge's information. While adjudicating the case, the judge receives a private signal $s \in \{y, n\}$ about the correct interpretation of the law. The signal may comprise information that she receives from witnesses and lawyers, as well as some "hard" evidence. The accuracy of the signal depends on the ability of the judge to interpret the law or to understand the evidence correctly. Let $\Pr(s = w \mid w) = t$ for $t \in [.5, 1]$. For example, if the judge's ability is $t = .5$, her signal is not informative about the true state of the world, whereas the higher is t, the more accurate is the signal.

The judge knows her own ability t; all the other players in the game know neither s nor t. The other players know only how the signal s is generated given t and w, and the prior distribution over t, which is assumed to be uniform over [.5, 1].

Given the prior q, and her own information (s, t), the judge forms the following posterior beliefs, according to Bayes' rule:

$$\Pr(w = y \mid s, t, q) = \begin{cases} \dfrac{tq}{tq + (1-t)(1-q)} & \text{if } s = y, \\ \dfrac{(1-t)q}{(1-t)q + t(1-q)} & \text{if } s = n, \end{cases} \tag{1}$$

where $\Pr(w = n \mid s, t, q) = 1 - \Pr(w = y \mid s, t, q)$.

The litigants' information. The information possessed by the litigants when they decide whether to appeal includes the judge's decision (d), the prior (q), and the cost (c). I assume that the litigants do not have private information about the state of the world. Although this is not very realistic, assuming that litigants have some information would just complicate the model without changing the qualitative results.

The higher court's information. The higher court H knows the judge's decision (d) and the prior (q). When the case is brought for an appeal, the higher court also receives some private information—a signal—about w. For simplicity, I assume that H's signal is perfect, i.e., he learns w and can make a fully informed reversal or affirmation decision once the case is brought for an appeal. The results are maintained, although the analysis is more complicated, if H receives an imperfect signal. It is also consistent, within the context of the model, to assume that H is on average more talented than J, since the higher-court judges are able judges who are promoted from lower courts.

The information of the evaluator. E's action is to form beliefs about the expected ability of the judge, given his prior belief (a uniform distribution). E also knows the prior about the case (q)

[10] When there is no common knowledge of preferences, judges may also attempt to signal their ideological views. The model is robust to the introduction of such concerns if they are weak enough (see the discussion in Section 5).

and the decision (d) of the judge. Finally, E can glean information from the judicial process. That is, when an appeal is brought, E can observe $d^h \in \{A, R\}$. This implies that E's information about w is endogenous. E can learn w only when an appeal is brought, an event that depends on the judge's and the litigants' behavior. Denote E's posterior beliefs about t by τ.

The evaluator represents those that the judge would like to impress; it can be the public, politicians, or the committee in charge of selecting judges for promotion. It is therefore assumed that these committee members, although smart and well-intentioned, are not legal experts. They do not have an independent knowledge about the interpretation of the law (or, it is too costly for them to acquire it) and can learn more about the case only after an appeal, when public opinions are published by the Supreme Court, for example.

To summarize the information structure, let us reconsider what is a judicial case in the model. Judicial cases may differ in the model only with respect to the information that they provide. In particular, they are currently distinguished only with respect to the parameter q, that is, the strength of previous decisions and evidence applying to them. The model is therefore general enough to encompass the possibility that the cases selected for trial versus settlement are biased in some sense. For example, such cases may be characterized by an initial low level of commonly known information. This would simply be reflected by one particular value or a biased set of values of the parameter q. More generally, cases may also be differentiated through the judge's private information, i.e., given a judge of a particular talent, some cases may provide more accurate signals than others. This can be captured by different functions describing the probabilities of receiving accurate signals. The model's results are robust as long as the assumption that more-talented judges are more likely to receive accurate signals is maintained.

□ **Objectives.** The judge is motivated by proving her competence. She therefore maximizes the expected beliefs about her ability t as perceived by E, and her objective function is $E(\tau)$.[11] Thus, siding with the plaintiff or the defendant is a cheap-talk action for the judge, since it has no direct bearing on her utility.

The higher court H maximizes the probability that the right decision is made (or, in other words, that the law is interpreted correctly). The assumption that the higher court is not careerist is for simplicity, in order to focus on the lower-court judge.

Let the litigants value a favorable decision at 1 and an unfavorable decision at 0. The litigants, when deciding whether to bring an appeal, evaluate the expected benefits of an appeal, that is, the probability that the decision d would be reversed, relative to its cost. Denote the probability of $d^h = R$ given a decision d and the information of L by $\Pr(R \mid d, q)$. Thus, the losing litigant perceives the function $\max\{0, \Pr(R \mid d, q) - c\}$ as his expected interim equilibrium utility and chooses whether to appeal accordingly.

I do not attribute any utility function to the evaluator. Rather, I assume that E updates his beliefs about the ability of the judge rationally, using Bayes' rule. This can be justified by the evaluator trying to promote the most-able judges.

Finally, for purposes of welfare analysis, I define the social utility function. Assume that society values a correct decision at 1 and places a weight 0 on an incorrect decision. From an *ex ante* point of view, the social utility function can be expressed as $\Pr(D = w) - 2\theta E(c)$, where $\theta \geq 0$ is a parameter capturing how much society cares about costs relative to taking the right decision and $2E(c)$ is the expected cost incurred by both sides. Society wishes therefore to maximize the probability that the final decision is correct minus its expected costs. Note that the litigants, who choose whether to appeal, do not necessarily do so according to the social utility but according to their own preferences.

□ **Timing, strategies, and equilibrium.** The structure of the game is as follows:

[11] In Zwiebel (1995), if a manager's type is perceived as too low, she is fired, whereas as long as she keeps her job, her income increases in the perception of the market on her type. Managers in his article thus care about being perceived as smart beyond some threshold, as well as about increasing their reputations. The possibility of firing seems irrelevant for the case of life-tenured judges.

Stage 1: J chooses $d \in \{y, n\}$.

Stage 2: L decides whether to appeal or not.

Stage 3: If L appeals, H takes an action $d^h \in \{A, R\}$.

Stage 4: E forms beliefs τ on t.

The strategy of J is a decision function $\sigma : (q, s, t) \to \{y, n\}$, whereas the strategy of L is a binary decision whether to appeal given q, d, and c. The strategy of H is $d^h : \{d, w\} \to \{A, R\}$. Finally, E has a belief updating function $\tau : \Omega^E \to [.5, 1]$, where Ω^E represents the information set of E. In particular, it is the prior uniform distribution over t, and either $\{q, d\}$ or $\{q, d, d^h\}$.

The equilibrium concept is that of a perfect Bayesian equilibrium. Beliefs are derived from the players' strategies and the strategies are best responses to these beliefs. I focus on informative equilibria, i.e., equilibria in which the judge's decision is contingent on her information and ignore "mirror" equilibria, in which the meaning of the actions is reversed.

3. Results

■ Let us consider first the actions of H and L in equilibrium. The optimal action of H, who is motivated by taking the correct decision, is to affirm the decision of J if $d = w$ and to reverse it otherwise.

Using backward induction, it is now easy to characterize the behavior of the losing litigant L. Anticipating the behavior of H, L knows that his appeal is successful if the decision is wrong. L therefore has to form his beliefs about whether d is right or wrong and compare them to the costs of appeal. To do so, he can update his beliefs about the state of the world given the judge's decision. Let $q_d(\sigma)$ denote the posterior probability, updated by L, that the state of the world is indeed d, given a decision d, previous decisions summarized by q, and σ, which is the conjecture of L about the strategy of J. $q_d(\sigma)$ is calculated as follows:

$$\begin{aligned} q_d(\sigma) &\equiv \Pr(w = d \mid q, d, \sigma) \\ &= \frac{\Pr(w = d \mid q) \cdot \Pr(d \mid w = d, \sigma)}{\Pr(w = d \mid q) \cdot \Pr(d \mid w = d, \sigma) + \Pr(w = d' \mid q) \cdot \Pr(d \mid w = d', \sigma)}, \end{aligned}$$

where $\Pr(w = d \mid \cdot)$ is shorthand for the probability with which w and d are the same and $\Pr(d \mid \cdot)$ is the probablility with which some decision d is made by the judge. Thus, the litigants view decision d as a signal about w with accuracy $q_d(\sigma)$. Since the probability that a decision is reversed is the probability that the judge is wrong, the litigants appeal if

$$1 - q_d(\sigma) > c.$$

Given any decision d, and the uniform distribution of costs on [0, 1], the probability of an appeal is $1 - q_d(\sigma)$, where in equilibrium, $q_d(\sigma)$ would be based on the correct conjecture of the judge's strategy. In other words, the assumptions about the appeal process imply that in equilibrium, an appeal is more likely the more the litigants believe that the judge is wrong.[12] Finally, we can compute the expected costs of an appeal, which are $E(c \mid c < 1 - q_d(\sigma)) = (1 - q_d(\sigma))/2$.[13]

□ **Benchmark: an efficient judge.** Before analyzing the equilibrium with a careerist judge, I analyze the behavior of an efficient judge. This can serve as a benchmark for the analysis.

An efficient judge adjudicates the case with the goal of maximizing social welfare, that is,

[12] Note that this type of result would follow also if the litigants would have some private information about the state of the world.

[13] When no confusion occurs, I write q_d and drop the conjecture σ.

maximizing the probability that the final decision is correct minus its expected costs. Note that when the judge is motivated only by social welfare, the evaluator plays no role in the game.

As a first step, consider what happens if no appeals are allowed and the decision of the judge is the final decision. When J rules y, her expected utility is the posterior probability that her decision is correct, $\Pr(w = y \mid q, s, t)$. Similarly, if she rules n, her expected utility is $\Pr(w = n \mid q, s, t)$. Thus, J decides y for all (s, t) such that $\Pr(w = y \mid q, s, t) \geq \Pr(w = n \mid q, s, t)$ and otherwise she decides n. By Bayes' rule, she decides y whenever $s = y$, or when $s = n$ and $t < q$. This is because when $s = n$ and $t = q$, $\Pr(w = y \mid q, s, t) = \Pr(w = n \mid q, s, t)$, since the judge's private information exactly offsets the prior.

This type of behavior can be summarized by a cutoff-point strategy, with a cutoff point (s^e, t^e), so that when $\Pr(w = y \mid q, s, t) \geq \Pr(w = y \mid q, s^e, t^e)$, the judge rules y and otherwise she rules n, where the index e in (s^e, t^e) stands for *efficiency* (for example, if no appeals are allowed, then the efficient judge uses the cutoff strategy $s^e = n$ and $t^e = q$). We can observe a cutoff strategy in Figure 1, which will accompany us throughout the analysis. The right part of the graph describes the judge's decision when $s = n$, for t ranging from .5 to 1. The left part of the graph describes the judge's decision when $s = y$, and t ranges from .5 (in the middle) to 1 (on the left). Thus, as we go from left to right, $\Pr(w = n \mid s, t)$ increases, from 0 at $s = y$ and $t = 1$, through $1/2$ at $t = 1/2$, and to 1 at $s = n$ and $t = 1$. The cutoff point, (s^e, t^e), is such that for all information (s, t) to the right of it, J rules n, whereas for all information (s, t) to its left, J rules y. Figure 1 describes then an example of a cutoff-point strategy for the judge, with $s^e = n$.

I can now analyze the more complicated case in which the judicial system allows for appeals, as the model assumes. The efficient judge understands the structure of the judicial system and the appeals process and takes it into consideration. The judge is still interested in making a decision d that she perceives as accurate, i.e., a decision with a higher $\Pr(w = d \mid q, s, t)$. But she is also forward looking and has to weigh the costs and benefits of an appeal. If an appeal occurs, the efficient judge knows that the final decision, made by the higher court, is correct for sure. This induces her to make the decision that is appealed more often, meaning the decision that is considered *less accurate* by the litigants. But appeals are costly, and the decision that is more often challenged also wastes more resources. The next lemma shows that despite this additional complexity, the behavior of the efficient judge can still be described by a cutoff-point strategy.

Lemma 1. In equilibrium, the efficient judge uses a cutoff-point strategy (s^e, t^e); she rules $d = y$ whenever $\Pr(w = y \mid q, s, t) \geq \Pr(w = y \mid q, s^e, t^e)$, and n otherwise.

Proof. Given the expected behavior of L and H, the efficient judge maximizes $\Pr(D = w) - 2\theta E(c \mid c < 1 - q_d(\sigma))$. Suppose that the judge rules d. Her expected utility can be expressed by

$$\Pr(w = d \mid q, s, t) + \Pr(w = d' \mid q, s, t)(1 - q_d) - \theta(1 - q_d)^2.$$

The first expression represents the probability that her decision is correct, and hence the final decision would be correct whether there is an appeal or not. The second expression represents the probability that her decision is wrong but corrected by the higher court, i.e., an appeal is brought. The last expression represents the expected costs of the decision (the expected costs of an appeal multiplied by its probability).

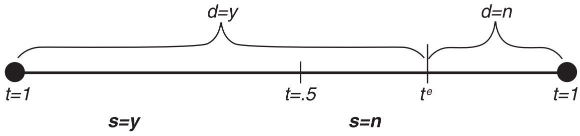
Thus, whenever the judge is indifferent between ruling y or ruling n, the above expression has to be equal for n and for y. Equating them and rearranging, I get the following condition:

$$\frac{\Pr(w = y \mid q, s, t)}{\Pr(w = n \mid q, s, t)} = \frac{q_y - \theta((1 - q_n)^2 - (1 - q_y)^2)}{q_n + \theta((1 - q_n)^2 - (1 - q_y)^2)}. \tag{2}$$

When the judge makes her decision, she takes as given the beliefs of the litigants, q_y and q_n. The right-hand side of (2) does not depend therefore on (s, t) but on the beliefs of the litigants, who have no knowledge of (s, t). The judge perceives it as constant for all (s, t). On the other hand,

FIGURE 1

A CUTOFF POINT STRATEGY WITH $s^e = n$



by Bayesian updating,

$$\frac{\Pr(w = y \mid q, s, t)}{\Pr(w = n \mid q, s, t)} = \begin{cases} \dfrac{tq}{(1-t)(1-q)} & \text{for } s = y \\ \dfrac{q(1-t)}{t(1-q)} & \text{for } s = n. \end{cases} \tag{3}$$

Hence, any different (s, t) yields a different value of $\Pr(w = y \mid q, s, t)/\Pr(w = n \mid q, s, t)$. This implies that there is (at most) a unique (s^e, t^e) that satisfies equation (2). Thus, there is a unique cutoff point (s^e, t^e), such that the judge rules y if and only if $\Pr(w = y \mid q, s, t) \geq \Pr(w = y \mid q, s^e, t^e)$. *Q.E.D.*

Equilibrium means that (2) is satisfied, subject to $0 \leq q_d \leq 1$ for $q_d(\sigma) = q_d(s^e, t^e)$, i.e., correct beliefs of the litigants. The parameters characterizing the equilibrium are $\{q, \theta\}$. Proposition 1 characterizes the equilibrium in part (i) and presents comparative statics analysis in part (ii):

Proposition 1 (i) When J maximizes social welfare, there is a unique informative equilibrium in which $s^e = n$. That is, J decides $d = y$ when $s = y$ or when $s = n$ and $t < t^e(q, \theta)$.

(ii) The cutoff point $t^e(q, \theta)$ increases in q and in θ, $t^e(q, \theta)_{q \to \frac{1}{2}} \to \frac{1}{2}$, $t^e(q, \theta)_{q \to 1} \to 1$, $t^e(q, \theta)_{\theta \to 0} < q$, and $t^e(q, \theta)_{\theta \to \infty} \to \tilde{t}(q)$, where $\tilde{t}(q)$ is a cutoff point that induces an equal probability of appeal for both decisions. For all parameters, an appeal is more likely when the judge contradicts previous decisions.

To understand the intuition for the equilibrium behavior of the efficient judge, let us consider the case in which $\theta = 0$, i.e., the judge (or society) does not care about the costs of appeal.

Recall that when no appeals are allowed, the equilibrium behavior has $s^e = n$ and $t^e = q$, and the judge simply issues a holding for the decision she believes most likely to be correct. Suppose that the judge were to use this cutoff point when appeals are actually allowed. Then, as the Appendix shows, $q_y > q_n$. That is, the litigants would believe that a decision that follows previous ones ($d = y$) is more likely to be correct compared with a decision that contradicts previous ones ($d = n$). This is a consequence of Bayesian updating and the prior leaning toward $d = y$. As a result, $d = n$ is challenged more than is $d = y$.

Given the above discussion, a judge with $s = n$ and $t = q$ would rather contradict previous decisions surely for $\theta = 0$ or more generally whenever θ is low enough. If no appeal is brought, both decisions, y and n, yield the same expected utility, since the judge believes that each of them is equally likely to be correct. If an appeal is brought, both decisions yield the same utility as well, because the higher court makes the correct decision. The tradeoff is solved, however, in favor of $d = n$ because its *probability* of appeal is higher, in which case the final decision is more likely to be correct. If such a judge with $t = q$ prefers $d = n$, the cutoff point then has to satisfy $t^e < q$.

Of course, decisions that are more likely to be challenged are also more expensive, because appeals are costly. Thus, if costs are important, such a judge opts for the cheaper decision and rules $d = y$. In this case, $t^e > q$. The cutoff point therefore increases with θ, which implies that the judge follows previous decisions more often the higher is θ. Finally, when $\theta \to \infty$, the judge cares only about costs and hence the cutoff point must be such that both decisions are challenged with the same probability.

□ **A careerist judge.** We are now ready to analyze the behavior of a careerist judge. Note that from the point of view of L and H, the exact motivation of the judge is not important. L just conjectures the strategy of the judge, whereas H simply reverses or affirms the decision given his information about the state of the world.

Recall that the careerist judge would like to impress an evaluator, who assesses the likelihood that she has accurate information. In particular, she is interested in maximizing the posterior beliefs τ of the evaluator on her expected ability. Let $\tau(d, w, \sigma)$ denote the updated belief of E about the expected type t of the judge, if E believes that J uses some strategy σ, the judge's decision is d, and E were to know the state of the world w. That is, $\tau(y, y, \sigma)$ denotes the beliefs of E when J decides y and she is correct. Similarly, $\tau(y, n, \sigma)$ denotes the updated belief of E when $d = y$ but E were to know that $w = n$. And so on.

The evaluator, however, does not observe the state of the world w and has to form beliefs about it or, in other words, about whether the judge is correct or not. Similarly to the litigants, the evaluator knows the decision d and the prior q, and has a conjecture about the strategy of the judge, σ. Thus, when no appeal is brought, the evaluator believes that the judge is correct with probability $q_d(\sigma)$. On the other hand, if an appeal is brought, the evaluator can also extract information about w from the decision of the higher court.

The evaluator will therefore attribute the reputation $\tau(d, d, \sigma)$ to the judge with the probability with which he thinks that $d = w$, i.e., that the decision and the state of the world are the same. Figure 2 helps to illustrate when this is indeed the case. The tree in the figure describes the possible "events" in the game (whether there is an appeal or not, whether the judge is perceived to be correct or not) and identifies the probabilities of each of these events, as perceived by the judge herself.

For example, the judge believes that she receives the reputation $\tau(d, d, \sigma)$ if an appeal is brought and she is found correct, which happens with probability $(1 - q_d)\Pr(w = d \mid q, s, t)$, or if no appeal is brought, but the evaluator believes that she is correct, which happens with probability q_d^2. We can therefore see how each of the judge's decisions, through the appeals process, induces a different probability distribution over the information about the state of the world that is available to the evaluator. This is why monitoring in this model is endogenous, that is, the judge can influence the evaluator's information about the state of the world and consequently about her type.

Using Figure 2 we can express the expected utility of J from a decision d as follows:

$$((1-q_d)\Pr(w = d \mid q, s, t)+q_d^2)\tau(d, d, \sigma)+((1-q_d)\Pr(w = d' \mid q, s, t)+q_d(1-q_d))\tau(d, d', \sigma). \tag{4}$$

As seen in (4), the belief of J that she would be perceived correct by E is increasing in $\Pr(w = d \mid q, s, t)$. Thus, J believes that her own information is correlated with that of E. The greater the probability she attaches to the event that $w = d$, the greater the probability she attaches to the event that E knows that $w = d$. This feature would discipline the judge to behave informatively, even if E does not know w for sure, and moreover his monitoring possibilities depend on the judge's decision. We can then establish the following lemma (where the superscript c stands for a careerist judge).

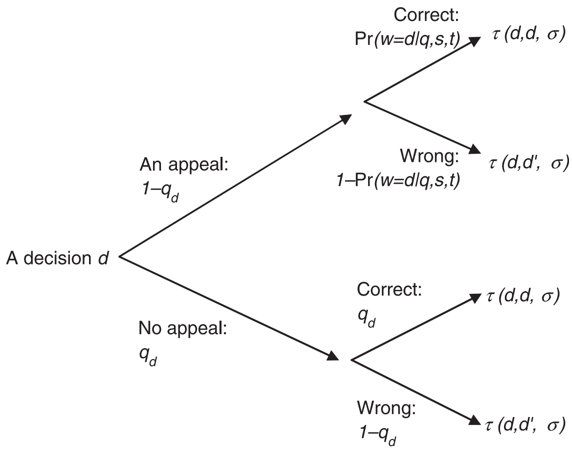
Lemma 2. In equilibrium, the careerist judge uses a cutoff-point strategy (s^c, t^c), that is, she rules y if and only if $\Pr(w = y \mid q, s, t) \geq \Pr(w = y \mid q, s^c, t^c)$.

The strategy of the careerist judge is similar to that of the efficient judge, who also uses a cutoff strategy. We therefore have to check which cutoff point the reputation incentives of the careerist judge induce her to choose. As a first step, we can impose more structure on the beliefs of the evaluator E whenever he conjectures that J uses some cutoff-point strategy $\sigma^c = (s^c, t^c)$ and if he were to know w.

Lemma 3. (i) For any action d, $\tau(d, d, \sigma^c) > \tau(d, d', \sigma^c)$.

(ii) If $s^c = n$, then $\tau(n, n, \sigma^c) > \tau(y, y, \sigma^c)$ and $\tau(n, y, \sigma^c) > \tau(y, n, \sigma^c)$.

FIGURE 2



(iii) If $s^c = y$, then $\tau(y, y, \sigma^c) > \tau(n, n, \sigma^c)$ and $\tau(y, n, \sigma^c) > \tau(n, y, \sigma^c)$.

Lemma 3 follows from Bayesian updating. The first part asserts that the reputation of J is higher if she makes the correct decision; this can arise as a signal on ability, since J is more likely to receive the correct signal when she is able, and her cutoff-point strategy is responsive to her signal.

In addition, Lemma 3 asserts that if $s^c = n$, the reputation that E attributes to those who rule n, whether they succeed or fail in making the right decision, is higher than the reputation they receive when they rule y. Intuitively, when $s^c = n$, J rules n only if $t > t^c$ (as in Figure 1). Hence, E knows that if $d = n$, it must be that $t > t^c$, whereas if $d = y$, J may admit a lower type, of $t < t^c$. The opposite happens when $s^c = y$. In this case, higher reputation is attributed to those who decide y.

The next lemma helps us to focus our analysis.

Lemma 4. In equilibrium, $s^c = n$.

By part (iii) of Lemma 3, if the evaluator believes that $s^c = y$, higher reputation is attributed to those who follow previous decisions compared with those who contradict, both when they are perceived as making the right decision and when they are perceived as making the wrong one. Moreover, types with $s = y$ have a higher probability of being perceived as making the right decision when they follow others since most of their evidence indicates that this is the right course of action. By part (i) of Lemma 3, being perceived as making the right decision also provides a higher reputation. This implies that the expected utility from ruling y is higher than from ruling n for these types. Hence no type with $s = y$ can be indifferent and the belief of the evaluator that $s^c = y$ cannot be sustained.

Given that $s^c = n$ in equilibrium, we can now find t^c. At the cutoff point t^c, the expected utility from each decision, as expressed in equation (4), has to be equal. This condition, along with correct conjectures of J's strategy by E and L and rational updating on their behalf, yield the next result; the first part of Proposition 2 characterizes the equilibrium, whereas the second part provides comparative statics results.

Proposition 2. (i) When J is careerist, there exists a unique informative equilibrium. In the equilibrium, the judge rules $d = y$ if $s = y$ or if $s = n$ and $t < t^c(q)$.

(ii) The cutoff point $t^c(q)$ increases with q, $t^c(q)_{q \to 1/2} \to 1/2$ and $t^c(q)_{q \to 1} \to \hat{t} < 1$, that is, $t^c(q)$ is bounded away from one for all q. In equilibrium, an appeal is more likely when the judge contradicts previous decisions.

Two types of signals emerge in equilibrium. The first signal is proving ability by contradicting previous decisions. This occurs because $s^c = n$ and hence, by Lemma 3, $\tau(n, n, \sigma^c) > \tau(y, y, \sigma^c)$ and $\tau(n, y, \sigma^c) > \tau(y, n, \sigma^c)$. The reason is that in equilibrium, only types with sufficient ability allow themselves to contradict previous decisions. These able judges may have private information that outweighs the informativeness of past verdicts. The second signal is proving ability by making the correct decision. A type who makes the correct decision is more likely to be able. Thus, a judge who is reversed has a lower reputation than a judge whose decision is reaffirmed. At the equilibrium cutoff point, the tradeoff between these two signals manifests itself: if this type of judge follows previous decisions, she is more likely to be correct but forgoes the possibility of using the signal of contradicting. If she contradicts, she receives high reputation for doing so but is more likely to err and be reversed.

Note that an informative equilibrium exists, even with endogenous monitoring. The least-able judges are not tempted to contradict previous decisions, although this provides high reputation, because in equilibrium such an action induces a higher probability of appeal. This is bad news, since the less-able judges may get "caught" by the higher court. Hence, the less-talented judges would rather follow previous decisions and be perceived as correct. The more-able judges, on the other hand, are encouraged to make decisions that are likely to be appealed, since this will affirm their ability. Even when they contradict precedent, they believe that their decision will not be overturned, since their own information is relatively accurate. Thus, although a judge's decision is a signal of her type and there are no exogenous costs in her utility function for ruling in favor of the plaintiff or in favor of the defendant, "costs" for making the wrong ruling are endogenously created in equilibrium.

Proposition 2 also characterizes the judge's behavior as a function of the parameter q. When q increases, the benefit from following previous decisions, all else equal, is higher. This is because the terms of the reputational tradeoff change; when one follows previous decisions, it becomes more likely that one will receive the (higher) reputation for making the correct decision. Hence, more types are inclined to follow previous decisions, that is, the cutoff point $t^c(q)$ increases.

However, a significant number of types choose to contradict previous decisions, as implied by the result that the value of $t^c(q)$ is bounded. In particular, I find that $t^c(q)$ is bounded by .625.[14] Thus, when $q \to 1$, all types in $(.625, 1)$ make the wrong decision, consciously and probably inefficiently.[15] To see why $t^c(q)$ is bounded, note that if the evaluator conjectures that $t^c(q)$ is very high, for example $t^c(q) \to 1$, then $\tau(n, \cdot, \sigma^c) > \tau(y, \cdot, \sigma^c)$. That is, the reputation from contradicting is higher than that from following regardless of the state of the world, since those who contradict previous decisions are only the most-able types, with $t \to 1$. In particular, $\tau(n, y, \sigma^c) > \tau(y, y, \sigma^c)$, i.e., even if the judge goes against her predecessors and is found wrong, her reputation is higher compared to the scenario in which she follows them and is found correct. Thus, if these are the beliefs of the evaluator, any judge would rather contradict previous decisions. This implies that such beliefs for the evaluator cannot be sustained, for any q. Consequently, there is an upper bound on the cutoff point. This feature will allow us to analyze the distortion due to career concerns, which I do next.

□ **Contrarian judges.** Both the efficient and the careerist judge behave in a relatively similar manner. That is, they both contradict previous decisions only if $s = n$ and t is high enough, in particular for $t > \{t^e(q, \theta), t^c(q)\}$ for the efficient and the careerist judge respectively.[16] I now compare the behavior of the two differently motivated judges. If, in equilibrium, $t^c(q) = t^e(q, \theta)$, then the careerist judge behaves efficiently. If $t^c(q) > t^e(q, \theta)$, it means that the careerist judge rules y more than the efficient judge, which I term *excessively following previous decisions*. If, on the other hand, the equilibrium value admits $t^c(q) < t^e(q, \theta)$, the careerist judge rules n more

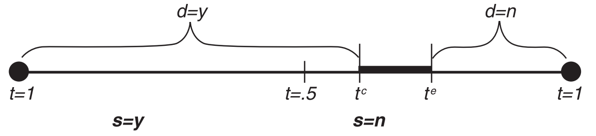
[14] This number is computed for the uniform prior distribution of the judges' types. The model is general and can be extended to other distributions. In this case, the upper bound would be different.

[15] I discuss the inefficiency of the careerist judge's decisions in Section 4.

[16] Moreover, it is easy to show that a judge who cares both for efficiency and reputation would use a cutoff point which is in between the respective cutoffs for each type of judge, the efficient and the careerist one.

FIGURE 3

A COMPARISON BETWEEN THE CAREERIST AND THE EFFICIENT JUDGE



than the efficient judge, which I term *excessively contradicting previous decisions*. The next result establishes that the judge tends to excessively contradict previous decisions.

Proposition 3. For any q, there exists $\theta(q)$ such that for all $\theta \geq \theta(q)$, the careerist judge excessively contradicts previous decisions, that is, $t^c(q) < t^e(q, \theta)$. Moreover, there exists $\bar{q}$ such that for all $q \geq \bar{q}$, $\theta(q) = 0$, i.e., the careerist judge excessively contradicts previous decisions for all values of θ.

Thus, careerist judges tend to be “creative” and to go against their predecessors, a behavior that induces an underutilization of previous decisions. The result that judges excessively contradict previous decisions is derived for high-enough values of q *or* for high-enough values of θ. Two remarks are due about these parameter values. First, in the context of our model, it is indeed consistent to assume that θ has high values. If θ is low and deliberation is not costly, the judicial system should enforce appeals or target almost all cases to the higher court. However, my analysis focuses on a judicial system that does not automatically transfer all cases to the higher courts; such a system must therefore embody high costs for a judicial review.

Second, numerical analysis shows that the judge actually contradicts previous decisions excessively for all other parameters as well. But this cannot be proven analytically because of the complex nature of efficient decision making in this model.

Figure 3 illustrates the distortion due to career concerns. Recall that the figure depicts the information of the judge, so that as we go from left to right, $\Pr(w = n \mid s, t)$ increases. The figure shows the area (t^c, t^e) in which the efficient judge decides $d = y$ whereas the careerist judge rules $d = n$.

Intuitively, the careerist judge gains excessive reputation from contradicting previous decisions, whereas the efficient judge treats each of these decisions symmetrically. Both judges face incentives to make the right decision, the efficient one because she cares about it directly and the careerist one because she cares about it indirectly, since it provides higher reputation. However, the careerist judge has an additional incentive to contradict previous decisions, since this too provides high reputation in equilibrium.

In more technical terms, the intuition for the result is as follows. For high values of q, the efficient judge follows previous decisions quite often because it is more likely to be the correct decision. On the other hand, the cutoff point of the careerist judge is bounded for any q by some $\hat{t}$, as established in Proposition 2. Thus, no matter how high q is, a significant portion of types contradicts previous decisions. For high values of θ, similarly, the efficient judge is inclined to follow previous decisions, since this is the cheaper course of action. The litigants view it as the more accurate decision and as a result challenge it less often. The judge who wishes to save on costs in this case follows others. For the careerist judge, this consideration is irrelevant, since she does not care about the costs of appeal.

Remark 1. The results reported in Propositions 2 and 3 yield the following empirical predictions. First, a judge who contradicts previous decisions and is affirmed is the most likely to be promoted. A judge who follows previous decisions and is reversed is the least likely to be promoted. Second, a careerist judge tends to contradict previous decisions more than an efficient judge. Thus, if incentives can be identified, stronger career concerns imply that when a judge contradicts previous decisions, she is more likely to be challenged by litigants and also more likely to be reversed.

In other contexts, several articles, such as Levy (2004), Trueman (1994), and Avery and Chevalier (1999), analyzed the behavior of careerist experts and showed that experts may behave inefficiently by excessively contradicting prior information. All the articles in this literature assume exogenous monitoring, i.e., that the evaluator knows what the right decision is independently of the decision itself. The contribution of the suggested judicial model to this literature is therefore the analysis of *endogenous monitoring*. It remains to be seen how endogenous and exogenous monitoring compare, a task I tackle next.

4. The design of judicial systems

■ It is now widely accepted that institutions matter (see, for example, North (1990)). That is, institutional rules affect individual incentives and, as a result, equilibrium outcomes. Since careerist judges distort their decisions, we can ask whether other, alternative judicial systems may change the careerist judge's incentives and mitigate her distortive behavior, so as to increase social welfare.

In this section I analyze the equilibrium of the model with a careerist judge, under two alternative institutions. The first one is the practice according to which judges are appointed by senior judges. This institution alters the amount of information that the evaluator may have when assessing the judge's type. The second one is the practice of binding precedent, which I analyze by changing the preferences of the higher court and hence its actions.

In both cases I maintain the structure of the legal process as outlined in Section 2; specifically, following the adjudication of the lower-court judge, the losing litigant decides whether or not to bring an appeal, and the high court rules upon appeals. This implies that for any of the alternative institutions that I analyze, the equilibrium is again unique and characterized by a cutoff point for the careerist judge.

The equilibrium cutoff point is the main determinant of social welfare; it affects the likelihood that the judge makes the correct decision, the likelihood of an appeal, and its expected costs. Thus, to evaluate the effect of different institutions on social welfare, I can compute the indirect social utility as a function of the cutoff, $t^c(q)$, and compare the *(ex ante)* social utility provided by each judicial institution by simply using the equilibrium value of the cutoff point $t^c(q)$ that each of them generates. Before doing so, I need to identify the relation between $t^c(q)$ and social welfare.

Recall that the litigants in the model do not necessarily behave efficiently, since they value costs differently compared with society in general. This implies that the behavior of the efficient judge is not necessarily optimal in terms of social welfare. It is possible to calculate social utility and observe that the efficient judge always induces a higher social utility than the careerist judge does; but because the social optimum may be different from the efficient judge's behavior, this still does not directly imply that social utility increases monotonically with $t^c(q)$ all the way up to $t^e(q, \theta)$, i.e., the more the careerist judge behaves like the efficient judge. The following lemma identifies when this is indeed the case.

Lemma 5. Whenever $t^c(q) < t^e(q, \theta)$ and θ is high enough, social welfare increases with $t^c(q)$, when $t^c(q)$ increases up to (and sometimes beyond) $t^e(q, \theta)$.

Lemma 5 establishes that for the relevant parameters specified in Proposition 3, social welfare increases with $t^c(q)$ at least up to $t^e(q, \theta)$. In fact, even for low values of θ, social welfare may increase locally with $t^c(q)$. However, Lemma 5 allows us to assess institutions that induce large changes in $t^c(q)$, for the purpose of the analysis in this article as well as for future analysis. Intuitively, for high-enough θ, the optimal cutoff point from the point of view of society is higher than $t^e(q, \theta)$. This is because the litigants bring too many appeals, and to compensate for that, the cheaper decision should be made more often. *A fortiori*, then, whenever $t^c(q) < t^e(q, \theta)$, and θ is high enough, social welfare increases if the careerist judge behaves more like the efficient judge and follows previous decisions more often.

In the rest of this section, I focus therefore on this range of parameters, that is, high-enough values of θ, and ask whether some alternatives for the judicial system described in this article can increase the tendency of the careerist judge to follow previous decisions.

□ **Who should select judges?** The procedure of judges' selection and promotion is heavily debated in many countries. In the United States, there is an ongoing debate about public election versus political appointment of judges.[17] In Britain, the Commission for Judicial Appointments, which advises the government on legal matters, has recently suggested a new procedure for judicial appointments.[18] According to the new procedure, all judges up to the level of the High Court would be selected by an independent panel of experts, most of whom would be drawn from outside the legal profession. This is in contrast to the tradition so far, in which judges are appointed by the government after consultation with existing judges and senior lawyers.

Often, different methods of judicial appointments are considered to have bearing on how independent judges' decisions are. Using the model suggested in this article, we can actually broaden the debate about different appointment procedures and assess the effect of these different procedures on career concerns and hence on the efficiency and quality of judicial decisions.

Indeed, the proposed new system for judicial appointments in Britain seems to accord with the model analyzed above, in which the evaluator "uses" endogenous monitoring. The evaluator in the model, who tries to assess the judge's ability in order to promote the able judges, can only glean information about the correct interpretation of the law from decisions of the High Court. Thus, the evaluator can represent a committee whose members have no legal experience. In other words, the costs of independently learning information about the right decision for this committee are prohibitive.

On the other hand, in the traditional selection system, lawyers and High Court judges select whom to promote among lower-court judges. But when High Court judges review a lower-court judge's file in order to decide whether to promote her or not, they may be able to determine whether she was right or wrong in each case without the need to wait for an appeal. These legal system insiders may know the correct interpretation of the law even if they do not observe the result of an appeal, and even if they do not adjudicate the case themselves. More precisely, the costs they have to incur in order to independently learn the true state should be relatively low.

This, more traditional, promotion system can be represented by a model with exogenous monitoring, i.e., a model in which the evaluator knows the state of the world independently of the judicial process, albeit at some cost. We can assume that the motivation to select the right judges is strong enough to overwhelm the cost parameter, so that such evaluators would always exert the effort or incur the cost to learn the state.

To compare between the two promotion systems I first analyze the equilibrium of the same model outlined in Section 2, but when the careerist judge is promoted by her superiors. That is, the only difference is that the evaluator learns the state w independently of the judicial process, while exerting some cost γ.[19] I then compare this equilibrium to the equilibrium in the model when the careerist judge is selected by the nonexperts committee, i.e., the evaluator knows the state w only if an appeal is brought (this equilibrium is characterized in Proposition 2).

A final note is due before presenting the result. I am comparing the two selection systems only with respect to social welfare as defined in the model: the costs and benefits from contemporary judicial decisions. Another relevant dimension is which judges they actually select, in other words, which system is better at identifying the able judges. These two dimensions relate to an intertemporal tradeoff, since the incentives of the careerist judge affect social utility from *present* decisions, whereas selection of able judges affects social utility from *future* decisions. This intertemporal tradeoff is beyond the scope of this article.

How, then, does each system affect the career-motivated judge? Intuitively, the more information the evaluator possesses, the harder it is for less-able judges to mimic the more-able judges. The next result characterizes the equilibrium when the evaluator learns w after any decision d, and shows that it is actually less efficient.

---

[17] For one of many articles on the issue, see Kopecky (1997).

[18] See the report in *The Observer*, August 31, 2003,available at http://observer.guardian.co.uk/uk_news/story/0,6903,1032749,00.html.

[19] The cost may vary with the decision. In any case, it has to be small enough so that the evaluator learns the state disregarding the decision.

Proposition 4. (i) When E knows w, the careerist judge decides y when $s = y$ or when $s = n$ and $t < t^f(q)$.[20]

(ii) The careerist judge follows previous decisions more often when E learns information from appeals than when E has full information, i.e., $t^f(q) < t^c(q)$. Social utility is therefore higher when the evaluator learns information only from appeals.

Increasing the amount of information available to the evaluator is even worse; it further distorts the decisions of the judge and, given Lemma 5, decreases social utility when $\gamma = 0$, let alone for positive values of γ. The implication of the proposition is that society may be better off if judges are selected by an independent committee with no legal expertise. In other words, even if it would have been costless for superior judges or lawyers to learn the information about the judges they contemplate promoting, social utility (from current judicial decisions) would be higher if the task of promoting judges were "outsourced" from the legal profession to a committee of nonexperts.

What is the intuition for this counterintuitive finding? Career concerns induce judges to prove their competence by excessively contradicting previous decisions. This becomes a signal, since only able judges have precise enough information to outweigh past verdicts. But when the evaluator learns from appeals, an important feature of the equilibrium is that the probability of appeal is higher after a decision that contradicts previous ones, as compared with a decision that follows previous ones. Contradicting previous decisions is therefore a "riskier" action for the less-able types, who are likely to be reviewed and found wrong. Following others becomes then a relatively "safe action" in equilibrium; less information about the judge's type is revealed. The less-able types indeed prefer to hide the truth about their ability, so their incentive to mimic able judges by contradicting previous decisions is mitigated. With exogenous monitoring, on the other hand, the evaluator knows the state of the world irrespective of the state and the judge's decision, and thus no such monitoring asymmetries occur. The judge in this case is more inclined to use the inefficient signal of contradicting previous decisions.[21]

Remark 2. The result of Proposition 4 implies the following empirical predictions. First, conditional on an appeal, a judge appointed by senior judges is more likely to be reversed when she contradicts previous decisions than is a judge selected by nonexperts. The opposite holds when the judge follows previous decisions. Second, a judge appointed by senior judges is more likely to be challenged by litigants when she contradicts previous decisions, compared with a judge elected by nonexperts.

□ **Binding precedent.** The intuition gained from the analysis above is that the judge behaves more efficiently if contradiction of previous decisions is penalized by a higher risk of appeal. This is because it reduces the distortive signalling incentives of the less-able judges. We can therefore look for institutions that increase the incentives of losing litigants to bring an appeal when the judge contradicts previous decisions.

For example, if the higher court could commit to a higher rate of reversal when the judge contradicts her predecessors, litigants could indeed be encouraged to bring an appeal more often in this case. This suggestion of different legal standards seems to be in accord with *stare decisis*, the praxis of the Common Law legal system. The concept of binding precedent indirectly implies that appeals courts are more likely to reverse a decision that contradicts precedent than one that follows precedent.

I model binding precedent in the following way; when the judge follows or contradicts previous decisions and is found wrong, the higher court reverses her decision as before. But

[20] The superscript f denotes that the evaluator has *full* information.

[21] In his seminal article about career concerns, Holmström (1999) exogenously assumes the existence of a "safe action," i.e., an action after which the state of the world is not revealed. Similarly, in Zwiebel (1995), actions are exogenously differentiated according to the information they produce to the market. Here, I derive a "safe action" and a "risky action" endogenously.

even when the judge contradicts previous decisions and is *correct*, the higher court overturns this decision with some probability. Thus, implicitly, this modelling embodies the idea that higher courts act in order to preserve the strength of previous decisions; a contrarian lower-court judge should be reversed even if she is correct.

If the higher court could commit to reverse a decision that goes against precedent even if it is correct, it would indeed increase the incentives of litigants to appeal when a judge contradicts previous decisions, as desired. The current specification of preferences of the high court judge would not enable him to commit to such a strategy, because it forces him to make a decision he perceives as the wrong one.

However, high court judges may be biased in their preferences. In other words, to induce such practice, one can nominate biased Supreme Court Justices. For example, if Supreme Court Justices are power seeking and do not tolerate departure from precedents, or have an ideology that concurs with current precedent, they will indeed find it optimal to reverse a contrarian judge.

Formally, consider again the main model outlined in Section 2, but suppose that we can optimally choose the strategy of the higher court and let $\phi \in [0, 1]$ denote the probability with which the judge is overruled when $d = n$ and $w = n$. The parameter ϕ can be interpreted as the degree of bias of the Supreme Court Justices in favor of precedent. When $d = n$ and $w = y$, the judge, as before, is reversed with probability 1. No changes are made when $d = y$, that is, the decision is affirmed if and only if $w = y$. Litigants appeal with a higher probability when $d = n$, because the probability of reversal is now $1 - q_n + q_n\phi$. The exercise is therefore to find what the optimal ϕ is, that is, the optimal bias of the high court, which maximizes social utility.

Proposition 5. When precedents bind,

(i) the careerist judge contradicts precedent more often, that is, $t^c(q)$ is lower for any $\phi > 0$ relative to the case in which $\phi = 0$.

(ii) Social utility is higher when $\phi = 0$ than when $\phi > 0$.

Surprisingly, binding precedents do not induce the judge to behave more efficiently, and even have the contrary effect. Intuitively, when previous decisions bind and the judge rules n, she may be perceived as being correct even if she is subsequently reversed: the evaluator knows that the higher court also reverses correct decisions. Hence, judges of low ability are induced to rule n and contradict previous decisions, since the loss from being reversed is not that great. Part (ii) of Proposition 5 establishes that indeed this practice decreases social utility; it is a combination of two effects. First, the judge contradicts previous decisions more often, which, by Lemma 5, decreases social utility. Second, there is a direct negative effect of this practice on social utility. This is because decisions at the high court are made inefficiently and because litigants distort the appeals process even more when $\phi > 0$, as they are encouraged to bring more appeals.

The result implies that promoting ideologically biased or power-seeking judges to the Supreme Court may reduce social welfare. Judges who are too conservative and wish to maintain precedent, or those whose ideological views are in favor of current precedent, will indirectly provide the wrong incentives for the lower-court judges and induce the latter to further distort their decisions.

5. Discussion

■ I conclude by discussing some of the assumptions of the model, to illustrate the robustness of its results as well as to highlight some possible extensions.

□ **Reputation for ideology.** There are many ways to think of reputation motives. In this article I have used ability as a desired trait for a judge. It is also possible that the judge is trying to prove to some evaluators, such as the public, politicians, or higher-court judges, that she shares their preferences regarding the interpretation of the law. In such an environment, the one who promotes the judge, for example a politician, is biased toward one of the parties to the conflict, and judges are promoted if they are considered sufficiently biased in this direction as well.

Such a model is analyzed in Morris (2001), who shows that when these concerns for proving the right ideology are too strong, an informative equilibrium may not exist. When such ideology concerns are sufficiently weak, an informative equilibrium exists and is characterized by the feature that decision makers bias their decisions to cater to their evaluator's preferences. In particular, even those who do have the "right" preferences would distort their decisions and behave inefficiently. This is because they need to accumulate reputation and hence to distinguish themselves from the types who have the "wrong" preferences.

Adding such considerations to the model presented in this article would not alter the results, if these concerns for acquiring a reputation for having the "right" ideology are weak enough. That is, career concerns (concerns for proving competence) would still induce judges to excessively contradict previous decisions and to behave more efficiently when promoted by outsiders, also when judges somewhat care to prove that they have some particular views.[22]

□ **Dynamics.** The model analyzes a one-shot game, i.e., the adjudication of one case. The main results should also hold in a dynamic context, in which the judge adjudicates a sequence of cases. When the game is prolonged, more information is revealed about the judge after each of her decisions. This only induces a different prior belief about the judge's type after each stage, and hence the results should be robust to this extension. There are other issues, however, that arise in a dynamic environment, which I now discuss.

□ **Reputation incentives for litigants.** When the game is repeated, a possible consideration for litigants is their reputation. If one of them is a repeated player, his actions in the current case may affect the perception of others about his type. For example, a defiant defendant, who appeals even when the costs of doing so are very high, may signal to future plaintiffs that it is better to settle than go to court.

In my model, if the plaintiff would have an excessive incentive to bring an appeal (compared with that of the defendant), the judge's decision would become more efficient. This is a consequence of the intuition gained in Section 4; going against precedent (i.e., against the plaintiff), would become more risky from the point of view of the judge. On the other hand, if the defendant is a litigant with additional incentives to bring appeals, then the judge would distort her decisions even more. From the point of view of the less-able judges who wish to hide their type, it would be even less attractive to follow previous decisions, due to the increased probability of this decision being challenged. Thus, such dynamic considerations may either ameliorate or exacerbate the inefficiencies identified in this model.

□ **Evolution of precedent.** We can use the model to analyze the evolution of the persuasive power of previous decisions. In particular, conditional on no appeal, a careerist judge's decision, when added to the body of previous decisions, would increase the persuasiveness of precedent, compared to an efficient judge's decision. To see the intuition for this, recall that precedent favors $d = y$, whereas the careerist judge's decisions—any of them—are stronger signals in favor of y; when the careerist judge rules y, she does it more selectively than the efficient judge. Similarly, when she rules n, because she does it more often than the efficient judge and for more signals that favor y, it indicates that the state of the world is y with a higher probability compared with a $d = n$ ruling of the efficient judge. Thus, previous decisions become more persuasive the more the judge has career concerns.[23]

---

[22] Maskin and Tirole (2004) also analyze a model in which a politician is motivated by proving that she has the right preferences. In their article, a judge is interpreted as an agent who has no career concerns.

[23] Kuran and Sunstein (1999), Daughety and Reinganum (1999), and Talley (1999) also discuss precedent in a dynamic context and in particular its interpretation as an information cascade.

Appendix

■ Proofs of Propositions 1–5 and Lemmas 2–5 follow.[24]

The following preliminary result, Lemma A1, concerns the Bayesian updating performed by the litigants (and the evaluator, when there is no appeal). In particular, when J uses a cutoff-point strategy, $\sigma^* = (s^*, t^*)$, we can write $q_d(s^*, t^*)$ in the following way:

$$q_y(n, t^*) = \frac{q\left(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv\right)}{q\left(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv\right) + (1-q)\left(\int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv\right)}$$

$$q_n(n, t^*) = \frac{(1-q)\int_{t^*}^{1} vf(v)dv}{(1-q)\int_{t^*}^{1} vf(v)dv + q\int_{t^*}^{1}(1-v)f(v)dv},$$

where $f(v)$ is the prior distribution over t. Similar expressions hold when $s^* = y$.

Lemma A1. When J uses a cutoff-point strategy (s^*, t^*), then

(i) $q_n(n, t^*)$ increases with t^* and $q_n(y, t^*)$ decreases with t^*;

(ii) $q_y(n, t^*)$ decreases with t^* and $q_y(y, t^*)$ increases with t^*;

(iii) $q_n(n, t^*) > q_n(y, t^*)$ and $q_y(n, t^*) < q_y(y, t^*)$;

(iv) $q_y(s^*, t^*) > q$ and $q_n(s^*, t^*) > 1 - q$;

(v) $q_y(s^*, t^*) > \Pr(w = y \mid q, s^*, t^*)$ and $q_n(s^*, t^*) > \Pr(w = n \mid q, s^*, t^*)$;

(vi) $\exists!(\tilde{s}, \tilde{t}(q))$ such that $q_y(\tilde{s}, \tilde{t}(q)) = q_n(\tilde{s}, \tilde{t}(q))$, where $\tilde{s} = n$ and $\tilde{t}(q) > q$.

For some cutoff point (s, t), define $\tilde{p}(d) = (1 - q_d)\Pr(w = d|q, s, t) + q_d^2$. Thus, when $s^* = n$ and for some t^* :

(vii) $\partial\tilde{p}(y)/\partial t^* < 0$, $\partial\tilde{p}(n)/\partial t^* > 0$, $\partial\tilde{p}(y)/\partial q > 0$, and $\partial\tilde{p}(n)/\partial q < 0$.

Proof of Lemma A1. The proof, which is purely technical, is provided in the web Appendix.

Proof of Proposition 1. An equilibrium is a solution to

$$\frac{\Pr(w = y \mid q, s^e, t^e)}{\Pr(w = n \mid q, s^e, t^e)} = \frac{q_y(s^e, t^e) - \theta\beta(s^e, t^e)}{q_n(s^e, t^e) + \theta\beta(s^e, t^e)}, \tag{A1}$$

where $\beta(s^e, t^e) = (1 - q_n(s^e, t^e))^2 - (1 - q_y(s^e, t^e))^2$, given true beliefs of L about J's strategy.

Step 1. An equilibrium exists.

Proof of step 1. When $s^e = y$,

$$\frac{\Pr(w = y \mid q, y, t^e)}{\Pr(w = n \mid q, y, t^e)} > \frac{q_y(y, t^e)}{q_n(y, t^e)} > \frac{q_y(y, t^e) - \theta\beta(y, t^e)}{q_n(y, t^e) + \theta\beta(y, t^e)}.$$

The first inequality holds for all $t > 1/2$ and $q > 1/2$ and follows from plugging in the expressions for $q_y(y, t^e)$ and $q_y(y, t^e)$. The second inequality follows because $q_y(y, t^e) > q_n(y, t^e)$. On the other hand, when $s^e = n$ and $t^e = \tilde{t}(q) > q$,

$$\frac{\Pr(w = y \mid q, n, \tilde{t}(q))}{\Pr(w = n \mid q, n, \tilde{t}(q))} < 1 = \frac{q_y(n, \tilde{t}(q)) - \theta\beta(n, \tilde{t}(q)))}{q_n(n, \tilde{t}(q)) + \theta\beta(n, \tilde{t}(q)))}.$$

Hence, there exists $s^e = n$ and $t^e \in (.5, \tilde{t}(q))$ that supports an equilibrium.

Step 2. The equilibrium is unique.

Proof of step 2. The proof is in the web Appendix.

Step 3. $t^e(q, \theta)$ increases with θ, $t^e(q, 0) < q$, and $t^e(q, \theta) \to \tilde{t}(q)$ when $\theta \to \infty$.

[24] Some technical parts of proofs of the results are provided in a web Appendix, available at http://personal.lse.ac.uk/levyg1/appendixb.pdf.

Proof of step 3. Since the equilibrium is unique, and for all θ there exists an equilibrium with $t^e < \tilde{t}(q)$, for $\tilde{t}(q)$ that satisfies $q_y(s, \tilde{t}(q)) = q_n(s, \tilde{t}(q))$, then in equilibrium, $q_y > q_n$, which implies that the probability of appeal is higher when $d = n$. Moreover, $\Pr(w = y \mid q, n, t^e)/\Pr(w = n \mid q, n, t^e)$ is constant for all θ, whereas $(q_y(n, t^e) - \theta\beta(n, t^e))/(q_n(n, t^e) + \theta\beta(n, t^e))$ decreases with θ whenever $\beta > 0$, i.e., for all $t^e(q, \theta) < \tilde{t}(q)$, which along with uniqueness implies that t^e increases with θ. When $\theta = 0$, then the left-hand side of (A1) equals one when $t = q$, whereas the right-hand side is greater than one, by part (vi) of Lemma A1. Hence, $t^e(q, 0) < q$. Finally, when $\theta \to \infty$, only costs matter. The judge can be indifferent between the two decisions only if the probability of appeal is equal for both, i.e., if $q_y(n, t^e) = q_n(n, t^e)$, which implies that $t^e(q, \theta) \to \tilde{t}(q)$.

Step 4. $t^e(q, \theta)$ increases with q, $t^e(q, \theta) \to_{q \to 1/2} 1/2$ and $t^e(q, \theta) \to_{q \to 1} 1$.

Proof of Step 4. In the web Appendix I show that $t^e(q, \theta)$ increases with q. That $t^e(q, \theta)_{q \to 1/2} \to 1/2$ and $t^e(q, \theta)_{q \to 1} \to 1$ simply follows from (A1).

This completes the proof of Proposition 1. *Q.E.D.*

Proof of Lemma 2. In an informative equilibrium, some types of J decide n, whereas some types decide y. This implies that there must be at least one type (s, t) who is indifferent between deciding n and y. That is, the following condition must hold for some (s, t):

$$\tilde{p}(y)\tau(y, y, \sigma) + (1 - \tilde{p}(y))\tau(y, n, \sigma) = \tilde{p}(n)\tau(n, n, \sigma) + (1 - \tilde{p}(n))\tau(n, y, \sigma), \tag{A2}$$

where

$$\tilde{p}(d) = (1 - q_d)\Pr(w = d \mid q, s, t) + q_d^2. \tag{A3}$$

Rearranging (A2), and plugging in the expressions for $\tilde{p}(n)$ and $\tilde{p}(y)$ from (A3), we get

$$\frac{\Pr(w = y \mid s, t, q)}{\Pr(w = n \mid s, t, q)} = \frac{\tau(n, n, \sigma)(q_n^2 + 1 - q_n) + \tau(n, y, \sigma)(q_n - q_n^2) - \tau(y, y, \sigma)q_y^2 - \tau(y, n, \sigma)(1 - q_y^2)}{\tau(y, y, \sigma)(q_y^2 + 1 - q_y) + \tau(y, n, \sigma)(q_y - q_y^2) - \tau(n, n, \sigma)q_n^2 - \tau(n, y, \sigma)(1 - q_n^2)}. \tag{A4}$$

The right-hand side of (A4) is fixed for all (s, t). As in the proof of Lemma 1, the left-hand side of (A4) changes with (s, t), and any different (s, t) yields a different value of $\Pr(w = y \mid q, s, t)/\Pr(w = n \mid q, s, t)$. This implies that there is (at most) a unique (s^c, t^c) that satisfies equation (A4). *Q.E.D.*

Proof of Lemma 3. (i) $\tau(d, w, \sigma^c)$ is an expectation over t, using an updated density function given the observations of d and w, and the knowledge of the cutoff-point strategy σ, i.e., $\tau(d, w, \sigma^c) = \int_{.5}^1 t f(t \mid d, w, \sigma^c)dt$. To show that

$$\int_{.5}^1 t f(t \mid d, d, \sigma^c)dt > \int_{.5}^1 t f(t \mid d, d', \sigma^c)dt,$$

we can use the MLRP property, i.e., it is enough to show that

$$\frac{f(t \mid d, d, \sigma^c)}{f(t' \mid d, d, \sigma^c)} \geq \frac{f(t \mid d, d', \sigma^c)}{f(t' \mid d, d', \sigma^c)}$$

for $t \geq t'$ with a strict inequality for at least one pair of values t and t'. It is easy to see that the MLRP is satisfied; whenever J uses a cutoff strategy with $s^c = n$, then

$$f(t \mid y, y, \sigma^c) = \begin{cases} \dfrac{2t}{\int_{.5}^{t^c} 2dt + \int_{t^c}^1 2tdt} & \text{if } t > t^c \\ \dfrac{2}{\int_{.5}^{t^c} 2dt + \int_{t^c}^1 2tdt} & \text{otherwise,} \end{cases} \qquad f(t \mid n, n, \sigma^c) = \begin{cases} \dfrac{2t}{\int_{t^c}^1 2tdt} & \text{if } t > t^c \\ 0 & \text{otherwise} \end{cases}$$

and

$$f(t \mid y, n, \sigma^c) = \begin{cases} \dfrac{2(1-t)}{\int_{.5}^{t^c} 2dt + \int_{t^c}^1 2(1-t)dt} & \text{if } t > t^c \\ \dfrac{2}{\int_{.5}^{t^c} 2dt + \int_{t^c}^1 2(1-t)dt} & \text{otherwise,} \end{cases} \qquad f(t \mid n, y, \sigma^c) = \begin{cases} \dfrac{2(1-t)}{\int_{t^c}^1 2(1-t)dt} & \text{if } t > t^c \\ 0 & \text{otherwise} \end{cases}$$

implying that

$$\frac{f(t \mid y, y, \sigma^c)}{f(t' \mid y, y, \sigma^c)} \geq \frac{f(t \mid y, n, \sigma^c)}{f(t' \mid y, n, \sigma^c)} \quad \text{and} \quad \frac{f(t \mid n, n, \sigma^c)}{f(t' \mid n, n, \sigma^c)} \geq \frac{f(t \mid n, y, \sigma^c)}{f(t' \mid n, y, \sigma^c)} \qquad \text{for } t \geq t'.$$

The analogous analysis holds when $s^c = y$.

(ii) Similarly to part (i), I can show that when $s^c = n$,

$$\int_{.5}^{1} tf(t \mid n, n, \sigma^c)dt > \int_{.5}^{1} tf(t \mid y, y, \sigma^c)dt$$

and

$$\int_{.5}^{1} tf(t \mid n, y, \sigma^c)dt > \int_{.5}^{1} tf(t \mid y, n, \sigma^c)dt,$$

by using the MLRP and in particular by showing that $f(t \mid n, y, \sigma^c)/f(t \mid y, n, \sigma^c)$ and $f(t \mid n, n, \sigma^c)/f(t \mid y, y, \sigma^c)$ increase with t. This follows directly from the specification of the updated density function in part (i).

(iii) The results for $s^c = y$ follow from symmetry and part (ii). *Q.E.D.*

Proof of Lemma 4. Given (A3), the expected utility from an action d is

$$\tilde{p}(d)\tau(d, d, \sigma^c) + (1 - \tilde{p}(d))\tau(d, d', \sigma^c).$$

When $s^c = y$, then $\tau(y, n, \sigma^c) > \tau(n, y, \sigma^c)$ and $\tau(y, y, \sigma^c) > \tau(n, n, \sigma^c)$, by Lemma 3. But when $s^c = y$, for any type $s = y$, also $\tilde{p}(y) > \tilde{p}(n)$:

$$\begin{aligned}\tilde{p}(y) &= q_y^2 + (1 - q_y)\Pr(w = y \mid q, y, t) > q_y q_n + (1 - q_y)\Pr(w = n \mid q, y, t)\\ &> q_n^2 + (1 - q_n)\Pr(w = n \mid q, y, t) = \tilde{p}(n).\end{aligned}$$

The first inequality follows because $q_y(y, t) > q_n(y, t)$ and $\Pr(w = y \mid q, y, t) > \Pr(w = n \mid q, y, t)$. The second inequality follows because $q_n > \Pr(w = n \mid q, y, t)$. This implies that the expected utility from ruling y is greater than the expected utility from ruling n for all $s = y$ because $\tau(y, y, \sigma^c) > \tau(y, n, \sigma^c)$ and $\tau(n, n, \sigma^c) > \tau(n, y, \sigma^c)$, and hence no type with $s = y$ can be indifferent. *Q.E.D.*

Proof of Proposition 2.

Step 1. An informative equilibrium exists.

Proof of step 1. If $s^c = n$ and $t^c = \tilde{t}(q)$, then $\tilde{p}(y) < \tilde{p}(n)$, while $\tau(n, n, \sigma^c) > \tau(y, y, \sigma^c)$ and $\tau(n, y, \sigma^c) > \tau(y, n, \sigma^c)$. This implies that the utility from ruling n is higher than the utility from ruling y. Along with Lemma 4, this assures existence.

Step 2. $t^c(q) < \hat{t}$ for all q.

Proof of step 2. The proof is in the web Appendix.

Step 3. Uniqueness.

Proof of step 3. Lemma A1 (vii) shows that when $s^c = n$, $\tilde{p}(y)$ decreases with t, whereas $\tilde{p}(n)$ increases with t and $\tilde{p}(y)$ increases with q, and $\tilde{p}(n)$ decreases with q. The proof of step 2 of this proposition (in the web Appendix) shows that $\tau(n, y, \sigma^c)$ is increasing in t, whereas similar analysis holds for $\tau(n, n, \sigma^c)$. (This is easy to see, since these are expectations over t for values of $t > t^c$. Thus, if t^c increases, the expectations increase also.) The proof of step 2 also shows that $\tau(y, y, \sigma^c)$ is decreasing in the range $[.5, \hat{t}]$, and analogous analysis holds for $\tau(y, n, \sigma^c)$. Thus uniqueness is assured, since the expected utility from ruling n is increasing for all t, and the expected utility from ruling y is decreasing for all $t < \hat{t}$.

Step 4. $t^c(q)$ increases with q, $t^c(q)_{q\to\frac{1}{2}} \to \frac{1}{2}$ and $t^c(q)_{q\to 1} \to \hat{t}$.

Proof of step 4. $\tilde{p}(y)$ increases with q, whereas $\tilde{p}(n)$ decreases with q and hence the utility from y increases for all t relative to the utility from ruling n, which implies that in equilibrium the judge has to rule y more often, i.e., $t^c(q)$ increases. The boundary results are trivial, using the equilibrium condition specified in Lemma 2.

This completes the proof of Proposition 2. *Q.E.D.*

Proof of Proposition 3. When the judge is efficient, she uses a cutoff point $t^e(q, \theta) \in [.5, \tilde{t}(q)]$. On the other hand, the careerist judge uses $t^c(q) < \min\{\tilde{t}(q), \hat{t}\}$. Since $t^e(q, \theta)$ is a continuous function that increases with q and θ, there exists $\bar{q}$ for which $t^e(\bar{q}, 0) = \hat{t}$. Hence, for all $q > \bar{q}$, $t^c(q) < t^e(q, \theta)$. For other values of q, since $t^e(q, \theta)$ increases with θ and converges to $\tilde{t}(q)$ when $\theta \to \infty$, and since $t^c(q) < \tilde{t}(q)$ and does not depend on θ, there exists $\theta(q)$ such that for all $\theta \geq \theta(q)$, the result holds. *Q.E.D.*

Proof of Lemma 5. We first have to define the expression for social utility, denoted by $U(t)$ (for brevity the index c from t^c is omitted):

$$U(t) = q\left(\int_{.5}^{1} 2v dv + \int_{.5}^{t} 2(1-v)dv + \int_{t}^{1} 2(1-v)dv(1-q_n(n,t))\right)$$
$$+ (1-q)\left(\int_{t}^{1} 2v dv + (\int_{.5}^{1} 2(1-v)dv + \int_{.5}^{t} 2v dv)(1-q_y(n,t))\right)$$
$$- \theta(1-q_y)^2\left(q\left(\int_{.5}^{1} 2v dv + \int_{.5}^{t} 2(1-v)dv\right) + (1-q)\left(\int_{.5}^{1} 2(1-v)dv + \int_{.5}^{t} 2v dv\right)\right)$$
$$- \theta(1-q_n)^2\left(q\int_{t}^{1} 2(1-v)dv + (1-q)\left(\int_{t}^{1} 2v dv\right)\right).$$

The first two lines express the social gain from making the right decision. This happens when the judge makes the correct decision, or when she does not but an appeal is brought. The remaining expressions measure the social loss from the costs of appeal. These are paid when an appeal is brought.

I will now show that for high-enough values of θ, in particular for $\theta > 1/2$, $\partial U(t)/\partial t > 0$:

$$\frac{\partial U(t)}{\partial t} = 2q(1-t)q_n - 2t(1-q)q_y + \theta(2t(1-q) + 2(1-t)q)((1-q_n)^2 - (1-q_y)^2)$$
$$+ 2(2\theta - 1)[(1-q)(1-q_n)^2 - q(1-q_y)^2].$$

The first three elements capture the change in the utility of the efficient judge when she uses a higher value of t. Thus, for any $t < t^e$ (or for any $\theta \geq \theta(q)$ defined in Proposition 3), the first three elements are positive. Also, for any $t < \tilde{t}(q)$ and $\theta > 1/2$, the last element is positive. Hence, for $\theta \geq \max\{1/2, \theta(q)\}$, $\partial U(t)/\partial t > 0$. *Q.E.D.*

Proof of Proposition 4. (i) For the proof of the first part, see Proposition 1 in Levy (2004).

(ii) The value of $t^f(q)$ solves

$$\Pr(w = y \mid q, n, t^f(q))\Gamma_y = \Pr(w = n \mid q, n, t^f(q))\Gamma_n + \Gamma, \tag{A5}$$

where $\Gamma_y = \tau(y, y, \sigma^f) - \tau(y, n, \sigma^f)$, $\Gamma_n = \tau(n, n, \sigma^f) - \tau(n, y, \sigma^f)$, and $\Gamma = \tau(n, y, \sigma^f) - \tau(y, n, \sigma^f)$. In the web Appendix, I show that at $t^f(q)$,

$$\tilde{p}(y)\Gamma_y > \tilde{p}(n)\Gamma_n + \Gamma \tag{A6}$$

for $\tilde{p}(d) = (1 - q_d)\Pr(w = d \mid q, n, t^f(q)) + q_d^2$, which implies that at $t^f(q)$, the utility from y is higher than the utility from n if appeals are allowed, meaning that the equilibrium solution $t^c(q)$ must admit $t^c(q) > t^f(q)$.

By Lemma 5, this implies that social welfare is lower under exogenous monitoring when $\gamma = 0$ and, *a fortiori*, for any $\gamma > 0$. *Q.E.D.*

Proof of Proposition 5. (i) I will calculate $\tilde{p}(n)$ for $\phi > 0$ and show that it is higher than the value of $\tilde{p}(n)$ when $\phi = 0$, whereas $\tilde{p}(y)$ does not change with ϕ. This means that the utility from n is higher for any t, and thus $t^c(q)$ must decrease.

The expression for $\tilde{p}(n) \mid_{\phi>0}$ is composed of two elements. First, there is the probability with which the judge is perceived correct, in her eyes, if there is an appeal. Appeal occurs with probability $1 - q_n + q_n\phi$. If the state is y, or if it is n, then with probability ϕ the judge is reversed. In this case, the evaluator believes that the state is actually n with probability $q_n\phi/(1 - q_n + q_n\phi)$, which is the updated probability given the strategy of the higher court. With the remaining probability, the judge is affirmed and the evaluator believes then that she is correct with probability 1. The second element is the beliefs when no appeal takes place, which are q_n. The probability of this event is $q_n(1 - \phi)$.

I now show that $\tilde{p}(n) \mid_{\phi>0} > \tilde{p}(n) \mid_{\phi=0}$:

$$\tilde{p}(n) \mid_{\phi>0} = (1 - q_n + q_n\phi)(\Pr(w = y \mid q, s, t) + \Pr(w = n \mid q, s, t)\phi)\frac{q_n\phi}{1 - q_n + q_n\phi}$$
$$+ (1 - q_n + q_n\phi)\Pr(w = n \mid q, s, t)(1 - \phi) + q_n(1 - \phi)q_n$$
$$> \Pr(w = n \mid q, s, t)(1 - q_n) + q_n^2 = \tilde{p}(n) \mid_{\phi=0}$$
$$\iff q_n - \Pr(w = n \mid q, s, t) > 0,$$

which holds by Lemma A1, since $q_n(s, t) > \Pr(w = n \mid q, s, t)$.

(ii) Social welfare in this case is a function of both the cutoff point of the judge $t^c(q, \phi)$ and of ϕ itself. Therefore,

$$\frac{dU(t^c, \phi)}{d\phi} = \frac{\partial U(t^c, \phi)}{\partial \phi} + \frac{\partial U(t^c, \phi)}{\partial t^c}\frac{\partial t^c}{\partial \phi}.$$

From part (i) and Lemma 5, we know that $(\partial U(t^c, \phi)/\partial t^c)(\partial t^c/\partial \phi) < 0$. It is left to check that $\partial U(t^c, \phi)/\partial \phi < 0$. But this is trivially the case for high-enough θ because the litigants, inefficiently, bring more appeals the higher is ϕ, and moreover, the high court does not make the right decision when $\phi > 0$. *Q.E.D.*

References

AVERY, C.N. AND CHEVALIER, J.A. "Herding Over the Career." *Economics Letters*, Vol. 63 (1999), pp. 327–333.

AYRES, I. AND VARS, F.E. "Determinants of Citations to Articles in Elite Law Reviews. " *Journal of Legal Studies*, Vol. 29 (2000), pp. 427–450.

BHATTACHARYA, M. AND SMYTH, R. "The Determinants of Judicial Prestige and Influence: Some Empirical Evidence from the High Court of Australia." *Journal of Legal Studies*, Vol. 30 (2001), pp. 223–252.

DAUGHETY, A.F. AND REINGANUM, J.F. "Stampede to Judgment: Persuasive Influence and Herding Behavior by Courts." *American Law and Economics Review*, Vol. 1 (1999), pp. 158–189.

——— AND ———. "Appealing Judgments." *RAND Journal of Economics*, Vol. 31 (2000), pp. 505–525.

IOSSA, E. AND PALUMBO, G. "Partisans and Discretion." Mimeo, Brunel University, 2004.

HOLMSTRÖM, B. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies*, Vol. 66 (1999), pp. 169–182.

KLEIN, D. AND MORRISROE, D. "The Prestige and Influence of Individual Judges on the U.S. Courts of Appeals." *Journal of Legal Studies*, Vol. 28 (1999), pp. 371–391.

KOPECKY, F.J. "Should Judges be Elected or Appointed: The Lingering Debate on Merit Selection." *Illinois Issues*, Vol. 3 (1997), pp. 14–17. Available at www.lib.niu.edu/ipo/ii771214.html.

KORNHAUSER, L.A. "Modeling Collegial Courts I: Path-Dependence." *International Review of Law and Economics*, Vol. 12 (1992), pp. 169–185.

KURAN, T. AND SUNSTEIN, C.R. "Availability Cascades and Risk Regulation." *Stanford Law Review*, Vol. 51 (1999), pp. 683–768.

LANDES W.M. AND POSNER, R.A. "The Independent Judiciary in an Interest-Group Perspective." *Journal of Law and Economics*, Vol. 18 (1975), pp. 875–901.

——— AND ———. "Legal Precedent: A Theoretical and Empirical Analysis." *Journal of Law and Economics*, Vol. 19 (1976). pp. 249–307.

LEVY, G. "Anti-Herding and Strategic Consultation." *European Economic Review*, Vol. 48 (2004), pp. 503–525.

MALTESE, J.A. "Confirmation Gridlock: The Federal Judicial Appointments Process Under Bill Clinton and George W. Bush." *Journal of Appellate Practice and Process*, Vol. 5 (2003), available at www.ualr.edu/∼japp/maltese.html.

MASKIN, E. AND TIROLE, J. "The Politician and the Judge: Accountability in Government." *American Economic Review*, Vol. 94 (2004), pp. 1034–1054.

MICELI, T.J. AND COŞGEL, M.M. "Reputation and Judicial Decision-Making." *Journal of Economic Behavior and Organization*, Vol. 23 (1994), pp. 31–51.

MORRIS, S. "Political Correctness." *Journal of Political Economy*, Vol. 109 (2001), pp. 231–265.

NORTH, D.C. *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press, 1990.

POSNER, R.A. "An Economic Analysis of the Use of Citations in the Law." *American Law and Economics Review*, Vol. 2 (2000), pp. 381–406.

RASMUSEN, E. "Judicial Legitimacy as a Repeated Game." *Journal of Law, Economics and Organization*, Vol. 10 (1994), pp. 63–83.

SALZBERGER, E. AND FENN, P. "Judicial Independence: Some Evidence from the English Court of Appeal." *Journal of Law and Economics*, Vol. 42 (1999), pp. 831–847.

SHAPIRO, F. "The Most-Cited Legal Scholars." *Journal of Legal Studies*, Vol. 29 (2000), pp. 409-426.

SCHARFSTEIN, D. AND STEIN, J. "Herd Behavior and Investment." *American Economic Review*, Vol. 80 (1990), pp. 465–479.

SHAVELL, S. "The Appeals Process as a Means of Error Correction." *Journal of Legal Studies*, Vol. 24 (1995), pp. 379–426.

———. "The Appeals Process and Adjudicator Incentives." *Journal of Legal Studies*, forthcoming.

SPITZER, M. AND TALLEY, E. "Judicial Auditing." *Journal of Legal Studies*, Vol. 29 (2000), pp. 649–683.

TALLEY, E. "Precedential Cascades: An Appraisal." *Southern California Law Review*, Vol. 73 (1999), pp. 87–137.

TRUEMAN, B. "Analyst Forecasts and Herding Behavior." *Review of Financial Studies*, Vol. 7 (1994), pp. 97–124.

WHITMAN, D.G. "Evolution of the Common Law and the Emergence of Compromise." *Journal of Legal Studies*, Vol. 29 (2000), pp. 753–781.

ZWIEBEL, J. "Corporate Conservatism and Relative Compensation." *Journal of Political Economy*, Vol. 103 (1995), pp. 1–25.