Careerist Judges*

Gilat Levy†

Abstract

In this paper I analyze how judges formulate their decisions using information they uncover during deliberations as well as relevant information they gather from previous decisions. I assume that judges have reputational concerns and try to signal to an evaluator that they have high aptitude for interpreting the law correctly. If an appeal is made, the appellate court's decision reveals whether the judge interpreted properly the law 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 inefficiently contradict previous decisions. I also show that the praxis of binding precedent may actually exacerbate the distortionary behaviour of careerist judges.

*I thank Oriana Bandiera, Michael Birenhak, Patrick Bolton, Gene Grossman, Alessandro Lizzeri, Wolfgang Pesendorfer, Ronny Razin and seminar participants in Tel Aviv University for helpful remarks.

†LSE and Tel Aviv University. Current address: Department of Economics, LSE, Houghton St., London WC2A 2AE. Email: g.levy1@lse.ac.uk.

1 Introduction

Judging from the numerous amount of recent papers published in leading Law and Economics Journals, who count 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 not new; but ways to measure features such as prestige or influence have become more sophisticated and as a result created a renewed interest in judicial reputation. As Posner (2000) writes,

> *"Citation analysis is growing mainly because it enables rigorous quantitative analysis of elusive but important social phenomenon such as reputation, influence, prestige, celebrity..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"*[2]

Judges may care about how others perceive and rank them for two reasons. First, this can influence their career. Although judges who have life tenure positions need not be in fear of losing their job, promotion to a better position in the judicial system may depend on whether others consider them as 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. These higher-echelon positions can increase both the judge's pay and her possibilities to influence other judges. Thus, trial judges may desire to become appellate judges, and that judges of intermediate appellate courts may aspire to become judges of courts of last resort.[3]

Moreover, even Supreme Court Justices who have already clinched the most desirable spot, may be motivated by reputation; these Justices may cater to groups such as the press, elite law professors, editors of law school casebooks and even law students as well as

[1] To name a few examples: The Prestige and Influence of Individual Judges on the US Courts of Appeals, by D. Klein and D. Morrisroe, *JLS 1999*. Determinants of Citations to Articles in Elite Law Reviews, by I. Ayres and F. Vars, *JLS 2000.* The Most Cited Legal Scholars, by F. Shapiro, *JLS 2000*. An Economic Analysis of the Use of Citations in the Law, by R. Posner, *ALER 2000.* The Determinants of Judicial Prestige and Influence: Some Empirical Evidence from the High Court of Australia, by M. Bhattaharya and R. Smyth, *JLS 2001.* Citation, Age, Fame, and the Web, by R. Posner and W. Landes, *JLS 2000*.

[2] Posner, R. (2000), An Economic Analysis of the Use of Citations in the Law, *American Law and Economic Review*, 381-406.

[3] Empirical research finds that the perceived quality of a judge plays a notable part in their promotion. For example, Salzberger and Fenn (1999) find that the promotion probability from the court of appeal to the house of Lords in England is significantly determined by a lower reversal rate of the judge's decision in the house of Lords. They interpret this finding by claiming that the house of Lords believes that a lower reversal rate indicates a better judge who deserves promotion.

historians.[4] Thus, a second reason to care about what others think about them, may be a humane concern for prestige and influence. In that respect, the judicial circle is similar to the scholarly academic world.

In this paper, I formalize the effect of reputation seeking behavior on judicial decision making, be it for advancing a career or simply for the need of peer recognition. In particular, I assume that the judge is interested in creating a reputation for high judicial ability, which is the ability to interpret the law correctly. To my knowledge, this is the first paper to formally analyze judicial decision making by judges who are motivated by career concerns. Traditionally, political and legal scholars assumed either that judges try to take 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). Alternatively, Gelly and Spiller (1990) assume that judges have reversal aversion. In a similar vain, Daughety and Reinganum (1999, 2000) assume that judges take the decision they believe the Supreme Court would have taken. In my model, these motivations as well as aversions arise endogenously from more fundamental preferences.

I focus the analysis on two features of the judicial system that may affect judicial decision making. The first feature is the availability of previous decisions in similar cases, i.e., non-binding precedent, that can assist the judge in her current decision. We can then ask how much weight a judge puts on previous decisions, compared to her views in the current case, when she is motivated by reputation. This follows the analysis of Daughety and Reinganum (1999) who investigate how judges, who are interested in taking the right decision, may learn some information from decisions of other courts, termed by 'persuasive influence'.[5]

The second feature of the judicial system is an endogenous appellate review. The judge is aware that she may be subject to review by an appeals court if the losing litigant believes that her decision is likely to be reversed. I therefore ask how careerist judges take decisions in light of such a possible appellate review. The court system is often modelled as an hierarchy, as for example in Spitzer and Talley (1998), Daughety and Reinganum (1999), and Shavell (1995).[6] In my model, an important feature is that the lower-court

[4] See Schauer (1997).

[5] Miceli and Cosgel (1994) and Rasmusen (1994) also analyze the use of previous decisions by judges in a repeated game. Rasmusen (1994) assumes that a judge is interested to maximize the influence of her decisions. Thus, when a judge follows others', she strengthens the institution of precedent, and thereby encourages others to follow her own decisions.

[6] Spitzer and Talley (1998) analyze an incomplete information model to examine the process of trial and appeal, in which the appellate court uses the auditing decision to trade-off ideological biases versus accuracy of prior decisions. In my model, there are no auditing decisions, and the courts do not differ in ideology but in ability. Daughety and Reinganum (2000) analyze a hierarchy in which lower courts concentrate on fact finding and appeals courts interpret the law. Each appeals court may glean information about what the Supreme Court would view as the correct decision from the decision of the litigants to

judges take into considerations how their decisions will affect the probability that the case will advance to a higher court.

I incorporate these two features, of appellate review and the availability of previous decisions, in a Bayesian signaling model. In the model, the judge receives some private information regarding the application of the law in a particular case. The accuracy of this information depends on the judge's ability. In particular, the more able is the judge, the more 'accurate' is her interpretation of the law.[7] Previous decisions, or (non-binding) precedent, provide additional information to the judge. The judge's decision is therefore based both on her private trial information and on the information provided by past decisions. In the next stage, the losing party appeals the verdict if the probability of reversal is greater than the cost of appeal. If an appeal is brought, an appeals court delivers a reversal or affirmation decision. I assume that the social goal is to attain the correct decision (through efficient aggregation of information) at the minimum cost.

A careerist judge, however, is not motivated by social efficiency. She cares about how an 'evaluator' perceives her ability. The evaluator can represent the public, who elects judges in many US states, politicians, who nominate judges in many countries, or the legal community. I assume that the evaluator observes the judicial process described above and subsequently forms an opinion about the ability of the judge.

I show that in equilibrium, the careerist judge tends to contradict previous decisions more than an efficient judge.[8] Contradicting previous decisions becomes a signal of the judge's ability, since able judges have accurate information of their own and do not need to rely on previous decisions. Since this signal increases reputation, other types of judges, and in particular less able types, tend to use it excessively and inefficiently. However, another equilibrium feature is that when the judge is reversed, her reputation is lower than when she is affirmed since taking the right decision is also a signal about ability. Thus, the least able types realize that if they contradict previous information they may get 'caught' by the appeals court and as a result will be viewed by the evaluator as less able. Therefore, they cannot fully mimic the behavior of the more able judges. This allows for an informative equilibrium and enhances the signal of contradicting previous decisions.

---

bring an appeal. But, as opposed to my model, an appeal court does not consider whether its decision will affect the probability of a further appeal. Finally, Shavell (1995) uses a mechanism design approach and shows that by making an appeals court more accurate than a trial court, and imposing a fee or subsidy on appeals, only cases of error will be appealed.

[7] An accurate interpretation is the decision the Supreme Court would acknowledge as correct.

[8] An efficient judge maximizes the social utility, that is, attaining the right decision at the minimum costs. Such a judge is forward looking and, besides attempting to take the right decision, also weighs the probability that each of his decisions would be corrected by the higher court versus the costs of an appeal.

The predictions of the model seem to accord with some common observations. First, it seems that judges do try to be original and innovative, and to create 'precedent' instead of following an existing one. Second, there is a common wisdom that judges do not like to be reversed. This "reversal aversion" is not necessarily a characteristic of judges' preferences. It arises endogenously in equilibrium, because reversal signals that the judge's decision was mistaken and reduces her reputation. Affirmation, on the other hand, means that the higher court acknowledges the judge's superior information and analytical ability.

The results of the model differ from those of Daughety and Reinganum (2000) who predict that judges may engage in inefficient herding, that is, they excessively follow previous decisions.[9] The reason is that in their analysis judges are interested only in taking the right decision (or, what the Supreme Court perceives as the right decision) whereas in my model judges are careerist. Thus, in my model, although judges do have some endogenous incentive to take the right decision, it may be in conflict with an endogenous incentive to contradict previous decisions.

In other contexts, several papers analyzed the behavior of careerist experts and showed that experts may behave inefficiently by excessively contradicting prior information (see Trueman (1994), Ottaviani and Sørensen (2002) and Levy (2000)).[10] The contribution of the suggested judicial model to this literature is the analysis of endogenous monitoring by the evaluator. In the papers mentioned above, the evaluator knows the state of the world - the correct decision - and hence can monitor the actions of the experts. In my model, the evaluator knows the state of the world only if an appeal is brought, which depends on the judge's decision. Hence, monitoring is *endogenous*, meaning, the judge can control the accuracy of the information that the evaluator would receive and hence affect the inference about her type.

In addition to the main results, I investigate issues of optimal design of the judicial system. Since different institutions provide different incentives, I try to assess the institutional features that could increase efficiency when judges are careerist. In particular, I find that the judge behaves more efficiently when monitoring is indeed endogenous compared to an exogenous monitoring system. That is, the careerist judge behaves more efficiently when the evaluator can only learn information about the correct interpretation of the Law from appeals to higher courts and does not know the correct decision independently. This may imply that the ones who should nominate judges should know *less*. We can therefore conclude that in some cases, judges should be nominated by members of the public, as in the US, and not by fellow superior judges such as Supreme Court Justices, which is more

[9] Their analysis follows the analysis of herding and informational cascades in financial markets as explored in Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1993).

[10] Efficiency, in the papers mentioned above, usually means taking the right decision. In the judicial decision making context however, efficiency is more involved since the judge is not the final decision maker and therefore has to weigh the costs and benefits of an appeal.

similar to the UK procedure.

The rest of the paper is organized as follows. In the next section I lay out the model. 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 nomination systems on the equilibrium outcomes and explore the effect of binding precedent on the distortive behavior of reputation seeking judges. Section 5 discusses some assumptions and future research. All proofs that are not in the text are relegated to an appendix.

2 The Model

The model describes a two-tier hierarchy of a judicial process, formed of a lower-court judge and a higher-court judge.[11] Previous decisions, which are common knowledge, may provide information that can assist the courts in adjudicating current cases. Given the lower-court's decision, the losing litigant can advance his case to the higher court by bringing an appeal. The adjudication process ends if no appeal is brought, or, after the higher court's ruling.[12] An evaluator observes the judicial process and forms beliefs on the ability of the lower-court judge. I analyze a one-shot game, i.e., the adjudication of one legal case.

Information structure and actions:

A judge, J, must make a dichotomous decision d, i.e., whether to accept the plaintiff's argument ($d = y$) or to reject it ($d = n$).[13] In order to determine which decision is the right one, she must decide what is the underlying state of the world w, which could be either y or n, with the interpretation that the correct decision in state y is y and in state n is n.

The judge must interpret the intentions of the lawmakers. While adjudicating the case, she receives a private signal $s \in \{y, n\}$ about the correct interpretation. The accuracy of the signal depends on the ability of the judge. Let $\Pr(s = w|w) = t$, where Pr stands for *probability* and $t \in [.5, 1]$. For example, if the judge's ability is $t = .5$, her signal is not

[11] Higher-echelon courts are often composed of more than one judge. I depart from collective decision making issues. For analysis of collective reputation problems, see Tirole (1994).

[12] The lower-court and the higher-court may represent either two appellate courts (intermediate and a court of last resort), or, it may represent a trial court and an appellate court. Daughety and Reinganum (1999) analyze a judicial system in which a trial court is engaged in fact finding whereas the appellate court interprets the law.

[13] Many judicial decisions are dichotomous in nature. Also, each decision may be viewed as a collection of binary decisions, i.e., whether such and such reasoning holds in this case and whether such and such evidence is valid or not. Thus, the model could be applied to any of these 'mini-decisions'. On the binary nature of judicial decisions, see Kornhauser (1992).

informative about the true state of the world. If $t = 1$, her signal is always accurate.[14]

In addition to her private information, the judge can make use of earlier decisions by other courts in similar cases. These prior decisions, provide information about the state of the world in the case at hand. Assume (without loss of generality) that previous decisions indicate that the state is likely to be y with probability $q \in (.5, 1)$. The prior belief about the state of the world is therefore $\Pr(w = y) = q$. 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 the previous case. 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.[15] I assume that the information about earlier decision(s) is common knowledge.[16]

Given the prior q, and her own information (s, t), the judge forms the following beliefs, according to Bayes rule:

$$\Pr(w = y|s, t, q) = \left\{ \begin{array}{l} \frac{tq}{tq+(1-t)(1-q)} \text{ if } s = y, \\ \frac{(1-t)q}{(1-t)q+t(1-q)} \text{ if } s = n \end{array} \right\} \tag{1}$$

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

After the judge's decision, the litigants, denoted by L, have the choice of challenging it by appealing to a higher court. Litigants have to bear the cost of an appeal. The cost, regardless of the decision and the state of the world, is a random variable c, $c \sim U[0, 1]$, where each side has to bear c. I assume that only the losing side can appeal. The costs, although not contingent on the decision, are not known prior to the lower-court judge's ruling. The costs are realized by the litigants only after the decision is made, since only then will the losing side be interested in estimating these costs.[17] The information that the litigants possess when they decide whether to appeal includes the judge's decision, d, the body of previous decisions, summarized by q, and the cost, c.[18]

[14] The judge can only make a binary decision, whereas her information is continuous. This restriction is justified by bounded rationality and time constraints. Ottaviani and Sorensen (2002) show that experts who have career concerns will report a binary message even if their signal space is continuous. I still impose this assumption for clarity of exposition.

[15] As explained in the introduction, these previous decisions represent non-binding precedent, that is, either decisions of lower-echelon courts or situations in which the judge has enough discretion to determine whether precedent applies. In the latter interpretation, q may also represent whether the precedent is broad or narrow in its scope.

[16] This is not an important assumption. Similar results would hold if all the players held some beliefs about q or if J only would know q.

[17] This assumption is made for simplicity. Qualitative results are maintained when c is observed by the judge.

[18] The litigants do not have information of their own in the model. The results would be similar if they have information, as long as it is not perfect information. The analysis would be unnecessarily

The higher court H adjudicates the case if an appeal is brought. I assume that H knows w.[19] H must decide then 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 $d^h = R$.

There is one additional player in the game, the evaluator, E, which represents the group that the judge would like to impress. The evaluator forms beliefs about the ability of the judge, i.e., beliefs about t. Denote the expected value of t by τ. The prior belief about the ability of the judge is captured by a uniform distribution on $[.5, 1]$ and is common knowledge. E also knows that previous decisions indicate that $\Pr(w = y) = q$, and obviously the action 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. This implies that $E's$ information about w is endogenous in the model. He can learn w only when an appeal is brought, an event which depends on the judge's and the litigants' behavior.[20]

Objectives:

The higher court H maximizes the probability that the right decision is taken (or, in other words, that the law is interpreted correctly). As a result, H simply takes $d^h = A$ if $d = w$ and $d^h = R$ otherwise.

Assume that the litigants value a favorable decision at 1 and an unfavorable decision at 0. Denote the probability of $d^h = R$ given a decision d by $\Pr(R|d)$. Thus, L appeals if $\Pr(R|d) > c$ and therefore the utility function that the losing litigant perceives is $\max\{0, E(\Pr(R|d)) - c\}$.

The judge J cares about her reputation. She maximizes the expected beliefs about her ability t as perceived by E. Her objective function is therefore $E(\tau)$.

I do not attribute any utility function to the evaluator. Rather, I implicitly assume that E updates his beliefs about the ability of the judge, using Bayes rule.[21]

Timing, strategies, and equilibrium:

The structure of the game is therefore as follows:

Stage 1: J chooses $d \in (y, n)$.

Stage 2: L learns c and decides whether to appeal. If L appeals, the game moves to stage 3. Otherwise, the game moves to stage 4.

---

complicated in this case. In particular, as will be clear soon, the information of the litigants would not affect the behavior of the higher court since the higher court knows the state of the world. It may affect the beliefs of the judge about the probabilities of appeal and the beliefs of the evaluator about the state of the world when no appeal is brought.

[19] It is either the case that the correct state is defined as what the higher court believes it to be, or that the judges in higher positions are more able and hence figure out the correct interpretation of the Law.

[20] The evaluator is likely to represent the public or politicians, who elect judges, but do not have an independent knowledge about the interpretation of the law. They can only learn from public opinions published for example by the Supreme Court.

[21] This can be justified by the need to elect the able judges, or simply by rationality.

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

Stage 4: E forms beliefs τ on t.

The strategy of L is a binary decision whether to appeal given q, d, and c. The strategy of H is $d^h : \{d, w\} \rightarrow \{A, R\}$. The strategy of E is a belief updating function $\tau : \Omega^E \rightarrow [.5, 1]$ where Ω^E represents the information set of E. In particular, it contains the prior uniform distribution over t, and either $\{q, d\}$ or $\{q, d, d^h\}$. Finally, $J's$ strategy is a decision function $\delta : (q, s, t) \rightarrow \{y, n\}$.[22]

The equilibrium concept is that of a Perfect Bayesian Equilibrium, where 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.

3 Results

We begin solving the model by using backward induction. The optimal action of H is obviously to affirm the decision of J if $d = w$ and to reverse it otherwise.

Consider now the litigants L. For any strategy of J in equilibrium, L updates his beliefs about the state of the world using Bayes rule and given the judge's decision d. Let q_d be the probability that the state of the world is d, given a decision d, the prior q, and the conjectured strategy of J, represented by δ. q_d is calculated as follows:

$$q_d \equiv \Pr(w = d|q, d, \delta) = \frac{\Pr(w = d|q) \cdot \Pr(d|w = d, \delta)}{\Pr(w = d|q) \cdot \Pr(d|w = d, \delta) + \Pr(w = d'|q) \cdot \Pr(d|w = d', \delta)}$$

In other words, a decision d is a signal about w with accuracy q_d.

In equilibrium, L anticipates $H's$ behavior in equilibrium and can compute the probability that a decision is reversed. This is simply the perceived probability that the judge is wrong, i.e., $1 - q_d$. Hence, the litigants appeal if

$$1 - q_d > c$$

Given any decision d, and the uniform distribution of costs on $[0, 1]$, the probability of an appeal is $1 - q_d$, and its expected cost is $E(c|c < 1 - q_d) = \frac{1-q_d}{2}$. Note that in equilibrium, q_d is based on the correct conjecture of the judge's strategy, which we explore next.

3.1 Benchmark: efficient judge

Before analyzing the equilibrium with a careerist judge, I analyze the behavior of an efficient judge, as a benchmark. An efficient judge adjudicates the case with the goal of maximizing social welfare, which is defined next. Assume that society values a correct

[22] Mixed strategies will turn out to have no role in equilibria, as shown below, and hence I assume already at this point that the players are using pure strategies.

decision at 1 and places a weight 0 on an incorrect decision. Thus, the objective function of an efficient judge is $E(\Pr(D = w) - 2\theta c)$, where $\theta \geq 0$ is a parameter capturing how much society cares about costs relative to taking the right decision.[23] The parameters characterizing the solution for an efficient judge are therefore $\{q, \theta\}$.

Consider a system with no appeals, i.e., the decision of the judge is the final decision. When the judge takes y, the expected social utility is simply the probability that the decision is correct, that is, $\Pr(w = y|q, s, t)$ whereas if she takes n, the expected social utility is $\Pr(w = n|q, s, t)$. The judge will therefore take $d = y$ for all (s, t) such that $\Pr(w = y|q, s, t) \geq \Pr(w = n|q, s, t)$ and otherwise she takes $d = n$. Thus, by Bayes rule, she takes y whenever $s = y$, or when $s = n$ and $t < q$. This behavior can be described as a cutoff point behavior, with a cutoff point (s^e, t^e), so that when $\Pr(w = y|q, s, t) > \Pr(w = y|q, s^e, t^e)$ the judge takes y and otherwise he takes n.[24] Thus, when no appeals are allowed, $s^e = n$ and $t^e = q$.

When the legal system allows for appeals, the judge is still interested in taking a decision d which she perceives as accurate, i.e., a decision with a higher $\Pr(w = d|q, s, t)$. However, she must also weight now the costs and benefits of an appeal. The benefit from an appeal is that the final decision will be correct since it is taken by the higher court. Thus, the judge is inclined to take a decision with is considered *less accurate* by the litigants, who would therefore bring an appeal with a higher probability. But of course, an appeal is costly. The next lemma shows that despite this additional complexity, the behavior of the efficient judge can still be characterized by a cutoff point strategy:

Lemma 1 *In equilibrium, the efficient judge uses a cutoff point strategy* (s^e, t^e), *i.e., she takes* $d = y$ *whenever* $\Pr(w = y|q, s, t) > \Pr(w = y|q, s^e, t^e)$ *and otherwise she takes* n.

Proof: Suppose that the judge rules d. Her expected utility can be expressed by:

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

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 costs of the decision, i.e., the costs of appeal multiplied by its probability.

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

[23] The costs are multiplied by 2 since both sides have to bear them.

[24] The index e for (s^e, t^e) represents *efficiency*.

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

The right-hand-side of (2) does not depend on (s,t) but on the beliefs of the litigants, who have no knowledge of (s,t). The judge perceives it therefore as constant for each (s,t). On the other hand, by Bayesian updating:

$$\frac{\Pr(w=y|s,t,q)}{\Pr(w=n|s,t,q)}=\left\{\begin{matrix} \frac{tq}{(1-t)(1-q)} \text{ for } s=y \\ \frac{q(1-t)}{t(1-q)} \text{ for } s=n \end{matrix}\right\} \tag{3}$$

Hence, any different (s,t) yields a different value of $\frac{\Pr(w=y|s,t,q)}{\Pr(w=n|s,t,q)}$. 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 if $\Pr(w=y|s,t,q)\geq\Pr(w=h|s^e,t^e,q)$ the judge takes y, and otherwise, she takes n.■

Figure 1 describes an example of a cutoff point strategy for the judge, with $s^e=n$. 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 (in the left). Thus, as we go from left to right, $\Pr(w=n|s,t)$ increases. The cutoff point, (s^e,t^e), is such that for all information (s,t) to the right of it, J takes n, whereas for all information (s,t) to its left, J takes y :

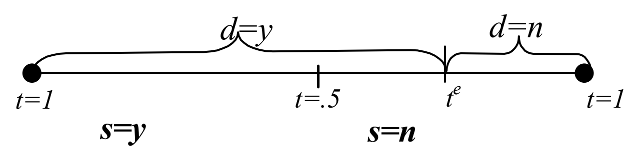

Figure 1: the cutoff point strategy of an efficient judge

Equilibrium means that (2) is satisfied, subject to $0\leq q_d\leq 1$ for $q_d=q_d(s^e,t^e)$ and is characterized in the first Proposition:

Proposition 1 *When J maximizes social welfare, there is a unique informative equilibrium, in which $s^e=n$. That is, J takes d=y when s=y or when s=n and t<$t^e(q,\theta)$. Moreover, $t^e(q,\theta)$ is an increasing function in q and in θ. In equilibrium, an appeal is more likely when $d=n$.*

Proof: see the appendix.■

To understand the intuition for Proposition 1, consider again the solution to the judge's problem when no appeals are allowed, i.e., when $s^e=n$ and $t^e=q$. If the judge were to use this cutoff strategy when appeals *are* allowed, then the litigants, conjecturing correctly her behavior, form beliefs about the possibility that each of her decisions is correct. In particular, the appendix shows that $q_y>q_n$, that is, they believe that a decision that follows previous ones is more likely to be correct. This implies that $d=y$ is cheaper, but less likely to be corrected. Thus, if θ is low enough, i.e., costs are not as important, the efficient judge's cutoff point satisfies $t^e < q$, because she rather spends the costs of an appeal to correct a wrong decision. If costs are important, she opts for the cheaper decision, and hence she takes y more often, even for higher values of t. In this case, $t^e > q$. Thus, the cutoff point increases with θ.

3.2 A careerist judge

We are now ready to analyze the behavior of a careerist judge. Recall that the careerist judge would like to impress an evaluator, who assesses the likelihood that she has accurate private information. Let $\tau(d, w, \delta)$ denote the updated belief of E about the type t of the judge, if E believes that J uses some strategy δ, the judge's decision is d and E were to know the state of the world w. That is, $\tau(y, y, \delta)$ denotes the beliefs of E when J takes y and she is correct. Similarly, $\tau(y, n, \delta)$ denotes the updated belief of E when $d = y$ but E were to know that $w = n$. And so on.

In the game however, the evaluator does not observe the state of the world but has to form beliefs about it. The evaluator will therefore attribute the reputation $\tau(d, d, \delta)$ to the judge only with some probability, the probability with which he thinks that $d = w$. Let $\tilde{p}(d)$ stand for the probability with which the careerist judge believes that the evaluator believes that her decision is correct. Formally, $\tilde{p}(d) = E(p(w = d|\Omega^E)|\Omega^J)$, where Ω^i stands for information held by player i. The expected utility of J from a decision d can be expressed as

$$\tilde{p}(d)\tau(d, d, \delta) + (1 - \tilde{p}(d))\tau(d, d', \delta). \tag{4}$$

If no appeal is brought, which happens with probability q_d, then E believes that the probability that the judge is correct is also q_d. If an appeal is brought, an event which happens with probability $1 - q_d$, the beliefs of E depend on the affirmation or reversal decision of H. If $d^h = A$, then E believes that the judge is correct with probability 1. J believes that this occurs with probability $p(w = d|q, s, t)$. If $d^h = R$, which occurs, in the eyes of J, with the probability that she is wrong, then E believes that the judge is correct with probability 0. Thus, we can express $\tilde{p}(d)$, the probability that J attaches to the event that E believes that d is correct given her own information (s, t), as

$$\tilde{p}(d) = (1 - q_d)p(w = d|q, s, t) + q_d^2 \tag{5}$$

We can see how each of the judge's decisions, through the appeal process, induces a different probability distribution over the information about the state of the world that the evaluator may gather. In other words, the judge, upon ruling y or n, decides which probability distribution over signals about the state of the world E will encounter at the end of the judicial process. If she rules y, with probability q_y the evaluator would believe that the state is y with probability q_y and with probability $1 - q_y$, the distribution over beliefs is degenerate over the right state. If she rules n, both the probabilities and the beliefs of E are different. The judge therefore controls the mean and the higher moments of the distribution over signals about the state of the world that E receives.[25]

Intuitively, the belief of J that she would be perceived correct by E is increasing in $p(w = d|q, s, t)$. Thus, J believes that her own information is correlated with that of E. The greater the probability she attaches to the event that she is correct, the greater the probability she attaches to the event that E knows that she is correct. This feature would discipline the judge to behave informatively, even if E does not know w for sure. We can then establish:

Lemma 2 *In an informative equilibrium the careerist judge uses a cutoff point strategy* (s^c, t^c), *that is she takes* y *whenever* $\Pr(w = y|q, s, t) \geq \Pr(w = y|q, s^c, t^c)$ *and otherwise she takes* n.[26]

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

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

re-arranging, and plugging the expressions for $\tilde{p}(n)$ and $\tilde{p}(y)$ from (5), we get:

$$\frac{\Pr(w=y|s,t,q)}{\Pr(w=n|s,t,q)} = \frac{\pi(n,n,\delta)(q_n^2+1-q_n)+\pi(n,y,\delta)(q_n-q_n^2)-\pi(y,y,\delta)q_y^2-\pi(y,n,\delta)(1-q_y^2)}{\pi(y,y,\delta)(q_y^2+1-q_y)+\pi(y,n,\delta)(q_y-q_y^2)-\pi(n,n,\delta)q_n^2-\pi(n,y,\delta)(1-q_n^2)} \qquad (6)$$

The right-hand-side of (6) is fixed for all (s, t), since these are the beliefs of the evaluator. The evaluator does not know (s, t) and hence cannot condition his beliefs on this information. The left-hand-side of (6), on the other hand, changes with (s, t), as in the Proof of Lemma 1. Hence, any different (s, t) yields a different value of $\frac{\Pr(w=y|q,s,t)}{\Pr(w=n|q,s,t)}$. This implies that there is (at most) a unique (s^c, t^c) that satisfies equation (6). Thus, there is a unique cutoff point (s^c, t^c), such that if $\Pr(w = y|q, s, t) \geq \Pr(w = y|q, s^c, t^c)$ the judge takes y, and otherwise, she takes n.■

The strategy of the careerist judge is therefore similar to that of the efficient judge, who also uses a cutoff strategy. We therefore have to check what cutoff point the reputational 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 a cutoff point strategy $\delta^c = (s^c, t^c)$ and if he were to know w:

Lemma 3 *(i) For any action* d, $\pi(d, d, \delta^c) > \pi(d, d', \delta^c)$, *(ii) If* $s^c = n$, *then* $\pi(n, n, \delta^c) > \pi(y, y, \delta^c)$ *and* $\pi(n, y, \delta^c) > \pi(y, n, \delta^c)$, *(iii) If* $s^c = y$, *then* $\pi(y, y, \delta^c) > \pi(n, n, \delta^c)$ *and* $\pi(y, n, \delta^c) > \pi(n, y, \delta^c)$.

[25] The feature that J controls the higher moments of the distribution of signals about the state of the world that E receives, does not exist in principal-agent models for example, where the agent controls the mean but not the variance of the signals that the principal receives.

[26] The superscript c denotes *careerist* judge.

Proof: see the appendix.■

The Lemma follows from Bayesian updating. The first part asserts that the reputation of J is higher if she takes 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 strategy is responsive to her signal. In addition, the lemma asserts the following; if $s^c = n$, the reputation that E attributes to those who take n, whether they succeed or fail in taking the right decision, is higher than the reputation they receive when they take y. Intuitively, when $s^c = n$, D takes n only if $t > t^c$ (as in Figure 1, which describes an efficient judge). 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 take y.

The next lemma helps us to focus our analysis:

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

Proof: see the appendix.■

Intuitively, if $s^c = y$, then previous decisions are contradicted too often. The evaluator may realize that a judge who contradicts is not necessarily an able one. Higher reputation would be attributed to those who follow previous decisions, i.e., $\pi(y, y, \delta^c) > \pi(n, n, \delta^c)$ and $\pi(y, n, \delta^c) > \pi(n, y, \delta^c)$. Moreover, types with $s = y$, are more likely to take the right decision when they follow previous decisions. That is, for these types, $\tilde{p}(y) > \tilde{p}(n)$. If they take n, not only it is considered inaccurate because too many types follow this route, also if an appeal is brought, they are likely to be found wrong. A decision for y is more likely to provide them with the reputation from being correct, and hence no such type with $s = y$ can be indifferent, implying that the cutoff point must admit $s^c = n$. In short, judges cannot contradict their predecessors 'too much'.

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 conditions on rational updating by E and L, yield the following characterization result.

Proposition 2 *When J is careerist, there exists a unique informative equilibrium. In the equilibrium, the judge takes $d = y$ if $s = y$ or if $s = n$ and $t < t^c(q)$. The cutoff point $t^c(q)$ increases with q, but is bounded, i.e., $t^c(q) < \hat{t} < 1$ for all q. In equilibrium, an appeal is more likely when $d = n$.*

Proof: see the appendix.■

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, $\tau(n, n, \delta^c) > \tau(y, y, \delta^c)$ and $\tau(n, y, \delta^c) > \tau(y, n, \delta^c)$. The reason is that in equilibrium only those types with sufficient ability allow themselves to contradict previous decisions. The able judges have private information that outweighs the informativeness of past verdicts. The second signal is proving ability by taking the correct decision. A type which takes the correct decision is more likely to be able because the judge's action is responsive to her signal when she uses a cutoff point strategy. Thus, a judge that is reversed has a lower reputation than a judge whose decision is re-affirmed. At the equilibrium cutoff point, the trade-off 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. A higher probability of appeal is bad news since they may get 'caught' by the higher court. Hence, they would rather follow others and be perceived as correct. A non-talented judge would take the risk of taking the wrong decision only if the probability of appeal is believed to be low enough. The more able judges, on the other hand, are encouraged to take decisions which are likely to be appealed, since this will affirm their ability. 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 created in equilibrium.

The Proposition also characterizes the judge's behavior as a function of the parameter q. When q increases, the benefit from following previous decisions, everything else being equal, is higher. This is because the terms of the reputational trade-off change; it becomes more likely to receive the (higher) reputation for taking the correct decision. Hence, more types are inclined to follow previous decisions, that is, the cutoff point t^c increases.

However, a significant portion of types always contradict previous decisions, since the value of $t^c(q)$ is bounded. For example, when $q \to 1$, the cutoff point t^c is strictly below 1 and there is a positive measure of types who contradict. In particular, I find that t^c is bounded by 0.625.[27] Thus, when $q \to 1$, all types in $(0.625, 1)$ take the wrong decision, consciously and probably inefficiently.[28]

To see why t^c is bounded, note that if the evaluator conjectures that t^c is very high, for example $t^c \to 1$, then $\pi(n, \cdot, \delta^c) > \pi(y, \cdot, \delta^c)$. 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, $\pi(n, y, \delta^c) > \pi(y, y, \delta^c)$, i.e., even if the judge goes against her predecessors and is found

[27] In particular, this number is computed for the uniform prior distribution of the judge's types. The model is general and can be extended to other distribution. In this case, the upper bound would be different.

[28] I discuss the inefficiency of the careerist judge's decisions in the next section.

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 we do next.

3.3 The distortion due to career concerns

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 $t > \{t^e, t^c\}$ for the efficient and the careerist judge respectively. I now compare the behavior of the two differently motivated judge. If in equilibrium, $t^c = t^e$, then it implies that the judge behaves efficiently. If $t^c > t^e$, it means that the judge takes y more than is efficient, which I term by *excessively following previous decisions*. If, on the other hand, the equilibrium value admits $t^c < t^e$, the judge takes n more then is efficient, which I term by *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 $q \geq \bar{q}$, $\theta(q) = 0$.*

Figure 2 describes the behavior of both judges in equilibrium. The figure shows the area (t^c, t^e) in which the efficient judge takes $d = y$ whereas the careerist judge takes $d = n$:

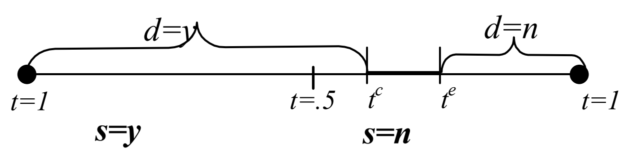

Figure 2: A comparison between the careerist and the efficient judge

The conclusion is that previous decisions, or non-binding precedent, tend to have less effect when the judge has reputational concerns, resulting in an underutilization of readily available information. The intuition is that the careerist judge gains excessive reputation from contradicting previous decisions whereas the efficient judge treats each of these decisions symmetrically. Both judges have induced incentives to take the right decision; the efficient one since she cares about it directly, and the careerist one since she cares about it indirectly, because it provides higher reputation. However, the careerist judge has an additional incentive to contradict previous decisions, since it provides high reputation as well in equilibrium.

The result is derived for high enough values of q, *or,* for high enough values of θ. For high values of q, the intuition is that the efficient judge has to follow previous decisions quite often since it is more likely to be the correct decision. The cutoff point of the careerist judge, on the other hand, is bounded for any q by some $\hat{t}$, as established in the previous section. Thus, no matter how high q is, a significant portion of types has to contradict previous decisions, otherwise, an informative equilibrium cannot be sustained. The reputation for contradicting would be too high, and all types would mimic this behavior. 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 is 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. Note that indeed it is reasonable in the context of our model that θ is high enough. Otherwise, if θ is low, deliberation is not costly, and the judicial system should enforce appeals, or target almost all cases to the higher court, instead of leaving this decision to the litigants.

4 Institutional design

It is now widely recognized that institutions matter; that is, the design of the judicial system affects the behavior of the judge through the incentives it creates. In this section I analyze the effect of different judicial systems on the behavior of a careerist judge and in particular, I analyze which system or institutions can increase social utility.

Social utility is defined in the model as the probability that the correct decision is taken, at the minimum costs. It is a function of the careerist judge's behavior, summarized by the cutoff point t^c. It can be expressed as:

$$U(t^c) = \sum_{w\in\{y,n\}} \Pr(w)\Pr(D = w|t^c, w) - \theta[\sum_{d\in\{y,n\}} (1 - q_d)\Pr(d)]$$

where

$$\Pr(D = w|t^c, w) = \Pr(d = w|t^c, w) + \Pr(d \neq w|t^c, w)(1 - q_d)$$

and

$$\Pr(d) = \sum_{w\in\{y,n\}} \Pr(w)\Pr(d|w)$$

The next lemma proves useful for the design analysis:

Lemma 5 *When the judge is careerist, then for high enough values of θ, social utility $U(t^c)$ increases when t^c increases.*

Proof: see the appendix.■

Given Lemma 5, we can look for instruments that increase the tendency of the judge to follow her predecessors, i.e., for methods that increase t^c.[29]

[29] The reason that we need Lemma 5, is that the litigants behave inefficiently in the model, that is, not

4.1 Who should nominate judges?

In several US states, judges are elected by the public. The Supreme Court Justices are nominated by the US President. The model that we described fits these systems, in the sense that the evaluator, be it the public or the politicians, indeed can only glean information about the correct interpretation of the law from appeals. In other Common Law countries, such as the UK, judges are promoted by their peers, usually Supreme Court Justices (the Law Lords).[30] This implies that the evaluator - the Supreme Court Justices - knows the correct interpretation of the law (i.e., the state of the world) even if they do not adjudicate the case. The reason is either that the correct interpretation of the law is what these Justices believe it to be, or that they are talented enough so they know the correct interpretation without the need for heavy deliberation. Thus, when the Supreme Court Justices review the judge's file and decide whether to promote her or not, they can determine whether she was right or wrong in each case. When the public or even the politicians review past behavior of the judge, they can only guess whether the judge was right or wrong, when no appeal was brought.

A different type of nomination system may provide different reputational incentives to the judge, who will seek to impress different audiences in each case. We can then ask what type of system can encourage efficient behavior and mitigate reputation incentives. I will now compare the behavior of the careerist judge when she is promoted by her superiors, i.e., an evaluator who knows the state w independently of the judicial process, and the behavior of the careerist judge when she is promoted by the public or politicians, i.e., an evaluator who knows the state w only if an appeal is brought.

Intuitively, the less information the evaluator possesses, the easier it is for less able judges to mimic the more able judges and thus create a greater distortion. The next result characterizes first the equilibrium when the evaluator knows w and shows that it may actually be less efficient.

Proposition 4 *(i) When E knows w, the careerist judge takes y when s=y or when s=n and* $t<t^f(q)$.[31] *(ii) The careerist judge follows precedents more often when E learns from appeals than when E has full information, i.e.,* $t^f(q) < t^c(q)$. *For high values of* θ, *social utility is higher when E learns w only from appeals, i.e., when the judge is elected by the public or politicians.*

The case of an evaluator who has full information about the state w is analyzed in other

according to the social preferences. Thus, given this additional distortion, we cannot use the comparison between the efficient judge and the careerist judge to directly deduce that social utility increases when the careerist judge behaves 'more like' the efficient judge. However, Lemma 5 establishes that this is indeed the case.

[30] More precisely, judges are promoted by a committee which contains fellow judges as well as lawyers and politicians.

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

contexts, in particular for managers or experts (see for example Levy (2000), Ottaviani and Sørensen (2002) and Trueman (1994)). The main result is similar to the case of endogenous monitoring; that is, decision makers, or judges, excessively contradict prior information. The intuition for the result is also similar; contradicting previous verdicts allows the judge to signal high ability since only able judges do so. The observation of w by the evaluator allows the evaluator to monitor the judge. In particular, when the judge takes the right decision, it also serves as a signal of ability and hence the judge refrains from contradicting too much, and an informative equilibrium can exist. Since the results are similar for both environments, what is then the effect of endogenous monitoring?

The second part of Proposition 4 states that $t^c(q) > t^f(q)$ that is, the judge follows previous decisions more often under endogenous monitoring. Lemma 5 has established that social utility increases with the cutoff point that the judge uses, when θ is high enough. This implies that social utility from present decisions is higher when the evaluator learns information only from appeals, since she uses a higher cutoff point, i.e., follows her predecessors more often.

The implication of this finding is therefore that when deliberation is too costly, increasing the amount of information available to the evaluator further distorts the decisions of the judge. The above result suggests that at least for some parameter values, judges should be elected by the public and not by the more knowledgable legal community.

What is the intuition for this counter-intuitive finding? When the evaluator learns from appeals, an important feature of the equilibrium is that the probability of appeal is lower after a decision that follows previous ones. Following others becomes a 'safe action' in equilibrium. When the judge follows others, less information about the state of the world will be revealed, therefore, less information about her type will be revealed. Contradicting previous decisions, on the other hand, induces a higher probability of appeal in equilibrium and it is therefore a 'risky' action for the less able types, who are likely to be reviewed and found wrong.[32] But the motivation of these types is to try and hide the truth about their ability; they rather follow others and thereby limit the information held by the evaluator.

With endogenous monitoring, the equilibrium feature of a risky and safe actions mitigates the distorted behavior of the judge; it becomes less rewarding to mimic able judges that contradict previous decisions because one is more likely to be reviewed by the higher court. This induces more types of the judge to behave efficiently. With exogenous monitoring, on the other hand, since the evaluator has full information about the state of the world, no 'safe' or 'risky' actions are possible. The evaluator's information is symmetric, i.e., he knows the state of the world irrespective of the state and the judge's decision.

In his seminal paper about career concerns, Holmström (1982) assumes the existence

[32] I use the words 'risky' and 'safe' although the judge is risk neutral. The results are strengthened if the judge is risk averse.

of a 'safe action'. In his framework, a manager could decide between two actions. One of them is risky, in the sense that the state of the world is revealed if it is taken. The other is safe, in the sense that the state of the world is not revealed. Here, I derive a 'safe action' and a 'risky action' endogenously. A decision that induces a lower probability of appeal becomes a safe action in equilibrium.

4.2 Binding precedent

The intuition gained from the pervious analysis is that the judge behaves more efficiently if contradiction of previous decisions is penalized by a higher risk of appeal. 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. This suggestion, of different legal standards, seems to be in accord with the praxis of the Common Law legal system. The concept of binding precedent 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, when the judge *contradicts previous decisions* and is *correct*, the higher court can also commit to overturn this decision with some probability. Thus, implicitly, this modelling implies that the higher courts may have an incentive to preserve the strength of previous decisions. A contrarian lower-court judge should be reversed even if she is correct. This will increase the incentives of litigants to appeal when a judge contradicts previous decisions. It is not clear how the higher court can indeed commit to implement such a strategy, which induces it to take decision which it perceives as wrong. For the time being, I depart from these considerations, and discuss them later on.

Formally, let $\phi \in [0,1]$ denote the probability with which the judge is overruled when $d = n$ and $w = n$. When $d = n$ and $w = y$ she 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 next result characterizes the effect of binding precedent on the judge's behavior.

Proposition 5 *When precedents bind, the judge contradicts precedent more often, that is, $t^c(q)$ is lower for any $\phi > 0$ relative to the case in which $\phi = 0$.*

Surprisingly, binding precedent do not induce the judge to behave more efficiently, and even have the contrary effect. The intuition for this result is as follows. When previous decisions bind, if the judge takes 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 take n and contradict previous decisions, because the loss from being reversed is now lower. Instead of following more often, they do the opposite. This implies, given Lemma 5, that social utility *decreases* when precedent bind.

Given the inefficiency of the praxis of binding precedent as modelled above, we should not worry about the higher court implementing such a strategy; in other words, even if the higher court could commit to behave in the manner described above, social efficiency considerations would induce it not to use such a praxis.

One explanation for the praxis of binding precedents is uniformity, i.e., the moral principle that *alike cases should be treated alike*. An alternative explanation is the economic principle of predictability. The parties can predict in advance the result of the judicial process and refrain from useless deliberation. My model shows that when judges are careerist, there may be a trade-off between these motivations and the efficiency considerations described here.

5 Discussion and Conclusion

In this paper, I show that judges with career concerns contradict previous decisions more than is efficient. By doing so, they pretend to have a more accurate information than the one supplied by their predecessors. The result is derived when the evaluator, the group whom the judge would like to impress, can learn information about the judge's type from the judicial process, i.e., an appeal to a higher court. The reversal or affirmation by the higher court allow the evaluator to monitor the judge. Thus, although the judge has some control over the level of monitoring the evaluator can exercise, through her decisions, I find that such an endogenous monitoring induces the judge to behave more efficiently relative to the case in which the judge is facing an evaluator who knows the correct decision irrespective of the judicial process. This finding allows us to realize what tools may be used to correct the inefficient behavior of lower-court judges. For example, this finding directly implies that judges should be promoted by the public and not by fellow superior judges.

When considering efficiency, I have concentrated on the social utility from present decisions. I have therefore ignored considerations such as electing the more able judges. Thus, when assessing which nomination system is better, an intertemporal trade-off should be established. If the public elects judges, the judge behaves more efficiently at present. But it is not clear if indeed the public elects the better judges, so as to increase the efficiency of future decisions. This intertemporal trade-off awaits future research.

Also, I have tried to assess which institutional tools would increase social utility by inducing the judge to behave more efficiently, taking as given the behavior of the litigants. Since the litigants also behave inefficiently in the model, i.e., bring appeals according to their own considerations and not according to social welfare considerations, it is possible to analyze tools that would increase social utility by correcting the behavior of the litigants (and through it, maybe affecting the judge as well). Such tools may be fees or subsidies for appeals, or a loser-pays-all system.

Finally, there are many ways to think of reputation motives. In this paper I have used ability as a desired trait for a judge. However, it is also probable that the judge is trying to prove to her evaluators, be it the public, politicians or higher-court judges, that she shares their preferences regarding the interpretation of the constitution. For example, Schauer (1997) claims that a judge of an intermediate appeals court who wants to be promoted to a position of judge of a court of last resort faces a trade-off. On the one hand, such a judge would be promoted if she adopts the views of the Supreme Court, i.e., follows its precedents. On the other hand, she is much more likely to receive academic recognition for formulating new doctrines by departure from existing precedents. The model presented here could be accommodated to explore the implication of such reputation motive.[33]

References

[1] Banerjee, A. (1992), "A Simple Model of Herd Behavior",*Quarterly Journal of Economics,* vol. 107, 797-817.

[2] Bikhchandani, S., D. Hirshleifer and I. Welch (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", *Journal of Political Economy,* vol. 100, 992-1026.

[3] Crawford, V. and J. Sobel (1982), "Strategic Information Transmission", *Econometrica,* vol. 50, 1431-1451.

[4] Daughety, A.F. and J.F. Reinganum (2000), "Appealing judgments", *Rand Journal of Economics 31,* 505-525.

[5] Daughety, A.F. and J.F. Reinganum (1999), "Stampede to Judgment: Persuasive Influence and Herding Behavior by Courts", *American Law and Economics Review 1*, 158-189.

[6] Gelly R. and P. Spiller (1990), "A Rational Choice Theory of Supreme Court Statutory Decisions With Applications to the State Farm and Grove city Cases", *Journal of Law, Economics and Organizations*, vol. 6, 265-268.

[7] Holmström, B. (1982), "Managerial Incentives Problems: A Dynamic Perspective", in *Essays in Economics and Management in Honor of Lars Wahlbeck*, Helsinki: Swedish School Econ.

[8] Kornhauser, L.A. (1992), "Modelling Collegial Courts I: Path-Dependence", International Review of Law and Economics 12, 169-185.

[9] Levy, G. (2000), "Strategic Consultation in the Presence of Career Concerns", Sticerd Discussion Paper, No. TE/00/404.

[33] The need of experts to accumulate reputation for having the same preferences as their principal is analyzed in Morris (2001).

[10] Miceli, T. and Cosgel M. (1994), “Reputation and Judicial Decision-Making”, *Journal of Economic Behavior and Organization*, 31-51.

[11] Milgrom, P.R. and R.J. Roberts (1986), “Relying on the Information of Interested Parties”, *Rand Journal of Economics 17*, 18-32.

[12] Morris, S. (2001), “Political Correctness”, *Journal of Political Economy* 109, 231-265.

[13] Ottaviani, M. and P. Sørensen (2002), “Professional Advice: The Theory of Reputational Cheap Talk”, mimeo, LBS and Copenhagen University.

[14] Posner, R. (2000), “An Economic Analysis of the Use of Citations in the Law”, *American Law and Economic Review, 381-406.*

[15] Posner, R.A. (1993), “What Do Judges Maximize (The Same Things Everybody Else Does)”, *Supreme Court Economic Review*, vol. 3, 1-41.

[16] Rasmusen, E. (1994), “Judicial Legitimacy as a Repeated Game”, *Journal of Law, Economics and Organization*, vol. 10, 63-83.

[17] Salzberger, E. and P. Fenn (1999), ”Judicial Independence: Some Evidence from the English Court of Appeal”, *Journal of Law and Economics*, 831-847.

[18] Schauer, F. (1997), “Incentives, Reputation, and Determinants of Judicial behavior”, presented in the 1997 *American Political Science Association* meetings.

[19] Shavell, S. (1995), “The Appeals Process as a Means of Error Correction”, *Journal of Legal Studies 24*, 379-426.

[20] Spitzer, M. and E. Talley (1998), “Judicial Auditing”, Working Paper no. 98-22, University of Southern California Law School.

[21] Tirole, J. (1999), “A Theory of Collective Reputation” in *The economics of corruption and illegal markets*, American International Distribution Corporation Williston, Vt.

[22] Trueman, B. (1994), “Analyst Forecasts and Herding Behavior”, *Review of Financial Studies,* vol. 7, 97-124.

6 Appendix

The next lemma is useful for the rest of the analysis:

Lemma 6 *When J uses a cutoff point strategy (s^*, t^*), such that she takes y whenever $\Pr(w = y|q, s, t) > \Pr(w = y|q, s^*, t^*)$, then: (i) $q_n(n, t^*)$ increases with t^* and $q_n(y, t^*)$ decreases with t^*, whereas $q_y(n, t^*)$ decreases with t^* and $q_y(n, t^*)$ increases with t^* : (ii) $q_n(n, t^*) > q_n(y, t^*)$ and $q_y(n, t^*) < q_y(y, t^*)$ (iii) $q_y(s^*, t^*) > prob(w = y) = q$ and $q_n(s^*, t^*) > 1 - q$ (iv) $q_y(s, t^*) > \Pr(w = y|q, s, t)$ and $q_n(s, t^*) > \Pr(w = n|q, s, t)$ (v) $\exists!(\tilde{s}, \tilde{t}(q))$ such that $q_y(\tilde{s}, \tilde{t}(q)) = q_n(\tilde{s}, \tilde{t}(q))$, where $\tilde{s} = n$ and $\tilde{t}(q) > q$.*

Proof of Lemma 6: When J uses a cutoff point strategy, $\delta^* = (s^*, t^*)$, we can write $q_d(s^*, t^*) = \Pr(w = d|q, s^*, t^*)$ in the following way:

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

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

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

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

$$q_n(y, t^*) = \frac{(1-q)(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv)}{(1-q)(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv) + q(\int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv)}.$$

(i) $q_n(n, t^*)$ increases with t^* :

$$sign\frac{\partial}{\partial t^*}q_n(n, t^*) =$$
$$sign(1-q)qf(t^*)[-t^*\int_{t^*}^{1}(1-v)f(v)dv + (1-t^*)\int_{t^*}^{1} vf(v)dv] =$$
$$sign - t^*\int_{t^*}^{1} f(v)dv + t^*\int_{t^*}^{1} vf(v)dv + \int_{t^*}^{1} vf(v)dv - t^*\int_{t^*}^{1} vf(v)dv =$$
$$sign(1-q)qf(t^*)[-t^*\int_{t^*}^{1} f(v)dv + \int_{t^*}^{1} vf(v)dv] > 0$$

where the last inequality follows since

$$\int_{t^*}^{1} vf(v)dv > \int_{t^*}^{1} t^*f(v)dv.$$

I will now show that $q_n(y, t^*)$ decreases with t^*.

$$sign\frac{\partial}{\partial t^*}q_n(s^* = y, t^*) =$$
$$sign(1-t^*)(\int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv) - t^*(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv) =$$
$$sign(1-q)qf(t^*)(\int_{.5}^{t^*}(1-2t^*)f(v)dv + \int_{t^*}^{1}(1-t^*-v)f(v)dv) < 0.$$

where the last inequality follows since $v > \frac{1}{2}$ and $t^* > \frac{1}{2}$. The proof of the second part of the claim is analogous.$\square$

(ii). $q_n(n, t^*) > q_n(y, t^*)$:

$$\frac{(1-q)\int_{t^*}^{1} vf(v)dv}{(1-q)\int_{t^*}^{1} vf(v)dv + q\int_{t^*}^{1}(1-v)f(v)dv} >$$
$$\frac{(1-q)(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv)}{(1-q)(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv) + q(\int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv)} \iff$$
$$\int_{t^*}^{1} vf(v)dv(1 - \int_{t^*}^{1} vf(v)dv) > \int_{t^*}^{1}(1-v)f(v)dv(\int_{t^*}^{1} vf(v)dv + \int_{.5}^{t^*} f(v)dv) \iff$$
$$\int_{t^*}^{1} vf(v)dv(1 - \int_{t^*}^{1} vf(v)dv) > \int_{t^*}^{1}(1-v)f(v)dv(1 - \int_{t^*}^{1}(1-v)f(v)dv).$$

Since

$$\int_{t^*}^{1} vf(v) > \frac{1}{2} > \int_{t^*}^{1} (1-v)f(v)dv,$$

the above inequality holds whenever

$$\int_{t^*}^{1} vf(v) < 1 - \int_{t^*}^{1} (1-v)f(v) \Leftrightarrow \int_{t^*}^{1} f(v) < 1$$

which is satisfied. The second part of the claim, i.e., that $q_y(n,t^*) < q_y(y,t^*)$, has an analogous proof.□

(iii)- $q_y(s^*,t^*) > q$: By the above claims, it is enough to show that $q_y(n,t^*) > q$:

$$q_y(n,t^*) > q \Longleftrightarrow$$

$$\frac{q(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv)}{q(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv) + (1-q)(\int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv)} > q \Longleftrightarrow$$

$$\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv > \int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv \Longleftrightarrow$$

$$\int_{t^*}^{1} vf(v)dv > \int_{t^*}^{1}(1-v)f(v)dv \Longleftrightarrow$$

$$\int_{t^*}^{1}(2v-1)f(v)dv > 0.$$

Again, the second part, i.e., that $q_n(s^*,t^*) > 1-q$, has an analogous proof which is therefore omitted.□

(iv) $q_y(s^*,t^*) > p(w=y|q,s^*,t^*)$:

We need to show that $q_y(n,t^*) > \Pr(w=y|q,n,t^*)$ and that $q_y(y,t^*) > \Pr(w=y|q,y,t^*)$:

$$q_y(n,t^*) > p(w=y|n,t^*) \Longleftrightarrow$$

$$\frac{q(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv)}{q(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv) + (1-q)(\int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv)} > \frac{q(1-t^*)}{q(1-t^*)+t^*(1-q)} \Longleftrightarrow$$

$$t^*(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv) > (1-t^*)(\int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv)$$

where the last inequality is satisfied since $t^* > 1-t^*$ and

$$(\int_{.5}^{1} vf(v)dv + \int_{.5}^{t^*}(1-v)f(v)dv) > (\int_{.5}^{1}(1-v)f(v)dv + \int_{.5}^{t^*} vf(v)dv)$$

as shown above.

$$q_y(y,t^*) > p(w=y|y,t^*) \Longleftrightarrow$$

$$\frac{q\int_{t^*}^{1} vf(v)dv}{q\int_{t^*}^{1} vf(v)dv + (1-q)\int_{t^*}^{1}(1-v)f(v)dv} > \frac{qt^*}{qt^*+(1-t^*)(1-q)} \Longleftrightarrow$$

$$\frac{\int_{t^*}^{1} vf(v)dv}{\int_{t^*}^{1}(1-v)f(v)dv} > \frac{t^*}{1-t^*},$$

but

$$\int_{t^*}^{1} vf(v)dv > t^* \text{ and } \int_{t^*}^{1}(1-v)f(v)dv < 1-t^*,$$

hence the above is satisfied. We can use the analogous proof to show that $q_n(s^*,t^*) > p(w = n|q,s^*,t^*)$.□

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

$q_y(y,\frac{1}{2}) > q_n(y,\frac{1}{2})$, and by the above claims this holds for all $t^* > \frac{1}{2}$ and $s^* = y$. On the other hand, $q_y(n,t^* \to 1) \to q$ and $q_n(n,t^* \to 1) \to 1$. Since $q_n(n,t^*)$ increases with t^* and $q_y(n,t^*)$ decreases with t^*, there exists a unique $\tilde{t}(q)$ such that $q_y(n,\tilde{t}(q)) = q_n(n,\tilde{t}(q))$ and for all $t^* < (>)\tilde{t}(q)$, $q_y(n,t^*) > (<) q_n(n,t^*)$. With the uniform distribution, i.e., $f(v) = 2$, then $q_y(n,q) = \frac{2-q}{3-2q} \geq \frac{1+q}{1+2q} = q_n(n,q)$ for all $q \geq .5$ which implies that $\tilde{t}(q) > q$.$\square$

This completes the proof of Lemma 6.$\blacksquare$

Proposition 1 *When J maximizes social welfare, there is a unique informative equilibrium, in which $s^e = n$. That is, J takes d=y when s=y or when s=n and t<$t^e(q,\theta)$, for $t^e(q,\theta) \in (.5, \tilde{t}(q))$ where $\tilde{t}(q)$ satisfies $q_y(n,\tilde{t}(q)) = q_n(n,\tilde{t}(q))$. Moreover, $t^e(q,\theta)$ is an increasing function in q and in θ and $t^e(q,\theta) \to_{\theta\to\infty} \tilde{t}(q)$.*

Proof of Proposition 1. An equilibrium is a solution to equation (1) in the text given true beliefs of L about the strategy of J. Thus, it is a solution to the fixed point equation in s^e, t^e :

$$\frac{\Pr(w=y|q,s^e,t^e)}{\Pr(w=n|q,s^e,t^e)} = \frac{q_y(s^e,t^e) - \theta\beta(s^e,t^e)}{q_n(s^e,t^e) + \theta\beta(s^e,t^e)} \tag{7}$$

where $\beta(s^e,t^e) = (1 - q_n(s^e,t^e))^2 - (1 - q_y(s^e,t^e))^2$.

Step 1. Existence and characterization.

When $s^e = y$,

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

The second inequality follows because $q_y(y,t^e) > q_n(y,t^e)$. The first inequality holds for all t iff:

$$\frac{qt}{(1-q)(1-t)} > \frac{\frac{q(1+t)}{q(1+t)+(1-t)(1-q)}}{\frac{(1-q)(2-t)}{(1-q)(2-t)+qt}}$$

which is satisfied for all $t > \frac{1}{2}$ and $q > \frac{1}{2}$. On the other hand, when $s^e = n$ and $t^e = \tilde{t}(q) > q$,

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

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

Step 2. Uniqueness:

I will show that at the equilibrium value of t^e, whenever $\frac{\partial}{\partial t}\frac{q_y(n,t)-\theta\beta(n,t)}{q_n(n,t)+\theta\beta(n,t)} < 0$, then:

$$|\frac{\partial}{\partial t}\frac{\Pr(w=y|q,n,t)}{\Pr(w=n|q,n,t)}| > |\frac{\partial}{\partial t}\frac{q_y(n,t) - \theta\beta(n,t)}{q_n(n,t) + \theta\beta(n,t)}|$$

which is a sufficient condition for uniqueness.[34]

Consider first $\frac{\Pr(w=y|q,n,t)}{\Pr(w=n|q,n,t)} = \frac{q(1-t)}{t(1-q)}$. Then:

[34] Note that $\frac{\partial}{\partial t}\frac{p(w=y|n,t)}{p(w=n|n,t)} < 0$, and that if $\frac{\partial}{\partial t}\frac{q_y(n,t)-\theta\delta(q_y(n,t),q_n(n,t))}{q_n(n,t)+\theta\delta(q_y(n,t),q_n(n,t))} > 0$, uniqueness is assured.

$$|\frac{\partial}{\partial t}\frac{\Pr(w=y|q,n,t)}{\Pr(w=n|q,n,t)}| = \frac{q}{(1-q)t^2}$$

Now consider

$$|\frac{\partial}{\partial t}\frac{q_y(n,t)-\theta(\beta(n,t)}{q_n(n,t)+\theta\beta(n,t)}|$$
$$=\frac{1}{q_n+\theta\beta}[\frac{\partial q_y}{\partial t}(-1+\theta\frac{\partial \beta}{\partial q_y}+\theta\frac{\partial \beta}{\partial q_y}\frac{q_y-\theta\beta}{q_n+\theta\beta})+\frac{\partial q_n}{\partial t}((1+\theta\frac{\partial \beta}{\partial q_n})\frac{q_y-\theta\beta}{q_n+\theta\beta}+\theta\frac{\partial \beta}{\partial q_n})]$$

but since

$$\frac{\partial \beta}{\partial q_y}>0, \frac{\partial \beta}{\partial q_n}<0\ , \frac{\partial q_n}{\partial t}>0\ and\ \frac{\partial q_y}{\partial t}<0,$$

it is enough to show that

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

which in equilibrium can be expressed by:

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

Let $q_x\in\{q_y,q_n\}$ such that $q_x(1-q_x)=\max\{q_y(1-q_y),q_n(1-q_n)\}$. It is therefore sufficient to prove that:

$$\frac{q}{t}>\frac{2q_x(1-q_x)}{q_n}[\frac{(1+t)(1-q)+q(2-t)}{(2-t)(1+t)}] \tag{8}$$

assume that $q_x=q_n$. We then have to show that (8) holds, i.e., that:

$$\frac{q}{t}>2(1-q_n)[\frac{(1+t)(1-q)+q(2-t)}{(2-t)(1+t)}]\Longleftrightarrow$$
$$\frac{q(2-t)(1+t)}{(1+t)(1-q)+q(2-t)}>\frac{2q(1-t)t}{(1+t)(1-q)+q(1-t)}\Longleftrightarrow$$
$$q(1+t)(1-q)(2-t+t^2)>q^2(2-t)(1-t)(t-1)$$

which is satisfied because the left-hand-side is greater than 0 whereas the right-hand-side is negative.

Now let $q_x=q_y$. Condition (8) reduces to:

$$\frac{q}{t}>\frac{2q_y(1-q_y)}{q_n}[\frac{(1+t)(1-q)+q(2-t)}{(2-t)(1+t)}]\Longleftrightarrow$$
$$(1-q)^2t^2(1-t^2+2t)+q^2(2-t)(2(1-t^2)+3t+t^3)>2q(1-q)(1+t)t(-2t^2-2+t^3)$$

which is satisfied because the left-hand-side is positive and the right-hand-side is negative.

Step 3. $t^e(q,\theta)$ increases with θ and $q_y(n,t^e)>q_n(n,t^e)$:

Since the equilibrium is unique, and for all θ there exists an equilibrium $t^e<\tilde{t}(q)$, for $\tilde{t}(q)$ that satisfies $q_y(s,\tilde{t}(q))=q_n(s,\tilde{t}(q))$, then in equilibrium, $q_y>q_n$. Moreover, $\frac{\Pr(w=y|q,n,t^e)}{\Pr(w=n|q,n,t^e)}$ is constant for all θ whereas $\frac{q_y(n,t^e)-\theta\beta(n,t^e)}{q_n(n,t^e)+\theta\beta(n,t^e)}$ 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. 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 $q_y(n,t^e)=q_n(n,t^e)\to t^e(q,\theta)\to\tilde{t}(q)$.

Step 4. $t^e(q,\theta)$ increases with q :

By total differentiation of the equilibrium condition:

$$\frac{dt}{dq}|_{t=t^e} = \frac{\frac{\partial}{\partial q}\frac{q_y(n,t)-\theta\beta(n,t)}{q_n(n,t)+\theta\beta(n,t)} - \frac{\partial}{\partial q}\frac{\Pr(w=y|q,n,t)}{\Pr(w=n|q,n,t)}}{\frac{\partial}{\partial t}\frac{\Pr(w=y|q,n,t)}{\Pr(w=n|q,n,t)} - \frac{\partial}{\partial t}\frac{q_y(n,t)-\theta\beta(n,t)}{q_n(n,t)+\theta\beta(n,t)}}|_{t=t^e}$$

I will show that when $\frac{\Pr(w=y|q,n,t)}{\Pr(w=n|q,n,t)} = \frac{q_y(n,t)-\theta\beta(n,t)}{q_n(n,t)+\theta\beta(n,t)}$,

$$\frac{\partial}{\partial q}\frac{\Pr(w=y|q,n,t)}{\Pr(w=n|q,n,t)} > \frac{\partial}{\partial q}\frac{q_y(n,t)-\theta(n,t)}{q_n(n,t)+\theta(n,t)}$$

which, along with step 2, implies that $\frac{dt}{dq}|_{t=t^e} > 0$.

As in step 2, it is enough to show the inequality for $\theta = 0$, i.e., we have to show that:

$$\frac{(1-t)}{t(1-q)^2} > \frac{1}{q_n}\left[\frac{\partial q_y}{\partial q} - \frac{\partial q_n}{\partial q}\frac{q(1-t)}{t(1-q)}\right] \iff$$

$$\frac{(1-t)}{t(1-q)^2} > \frac{1}{q_n}\left[\frac{t(2-t)}{((1-q)t+q(2-t))^2} + \frac{(1+t)(1-t)}{((1-q)(1+t)+q(1-t))^2}\frac{q(1-t)}{t(1-q)}\right] \iff$$

$$\frac{(1-t)}{t(1-q)^2} > \frac{1}{q_n}\left[\frac{q_y(1-q_y)}{q(1-q)} + \frac{q_n(1-q_n)(1-t)}{t(1-q)^2}\right]$$

but since $q_y > q_n$ in equilibrium, and for all q and t, $q_y > 1-q_n$, it is sufficient to show that:

$$\frac{(1-t)}{t(1-q)^2} > \frac{q_n(1-q_n)}{q_n}\left[\frac{1}{q(1-q)} + \frac{(1-t)}{t(1-q)^2}\right] \iff$$

$$\frac{q(1-t)}{t(1-q)+q(1-t)} > \frac{q(1-t)}{(1-q)(1+t)+q(1-t)}$$

which holds for all $t, q \in [.5, 1]$. This implies that $\frac{dt}{dq}|_{t=t^e} > 0$ which completes the proof.$\blacksquare$

Lemma 3 *(i) For any action d, $\pi(d,d,\delta^*) > \pi(d,d',\delta^*)$, (ii) If $s^* = n$, then $\pi(n,n,\delta^*) > \pi(y,y,\delta^*)$ and $\pi(n,y,\delta^*) > \pi(y,n,\delta^*)$, (iii) If $s^* = y$, then $\pi(y,y,\delta^*) > \pi(n,n,\delta^*)$ and $\pi(y,n,\delta^*) > \pi(n,y,\delta^*)$.*

Proof of Lemma 3. $\tau(d,w,\delta^*)$ 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 δ, i.e., $\tau(d,w,\delta^*) = \int_{.5}^{1} tf(t|d,w,\delta^*)dt$. To show that

$$\int_{.5}^{1} tf(t|d,d,\delta^*)dt > \int_{.5}^{1} tf(t|d,d',\delta^*)dt$$

we can use the MLRP property, i.e., show that

$$\frac{f(t|d,d,\delta^*)}{f(t'|d,d,\delta^*)} \geq \frac{f(t|d,d',\delta^*)}{f(t'|d,d',\delta^*)}$$

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 in this case, since whenever J uses a cutoff strategy, then:

$$f(t|y,y,\delta^*) = \left\{\begin{matrix} t \text{ if } t > t^* \\ 1 \text{ otherwise} \end{matrix}\right\},\ f(t|n,n,\delta^*) = \left\{\begin{matrix} t \text{ if } t > t^* \\ 0 \text{ otherwise} \end{matrix}\right\}$$

and

$$f(t|y,n,\delta^*) = \left\{\begin{matrix} t \text{ if } t > t^* \\ 1 \text{ otherwise} \end{matrix}\right\},\ f(t|n,y,\delta^*) = \left\{\begin{matrix} 1-t \text{ if } t > t^* \\ 0 \text{ otherwise} \end{matrix}\right\}$$

implying that

$$\frac{f(t|d,d,\delta^*)}{f(t|d,d',\delta^*)} = \left\{ \begin{array}{c} \frac{t}{1-t} \text{ if } t > t^* \\ 1 \text{ otherwise} \end{array} \right\}$$

for all d, which increases with t.

(ii) When $s^* = n$, $\tau(y,y,\delta^*) < \tau(n,n,\delta^*)$ and $\tau(y,n,\delta^*) < \tau(n,y,\delta^*)$:

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

$$\int_{.5}^{1} tf(t|n,n,\delta^*)dt > \int_{.5}^{1} tf(t|y,y,\delta^*)dt$$

and that

$$\int_{.5}^{1} tf(t|n,y,\delta^*)dt > \int_{.5}^{1} tf(t|y,n,\delta^*)dt$$

by using the MLRP and in particular showing that $\frac{f(t|n,y,\delta^*)}{f(t|y,n,\delta^*)}$ and $\frac{f(t|n,n,\delta^*)}{f(t|y,y,\delta^*)}$ increase with t. Since

$$\frac{f(t|n,n,\delta^*)}{f(t|y,y,\delta^*)} = \frac{f(t|n,y,\delta^*)}{f(t|y,n,\delta^*)} = \left\{ \begin{array}{c} 1 \text{ if } t > t^* \\ 0 \text{ otherwise} \end{array} \right\}$$

the result follows. (iii) The results for $s^* = y$ follow from symmetry and part (ii).■

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

Proof of Lemma 4: The expected utility from an action d is

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

When $s^* = y$, then $\tau(y,n,\delta^*) > \tau(n,y,\delta^*)$ and $\tau(y,y,\delta^*) > \tau(n,n,\delta^*)$, by Lemma 3. I will now show that when $s^* = y$, for any type $s = y$, also $\tilde{p}(y) > \tilde{p}(n)$. This implies that the expected utility from ruling y is greater than the expected utility from ruling n for all $s = y$ because $\tau(y,y,\delta^*) > \tau(y,n,\delta^*)$ and $\tau(n,n,\delta^*) > \tau(n,y,\delta^*)$, and hence no type with $s = y$ can be indifferent:

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

The first inequality follows because $q_y(y,t) > q_y(n,t)$ and $\Pr(w=y|q,y,t) > \Pr(w=n|q,y,t)$. The second inequality follows because $q_n > \Pr(w=n|q,y,t)$. This completes the proof.■

Proposition 2 *When J is careerist, there exists a unique informative equilibrium. In the equilibrium, the judge takes $d = y$ if $s = y$ or if $s = n$ and $t < t^*(q)$, for $t^*(q) \in (.5, \tilde{t}(q))$. The cutoff point $t^*(q)$ increases with q, but is bounded, i.e., $t^*(q) < \hat{t} < 1$ for all q.*

Proof of Proposition 2. If $s^* = n$ and $t^* = \tilde{t}(q)$, then $\tilde{p}(y) < \tilde{p}(n)$ while $\tau(n, n, \delta^*) > \tau(y, y, \delta^*)$ and $\tau(n, y, \delta^*) > \tau(y, n, \delta^*)$. This implies that the utility from ruling n is higher than the utility from ruling y. Hence, along with Lemma 4, this assures existence. To show the rest we now prove the following Lemma:

Lemma 7 *(i) When $s^* = n$, $\tilde{p}(y)$ decreases with t whereas $\tilde{p}(n)$ increases with t and $\tilde{p}(y)$ increases with q whereas $\tilde{p}(n)$ decreases with q (ii) For all q, $t^*(q) < \hat{t} < 1$. (iii) $\tau(n, y, \delta^*)$ and $\tau(n, n, \delta^*)$ increase with t whereas $\tau(y, n, \delta^*)$ and $\tau(y, y, \delta^*)$ decrease with t.*

Proof of Lemma 7. (i) Recall that

$$\tilde{p}(y) = q_y^2 + (1 - q_y)\Pr(w = y|q, s, t),$$

Hence:

$$\frac{\partial \tilde{p}(y)}{\partial t}|_{s^*=n} = (2q_y - \Pr(w = y|q, n, t))\frac{\partial q_y}{\partial t} + (1 - q_y)\frac{\partial \Pr(w = y|q, n, t)}{\partial t}$$

but when $s^* = n$, $\frac{\partial q_y}{\partial t} < 0$. Also, $\frac{\partial \Pr(w=y|q,n,t)}{\partial t}$ and $2q_y - \Pr(w = y|q, s, t) > 0$ by Lemma 5. Similarly,

$$\frac{\partial \tilde{p}(n)}{\partial t} = (2q_n - \Pr(w = n|q, n, t))\frac{\partial q_n}{\partial t} + (1 - q_n)\frac{\partial \Pr(w = n|q, n, t)}{\partial t} > 0.$$

An analogous analysis holds for the derivatives w.r.t. q.

(ii) Let s* $= n$ and $t = t^*$. I will show that there is a unique $\hat{t} < 1$ satisfying $\tau_{\hat{t}}(y, y, \delta^*) = \tau_{\hat{t}}(n, y, \delta^*)$, and that for all $t > \hat{t}$, $\tau(y, y, \delta^*) < \tau(n, y, \delta^*)$. This implies that an equilibrium with cannot exist, since then the expected utility from ruling n, an average over $\tau(n, y, \delta^*)$ and $\tau(n, n, \delta^*)$ where $\tau(n, n, \delta^*) > \tau(n, y, \delta^*)$, is higher than the expected utility from ruling y, an average over $\tau(y, n, \delta^*)$ and $\tau(y, y, \delta^*)$ where$\tau(y, y, \delta^*) > \tau(y, n, \delta^*)$.

The expression for $\tau(y, y, \delta^*)|_{s^*=n}$ is

$$\tau(y, y, \delta^*)|_{s^*=n} = \frac{\int_{.5}^{1} t^2 f(t)dt + \int_{.5}^{t^*} t(1 - t)f(t)dt}{\int_{.5}^{1} tf(t)dt + \int_{.5}^{t^*}(1 - t)f(t)dt}$$

Taking the derivative of $\tau(y, y, \delta^*)|_{s^*=n}$ w.r.t t^*, it is

$$\frac{d\tau(y, y, \delta^*)}{dt^*}|_{s^*=n} = \frac{(1 - t^*)f(t^*)(t^* - \tau(y, y, \delta^*))}{(\int_{.5}^{1} tf(t)dt + \int_{.5}^{t^*}(1 - t)f(t)dt)^2}$$

Therefore, $\tau(y, y, \delta^*)$ is a monotonically decreasing function as long as $t^* < \tau(y, y, \delta^*)$ and a monotonically increasing function when $t^* > \tau(y, y, \delta^*)$. When $t^* \to .5$, $t^* < \tau(y, y, \delta^*)$ and when $t^* \to 1$, $\tau(y, y, \delta^*) < t^*$. Therefore, there exists t' such that $t' = \tau_{t'}(y, y, \delta^*)$. Moreover, t' is unique, since when $t^* > \tau(y,y,\delta^*)$, $\frac{d\tau(y,y,\delta^*)}{dt^*} < 1$.[35] Thus, for all $t^* < (>)t'$, $t^* < (>)\tau(y,y,\delta^*)$ and $\frac{d\tau(y,y,\delta^*)}{dt^*} < (>)0$.

On the other hand, $\tau(n,y,\delta^*)$ is an average over t for $t > t^*$ and thus increases with t^* for all $t^* > .5$. Also, since only values of $t > t^*$ are included in the computation of $\tau(n,y,\delta^*)$, then $\tau(n,y,\delta^*) > t^*$ for all t^*. By the above, when $t^* \to 1$, $\tau(n,y,\delta^*) > t^* > \tau(y,y,\delta^*)$. When $t^* = .5$, by Lemma 3, $\tau(y,y,\delta^*) = \tau(n,n,\delta^*) > \tau(n,y,\delta^*)$. Then, there must exist some $\hat{t} \in (.5,1)$ satisfying $\tau_{\hat{t}}(y,y,\delta^*) = \tau_{\hat{t}}(n,y,\delta^*)$. Moreover, it must be that $\hat{t} < t'$ because for all $t^* > t'$, $\tau(n,y,\delta^*) > t^* > \tau(y,y,\delta^*)$. Because $\tau(y,y,\delta^*)$ decreases monotonically for $t^* < \hat{t}$ and $\tau(n,y,\delta^*)$ increases monotonically in t^*, $\hat{t}$ is unique.

(iii) The proof of part (ii) showed that $\tau(n,y,\delta^*)$ is increasing in t, whereas similar analysis holds for $\tau(n,n,\delta^*)$. It also showed that $\tau(y,y,\delta^*)$ is decreasing in the range $[.5,\hat{t}]$, and analogous analysis holds for $\tau(y,n,\delta^*)$.■

We are now ready to prove Proposition 2. 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, by Lemma 3 and step (i) of Lemma 7. Also, part (ii) of Lemma 7 showed that the solution $t^*(q)$ is bounded for all q. Finally, to see that $t^*(q)$ increases with q, Lemma 6 showed that $\tilde{p}(y)$ increases with q whereas $\tilde{p}(n)$ increases 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., $t^*(q)$ increases.■

Proposition 3 *For any q, there exists $\theta(q)$, such that for all $\theta \geq \theta(q)$, the careerist judge contradicts precedent more than is efficient, that is, $t^*(q) < t^e(q,\theta)$. Moreover, there exists $\bar{q}$, such that for all $q \geq \bar{q}$, $\theta(q) = 0$.*

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

Lemma 5 *When the judge is careerist, then the social utility $U(t^*)$ increases when t^* increases.*

Proof of Lemma 5 We first have to define the expression for social utility, denoted by $U(t^*)$:

$$U(t^*) =$$

$$q(\int_{.5}^{1} 2v dv + \int_{.5}^{t^*} 2(1-v)dv + \int_{t^*}^{1} 2(1-v)dv(1-q_n(n,t^*)))$$

[35] To see that note that $\sqrt{(1-t^*)f(t^*)t^*} < \int_{.5}^{1} tf(t)dt$ since $\sqrt{(1-t^*)f(t^*)t^*} < \frac{1}{\sqrt{2}} < \frac{3}{4} = \int_{.5}^{1} tf(t)dt$ for a uniform $f(t)$ on $[.5,1]$. Hence, $\sqrt{(1-t^*)f(t^*)(t^*-\tau(y,y))} < (\int_{.5}^{1} tf(t)dt + \int_{.5}^{t^*}(1-t)f(t)dt)$ which implies that $(1-t^*)f(t^*)(t^*-\tau(y,y)) < (\int_{.5}^{1} tf(t)dt + \int_{.5}^{t^*}(1-t)f(t)dt)^2 \to \frac{d\tau(y,y)}{dt^*}|_{s^*=n} < 1$.

$$+(1-q)(\int_{t^*}^{1} 2vdv + (\int_{.5}^{1} 2(1-v)dv + \int_{.5}^{t^*} 2vdv)(1-q_y(n,t^*)$$
$$-\theta(1-q_y)^2(q(\int_{.5}^{1} 2vdv + \int_{.5}^{t^*} 2(1-v)dv) + (1-q)(\int_{.5}^{1} 2(1-v)dv + \int_{.5}^{t^*} 2vdv))$$
$$-\theta(1-q_n)^2(q\int_{t^*}^{1} 2(1-v)dv + (1-q)(\int_{t^*}^{1} 2vdv)$$

The first two lines express the social gain from taking the right decision. This happens when the judge takes 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. This are paid when an appeal is brought, which happens with the probability that a judge rules d multiplied by $1-q_d$.

I will now show that for high enough values of θ, in particular for $\theta > \frac{1}{2}$, $\frac{\partial U(t^*)}{\partial t^*} > 0$:

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

Plugging for the expressions of:

$$\frac{\partial q_y}{\partial t} = \frac{-2q(1-q)}{(q(2-t)+(1-q)t)^2}; \frac{\partial q_n}{\partial t} = \frac{2q(1-q)}{((1-q)(1+t)+q(1-t))^2}$$

The expression becomes:

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

Note that when $t = t^e$, then the first two elements are 0, whereas for any $t < t^e$, the first two elements are positive (see the proof of Proposition 1). Finally, when $\theta > \frac{1}{2}$, the last element is positive since for any $t < \tilde{t}(q)$,

$$(1-q)(1-q_n)^2 - q(1-q_y)^2 > 0.$$

Hence, $\frac{\partial U(t^*)}{\partial t^*} > 0.$■

Proposition 4 *(i) When E knows w, the careerist judge takes y when s=y or when s=n and t<t^f(q). (ii) The careerist judge follows precedents more often when E learns from appeals than when E has full information, i.e., $t^f(q) < t^*(q)$. For high values of q or θ, social utility is higher when the judge is elected by the public or politicians.*

Proof of Proposition 4. (i) For the proof of the first part, see Proposition 1 in Levy (2002). (ii) I will now show that $t^f(q) < t^*(q)$. $t^f(q)$ solves:

$$\Pr(w = y|q, n, t^f(q))\Gamma_y = \Pr(w = n|q, n, t^f(q))\Gamma_n + \Gamma \tag{9}$$

where $\Gamma_y = \tau(y,y,\delta^f) - \tau(y,n,\delta^f)$, $\Gamma_n = \tau(n,n,\delta^f) - \tau(n,y,\delta^f)$ and $\Gamma = \tau(n,y,\delta^f) - \tau(y,n,\delta^f)$.[36] We will show that at $t^f(q)$,

$$\tilde{p}(y)\Gamma_y > \tilde{p}(n)\Gamma_n + \Gamma \tag{10}$$

for $\tilde{p}(d) = (1-q_d)\Pr(w=d|q,n,t^f(q)) + q_d^2$, 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 $t^*(q)$ must admit $t^*(q) > t^f(q)$.

Plugging (9) into (10), we have to show that:

$$\begin{aligned} q_y(q_y - \Pr(w=y|q,n,t^f(q)))\Gamma_y > q_n(q_n - \Pr(w=n|q,n,t^f(q)))\Gamma_n \Leftrightarrow \\ \frac{\Gamma_y}{\Gamma_n} > \frac{q_n(q_n - \Pr(w=n|q,n,t^f(q)))}{q_y(q_y - \Pr(w=y|q,n,t^f(q)))} \end{aligned}$$

However, for all values of t, $\frac{\Gamma_y}{\Gamma_n} > 1$, whereas for all values of t, $\frac{q_n(q_n - \Pr(w=n|q,n,t^f(q)))}{q_y(q_y - \Pr(w=y|q,n,t^f(q)))} < 1$. The proofs for these two claims is omitted and available upon request. This completes the proof.$\blacksquare$

Proposition 5 *When precedents bind, the judge tends to contradict precedent more often, that is, $t^a(q)$ is lower for any $\phi > 0$ relative to the case in which $\phi = 0$.*

Proof of Proposition 5 We will calculate $\tilde{p}(n)$ and show that it increases compared to the case in which $\phi = 0$, whereas $\tilde{p}(y)$ does not change with ϕ. This means that the utility from n is higher for any t, and thus $t^a(q)$ must decrease.

$$\begin{aligned} &\tilde{p}(n)|_{\phi>0} = \\ &(1-q_n+q_n\phi)(p(w=y|s,t)+p(w=n|s,t)\phi)\frac{q_n\phi}{1-q_n+q_n\phi} + \\ &(1-q_n+q_n\phi)p(w=n|s,t)(1-\phi)) + q_n(1-\phi)q_n \end{aligned}$$

The first element 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 the state is n, and then with probability ϕ, the judge is reversed. In this case, the evaluator believes that the state is actually n with probability $\frac{q_n\phi}{1-q_n+q_n\phi}$, 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 describes the beliefs when no appeal takes place, which are q_n.

We will show that $\tilde{p}(n)|_{\phi>0} > \tilde{p}(n)|_{\phi=0} \rightarrow$

$$\begin{aligned} &(1-q_n+q_n\phi)((p(w=y|s,t)+p(w=n|s,t)\phi)\frac{q_n\phi}{1-q_n+q_n\phi} + \\ &\qquad p(w=n|s,t)(1-\phi)) + q_n(1-\phi)q_n > p(w=n|s,t)(1-q_n) + q_n^2 \Longleftrightarrow \\ &q_n\phi(1-q_n) - p(w=n|s,t)\phi(1-q_n) > 0 \end{aligned}$$

which holds since $q_n(s,t) > p(w=n|s,t)$ as established in Lemma 6.$\blacksquare$

[36]The values of Γ_y, Γ_y and Γ_y depend of course on $t^f(q)$.