

University of Virginia Press  
© 2006 by the Rector and Visitors of the University of Virginia  
All rights reserved  
Printed in the United States of America on acid-free paper

*First published 2006*

9 8 7 6 5 4 3 2 1

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Institutional games and the U.S. Supreme Court / edited by James R. Rogers, Roy B. Flemming, and Jon R. Bond.

p. cm. — (Constitutionalism and democracy)

Includes bibliographical references and index.

ISBN 0-8139-2527-4 (cloth : alk. paper)

1. United States. Supreme Court. 2. United States. Congress—Powers and duties.  
3. Separation of powers—United States. 4. Political questions and judicial power—  
United States. I. Title: Institutional games and the US Supreme Court. II. Rogers,  
James R., 1960– III. Flemming, Roy B. IV. Bond, Jon R. V. Series.

KF8742.A5I57 2006

347.73'26—dc22

2005029109

# Institutional Games and the U.S. Supreme Court

EDITED BY

James R. Rogers, Roy B. Flemming, and Jon R. Bond

## Appeals Mechanisms, Litigant Selection, and the Structure of Judicial Hierarchies

CHARLES M. CAMERON AND LEWIS A. KORNHAUSER

The previous essay elaborated on the theme of compliance in a judicial hierarchy. This essay pulls back to ask broader questions about judicial hierarchies that inquire into the logic of hierarchies as ways of minimizing and correcting errors. Three models are developed. The first approaches the question from a “macro” perspective of the adjudicatory system. It identifies conditions on the relative rates at which wrongly decided cases are appealed and the rates at which errors are corrected and introduced by an appellate process to determine when the addition of another appellate tier would be desirable. The essay then provides two *team* models of appeal that provide microfoundations for an analysis of hierarchy. The first of these team models shows the power of litigant selection of cases to appeal in the determination of the structure of the hierarchy when courts simply correct errors. With perfect selection by litigants, the optimal hierarchy in an error-correcting judiciary has exactly three tiers. The second of these team models shows that litigants will appeal only hard cases and that the rate of appeal will be a function of the quality of the court.

Judicial systems are typically organized as hierarchies, with rights of appeal from level to level. Why is this? Because appeals mechanisms are ubiquitous in court systems, this question is fundamental for our understanding the structure and operation of judicial organizations.

We begin this essay with a structural analysis of appellate processes. The analysis characterizes the appellate process in terms of systemic features rather than individual behavior. In particular, we identify five key variables that allow a characterization of the performance of appeals processes. We then turn to consider specific ways of organizing appeals and focus in particular on appeals initiated by the litigants themselves, a method we call “litigant selection.” Analysts of judicial systems, we believe, need to build models of appeals processes where the conduct of the actors is endogenous; these models typically will be game theoretic. Given

a game-theoretic model of judicial and litigant conduct, we can use the structural apparatus in the first part of the essay to characterize the performance of the system. In this way, we explore how institutional design structures individual conduct, which in turn determines systematic performance. We next illustrate this style of analysis by presenting a simple game theoretic model of appeals and relate this model to the structural systemic analysis presented earlier.

How does this essay relate to others that study appeals processes? Broadly speaking, the theoretical study of judicial appeals remains in its infancy. Steven Shavell (1995) outlined the first and only structural analysis of appeals and hierarchy. In his analysis, Shavell assumes that error rates at both the trial and appellate levels are functions of litigant and state resources invested in the case. The conditions he places on these error rates invoke, but are not derived from, individually rational behavior. He argues that the costs of hierarchy are justified if litigants select the “right” cases—that is, the wrongly decided ones—to appeal. He assumes a system of fees and penalties will assure this outcome.<sup>1</sup> Andrew Daughety and Jennifer Reinganum (2000), in contrast, use a game-theoretic model to examine the informational properties of litigant selection of appeals. In this interesting model, the defendant and the appellate court both receive signals about the legal preferences of a superior court; the appellate court can make deductions about the defendant’s private information from the appeal. A handful of theoretical papers examine other judicial appeals mechanisms.<sup>2</sup>

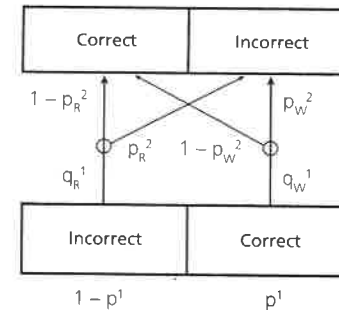
### The Structure of an Appellate Process

Figure 1 displays a simple appellate system. In the figure, cases enter the system and are probabilistically sorted by a trial judge, either correctly or incorrectly (this is the bottom box in the figure). These cases—correctly or incorrectly adjudicated—may be appealed, at the indicated rates (the arrows from the lower box to the shaded circle). In turn, appealed cases may be correctly or incorrectly sorted at the appellate stage (as indicated by the various arrows and corresponding probabilities). This yields a sorting of cases after the appeals process, shown in the upper box in the figure.

#### SOME DEFINITIONS

We can characterize an appeals process from level  $t$  to level  $t + 1$  by five parameters: (1)  $p^t$ , the error rate at  $t$  (with  $1 - p^t$  as the nonerror rate at  $t$ ); (2)  $q'_R$ , the probability that a case rightly decided at level  $t$  will be appealed

FIGURE 1 Structural parameters in an appeals system



to level  $t + 1$ ; (3)  $q'_W$ , the probability that a case wrongly decided at level  $t$  will be appealed to level  $t + 1$ ; (4)  $1 - p_R^{t+1}$ , the probability that, on appeal, a correctly decided case will be upheld at  $t + 1$ ; and, (5)  $1 - p_W^{t+1}$ , the probability that, on appeal, a wrongly decided case will be reversed at  $t + 1$ . We will refer to these probabilities by name in the following ways. We shall call  $p_R^{t+1}$  the *conditional error introduction rate at  $t + 1$*  because it reflects the conditional probability that the appellate court will introduce an error. And we call  $(1 - p_W^{t+1})$  the *conditional error correction rate at  $t + 1$* .<sup>3</sup>

In addition, we call  $q'_W/q'_R$  the *selection ratio at  $t$* ; it is the ratio of the probability a wrongly decided case is appealed to the probability a correctly decided case is appealed. An appeals process is *selective*, *selection-neutral*, or *anti-selective* as this ratio is greater than, equal to, or less than 1 respectively.

We call  $(1 - p_W^{t+1}/1 - p^t)$  the *error correction ratio at  $t + 1$* ; it is the ratio of the conditional error correction rate at  $t + 1$  to the nonerror rate at  $t$ . An appeals process is *error correcting*, *error-correction neutral*, or *anti-error correcting* as this ratio is greater than, equal to, or less than 1 respectively.

We call  $(p_R^{t+1}/p^t)$  the *error introduction ratio at  $t + 1$* ; it is the ratio of the conditional error introduction rate at  $t + 1$  to the error rate at  $t$ . An appeals process is *error introducing*, *error introduction neutral*, or *anti-error introducing* as this ratio is greater than, equal to, or less than 1 respectively.<sup>4</sup>

#### A FUNDAMENTAL THEOREM

In appendix A to this essay, we use these definitions to prove a theorem about the desirability of an add-on appeals process. Our metric for evaluating a judicial hierarchy is the total number of errors remaining after the operation of the hierarchy.

*Theorem 1:* An add-on hierarchy is superior to the underlying court structure if and only if (1) the add-on hierarchy is sufficiently selective at  $T$ , or (2) the add-on hierarchy is sufficiently error correcting at  $T + 1$ , or (3) the addition of the  $T + 1$  level sufficiently improves the performance in the underlying court structure.

Two simple corollaries follow immediately from the theorem (again, proofs are in appendix A).

*Corollary 1.1:* An add-on hierarchy that is error-introduction neutral and error-correction neutral is superior to the underlying court structure if and only if the added appeals process is sufficiently selective.

*Corollary 1.2:* Consider an add-on appeals process that is selection neutral. If the new tier does not reduce errors in the tiers below, the add-on hierarchy is superior to the underlying court structure if and only if the error-correction ratio exceeds the error-introduction ratio for the add-on appeals process.

The corollaries indicate the following. If the new tier improves performance below, selectivity in the new tier is not necessary to guarantee the superiority of the new hierarchy. Otherwise, it is a necessary condition, at least when the appeals process is error-introduction and error-correction neutral. However, selectivity in the new tier is not sufficient to guarantee the superiority of the new hierarchy, even if performance below degrades. Rather, when performance below degrades, the additional appeals process must be *sufficiently* selective to overcome the degradation in the quality of adjudication at the lower tiers.

Corollaries 1.1 and 1.2 make clear that the desirability of hierarchy does not require, as Shavell (1995, 394) assumes, that the error-introduction ratio be less than the error-correction ratio. Examination of the critical equation (H1), in appendix A, shows that an add-on hierarchy may be superior to the underlying hierarchy even when the error-introduction rate exceeds the error correction rate. This can occur, for example, if the appeals rate from  $T$  to  $T + 1$  is sufficiently low for cases correctly decided at  $T$ .<sup>5</sup> Of course, this appeals rate might not be sufficiently low when economically rational litigants face a high error-introduction rate. It can also occur if the addition of the new tier sufficiently improves performance in the tiers below, for example, by improving litigant selection. But why should a new tier affect behavior below? We return to this rather subtle point later.

## Motives, Information, and Mechanisms

The preceding structural analysis of judicial hierarchy, which identified five key variables in an appeals process, treated these variables parametrically. In actual appellate systems, the key probabilities are endogenous. Selectivity and other rates reflect the choices of litigants and judges operating within a specific institutional setting, a setting that structures the actors' incentives and brings to bear specific technologies of information revelation (e.g., the rules of evidence and judicial inference). Thus, working models of judicial hierarchies need to show how different institutional structures and procedural rules lead to different conduct by the actors and thus different performance of the system, in terms of the five key variables. Working models of this kind will typically be game theoretic.

This section acts as a bridge to game-theoretic models of appeals mechanisms, by discussing two of the more problematic elements in models of appeals: assumptions about the motives of judges and assumptions about information.

### THE SOURCES OF JUDICIAL ERROR

The ability of different mechanisms to detect and correct errors depends on the underlying cause of errors.<sup>6</sup> Two different approaches—the “principal-agent” (or just “agency”) approach and the “team” approach—make very different assumptions about the sources of judicial errors and consequently lead to different models of appeals mechanisms.

Principal-agent models of adjudication focus on conflicting interests among judges to explain the existence of hierarchy. Each judge has a preference relation among policies (or, more simply, over case outcomes) and seeks to implement her preferences through her decisions. Hence, “errors” reflect deliberate malfeasance or rebellion by judges. From this perspective, the real source of error in judicial systems is a dearth of reliable judges.

Because the agency approach views judges as political actors, each striving to promote her policy interests, the design of the system will focus on controlling those conflicts rather than efforts to achieve correct answers. Hence, the agency approach to hierarchy directs attention to appeals mechanisms incorporating strategic auditing by superiors, structured competition among lower-court judges to induce compliance (tournaments), “whistle-blowing” by dissenters on collegial courts, and

litigant decisions to appeal, the latter a device to reveal litigants' knowledge of malfeasant actions. It is worth noting that some features of the U.S. judicial hierarchy seem difficult to reconcile with an agency approach. An example is the specialization of trial courts in fact-finding. The underlying assumptions of the agency approach would seem to suggest a system in which appellate courts have the authority to redetermine the facts of a case. Otherwise, a trial judge can partly achieve her desired outcomes through her assessment of the facts.

The team approach presents a very different view of adjudication. In this approach, members of the judiciary are treated as if they are a member of team, in the sense of Ray Radner: they all share a common goal and an identical utility function (Marshak and Radner 1972). The most obvious goal is deciding cases "correctly." This aspiration makes sense in a context in which all judges *agree* on what constitutes correct answers, since they share the same utility function. One might imagine other goals as well, for example, minimize the number of cases brought before the judiciary, or maximize certainty in the law. But "maximize the number of correctly decided cases" seems particularly sensible in the team context.

In the team approach, the principal source of judicial error is hidden information. If judges knew *all* relevant information about the cases, there would be no errors. However, some knowledge is very costly to acquire or verify, so error is inevitable in a world with resource constraints (and, possibly, bounded rationality and variable degrees of judicial skill). In terms of appeals the team approach stresses litigant selection of appeals, as a device to reveal information hidden to judges but known by litigants.

Which approach—agency or team—is the "right" or "best" way to model adjudication? The answer, of course, depends on the modeler's interests. If the modeler is interested in value conflicts and their consequences, then the agency approach is obviously more appropriate. But there is more to the story of adjudication than value conflict—in fact, from a purely positive perspective, most litigation most of the time involves little conflict over fundamental values. To the extent that the modeler is interested in the political economy of "normal" cases rather than extraordinary ones, the team approach is very appealing. Finally, the team approach offers an attractive avenue for creating real (as opposed to straw man) versions of what political scientists like to call "the legal model." Having both agency (political) and team (legal) models of the same phenomenon may be extremely useful for structuring empirical work and, ultimately, understanding why judicial systems are organized the way they are and operate the way they do.

#### INFORMATION STRUCTURES IN APPELLATE PROCESSES

In game-theoretic models of appeals, what information structures are worth considering? The answer depends on one's view of litigation. Two stand out: the "implementation" view of adjudication and the "law-creation" view.

In the "implementation" account of adjudication, an initial trial is about determining facts asymmetrically known by the litigants.<sup>7</sup> Then, given the relevant and admissible facts that emerge, the trial court applies a definite legal rule to determine judgment (Kornhauser 1992). Within this framework, appeals involve correcting lower courts' errors in implementing legal rules. Within a team model, these errors might occur because relevant information isn't revealed at trial, information that might emerge at appeal, thus allowing error correction, or, errors might occur because the trial judge simply makes a mistake in applying law to facts, a mistake that might be corrected through more careful review at the appellate stage. Within the U.S. system of appeals, in which the fact-finding role of trial courts is privileged and little if any additional factual material emerges on appeal, only the second possibility would seem to deserve serious consideration.

In contrast, in the "law-creation" account of adjudication, the key issue is really the absence of a definitive legal rule—there is a "gap" in the law. Of course, an initial trial may *still* be about determining facts. However, the thorny problem for the trial judge, even after the facts emerge, is to make a judgment in the absence of a definitive rule: she must "guess" the "best" rule and apply it. (Note that in a team model, every judge in the same position with the same information would proceed similarly, and the "best" rule—once discovered—will be compelling for all.) Within this account, appeals allow additional or more-intense deliberation about the "best" rule, correcting faulty inferences about the best rule.

This view of litigation directs attention to a handful of information structures, among the multitude of possible ones. In particular, within the team framework:

##### 1. Implementation of Existing Rules

A. Losing litigant has unrevealed private information about the trial judge's fact-law match (litigant knows the judge made a mistake), so that the act of appeal may signal judicial error. The appellate judge (1) may or (2) may not receive additional information about the correct fact-law match during the appeal hearing itself.

B. The losing litigant does not have private information about the trial judge's fact-law match (litigant has no special knowledge of judicial error). In this case, the act of appeal is simply a lottery rather than a signal. As before, additional information about the fact-law match (1) can or (2) cannot be revealed to the appellate judge during the appeal hearing itself.

2. Creation of New Rules

A. Losing litigant has unrevealed private information about the "best rule" for filling the gap, so that the act of appeal is a signal about this information. Then, additional information about the best rule (1) can or (2) cannot be revealed to the appellate judge during the hearing itself.

B. Losing litigant does not have unrevealed private information about the "best rule" for filling the gap, so that the act of appeal is the act of appeal is simply a lottery rather than a signal. Then, additional information about the best rule (1) can or (2) cannot be revealed to the appellate judge during the hearing itself.

As an example, the model in Daugherty and Reinganum 2000 analyzes the information structure in 2.A.(1). In their model, the losing litigant receives a noisy signal about the likely rule ultimately to be chosen by the highest court (the "best rule"), but so does the judge during the appeals process. The task for the appellate judge is to best use both signals efficiently.

In our view, an important research task for analysts of judicial systems is to explore each of these information structures and show how different institutional designs lead to different conduct by the actors and different performance by the system, especially in terms of the five variables identified earlier.

Litigant Selection and Appeals

We now formulate a game in which selection, error-correction, and error-introduction ratios are determined endogenously. We focus on the team context, to create a legal model of hierarchy. And, we consider legal implementation, focusing on information structure 1.A (1). We provide a particularly simple model, not as a final word on the subject, but to illustrate some of the relevant ideas.

PRELIMINARIES

There are two classes of litigants—plaintiffs and defendants—and, depending on the game, one, two, or three tiers of judges. Defendants have a type  $\beta \in \{l, \bar{l}\}$ , (liable and not liable, respectively). Nature selects Defendant's type as  $l$  with common knowledge probability  $p_0$ . Plaintiff and defendant each have two actions open to them in the event the judicial system has more than one tier. Suppose the system has  $T$  tiers. If a judgment at tier  $t < T$  is adverse to its interest, losing litigant  $j$  at level  $t$  may either appeal ( $s_j^t = 1$ ) or not appeal ( $s_j^t = 0$ ). (A judgment at tier  $T$  cannot be appealed.) Let  $\sigma_j^t$  denote the probability of an appeal by losing litigant  $j$  at tier  $t$ . A judge  $i$  at tier  $t$  reaches a judgment  $v_i^t \in \{l, \bar{l}\}$  (defendant held liable or not liable, respectively). Let  $p_i^t$  denote the probability that judge  $i$  at tier  $t$  reaches judgment  $v_i^t = l$ . And, let  $v_F$  denote the final judgment prevailing in the judicial system; that is,  $v_F$  is the decision of the judge at the highest tier in the system to hear the case.

In this team model, all judges wish to maximize the expected number of rightly decided cases in the system. In addition, lower court judges wish to avoid reversal; they suffer a loss  $\epsilon$  if a case is reversed. The utility of judge  $i$  at level  $t$  is then given by:

$$u_i^t = \begin{cases} 1 - J_i \epsilon & \text{if } v_F = \beta \\ 0 - J_i \epsilon & \text{if } v_F \neq \beta \end{cases}$$

where  $J_i$  is an indicator variable taking the value 1 if  $i$ 's judgment is ultimately reversed and 0 otherwise.

The defendant pays damages  $d$  in the event that  $v_F = l$  (that is, she is held liable in the end); otherwise she pays 0. In addition, a litigant incurs a cost  $c$  each time she appeals.

The defendant's utility is then given by

$$u_D = \begin{cases} -d - I_D c & \text{if } v_F = l \\ 0 - I_D c & \text{if } v_F = \bar{l} \end{cases}$$

where  $I_D$  equals one plus the number of appeals the Defendant makes.<sup>8</sup>

The plaintiff suffers a loss  $\lambda$ ; but this occurs regardless of the play of the game and so can be normalized to zero. Should the plaintiff prevail in litigation he will receive damages  $d$  from the defendant in the event that  $v_F = l$ ; otherwise he receives 0. The plaintiff's utility is then given by

$$u_p = \begin{cases} d - I_p c & \text{if } v_p = l \\ 0 - I_p c & \text{if } v_p = \bar{l} \end{cases}$$

where  $I_p$  equals the number of appeals the plaintiff makes.

The information structure is the following. Initially, the defendant's type  $\beta$  is private information; hence, it may be rational for the defendant and plaintiff to engage in litigation (however, we do not actually model the pretrial settlement process). At the end of the trial, the defendant's type is revealed to *both* litigants. This assumption reflects the operation of discovery and the litigants' special knowledge of their circumstances. However, the defendant's type is revealed to judge  $i$  at tier  $t$  only with probability  $\pi_i^t(c/n^t)$ , that is, with a probability that is a function of the judge's caseload.<sup>9</sup> More specifically, as a result of the trial and the judge's deliberation concerning matters of fact and law, the judge at tier  $t$  receives a "hard signal"  $x^t \in \{0, \beta\}$ , where 0 denotes a noninformative hard signal.<sup>10</sup> We assume that if a defendant's type is ever revealed by a "hard signal"—verifiable and legally admissible evidence combined with judicial reasoning that team members see as determinative—then all subsequent judges also know the defendant's type. We assume  $\pi^{t'} < 0$ . Let  $\mu_i^t$  be the belief of judge  $i$  at tier  $t$  that defendant is liable ( $\beta = l$ ).

FLAT ORGANIZATION

As a benchmark, we first consider a flat organization. The "equilibrium" is straightforward, because, in effect, there is only one player, the judges at the first tier, each of whom acts unilaterally. If the judge learns the defendant's type from the hard signal, it rules accordingly. Otherwise, it relies on its prior beliefs. A strategy  $\rho_i^1(x^1, p_0)$  for judge  $i$  is a probability of ruling "l", given the hard signal and the initial probability the defendant is liable.

*Proposition 1:* In a one-tier hierarchy, the following characterizes judge  $i$ 's strategy:

$$\rho_i^1(x^1, p_0) = \begin{cases} 1 & \text{if } \mu_i^1 > \frac{1}{2} \\ \alpha \in [0, 1] & \text{if } \mu_i^1 = \frac{1}{2} \\ 0 & \text{if } \mu_i^1 < \frac{1}{2} \end{cases}$$

*Proof:* The expected value of  $v^1 = l$  is  $\mu_i^1(1) + (1 - \mu_i^1)(0)$ , while the expected value of  $v^1 = \bar{l}$  is  $\mu_i^1(0) + (1 - \mu_i^1)(1)$ . The former will be

greater than the latter as  $\mu_i^1 \geq (1/2)$ . Note that if the hard signal reveals the defendant's type, then  $\mu_i^1(x^1, p_0) = 1$  or  $\mu_i^1(x^1, p_0) = 0$  as  $x_1 = l$  or  $x_1 = \bar{l}$ , respectively. Absent an informative hard signal,  $\mu_i^1(x^1, p_0) = p_0$ .

In expectation, the proportion of correctly decided cases is  $\Pi(c^1/n^1) = \pi^1(c^1/n^1) + (1 - \pi^1(c^1/n^1))\max\{p_0, 1 - p_0\}$ . This proportion is higher as  $\pi^1$  increases, so that informative hard signals are more likely. It also increases as  $p_0$  tends toward 0 or 1, so that relying on prior beliefs tends to sort the litigants properly.

TWO-TIER HIERARCHY

We now consider a two-tier hierarchy. In this game, a losing litigant may appeal to a higher court, which is obliged to hear the appeal. This is a signaling game, since the appeal may reflect the hidden information of the appellant. There are three types of equilibria to consider: a separating equilibrium; a hybrid, or partial pooling equilibrium; and a pooling equilibrium.

**SEPARATING EQUILIBRIUM.** In this equilibrium, the litigants separate in appeals: if the defendant's type was not revealed at trial, the losing litigant does not appeal in correctly decided cases ( $v^1 = \beta$ ), thus ending the game; but the losing litigant does appeal (so  $s_j^1 = 1$ ) in incorrectly decided ones ( $v^1 \neq \beta$ ). Given this behavior by the litigants, the appellate court always reverses the trial court (in the absence of a hard signal that the losing litigant has improperly appealed—an out-of-equilibrium event). But a hard signal upon appeal must be rather likely, for it is this possibility that drives the correctly losing litigants to separate from incorrectly losing ones.

Conceptually, an equilibrium of this game requires the specification of strategies for seven players: the trial judge, the appellate judge, a losing defendant of type  $l$ , a losing defendant of type  $\bar{l}$ , a losing plaintiff who knows the defendant's type is  $l$ , and a losing plaintiff who knows the defendant's type is  $\bar{l}$ . However, in the separating equilibrium, we may describe the strategy of all losing litigants identically, conditioning only on whether the trial judgment was correct or incorrect, given the defendant's type. Consequently, as the following proposition shows, the separating equilibrium is quite simple.

*Proposition 2a:* In a two-tier hierarchy, if  $\pi^2(c^2/n^2) \geq 1 - (c/d)$ , then the following is an equilibrium: judge  $i$  at tier 1 adopts the strategy

$$\rho_j^1(x^1, p_0) = \begin{cases} 1 & \text{if } \mu_i^1 > \frac{1}{2} \\ \alpha \in [0,1] & \text{if } \mu_i^1 = \frac{1}{2} \\ 0 & \text{if } \mu_i^1 < \frac{1}{2} \end{cases}$$

A losing litigant  $j$  at tier 1 adopts the strategy

$$\sigma_j^1(v^1, x^1, \beta) = \begin{cases} 1 & \text{if } v^1 \neq \beta \\ 0 & \text{if } v^1 = \beta \end{cases}$$

and an appellate judge  $i$  at tier 2 adopts the strategy

$$\rho_i^2(s_j^1, x^2, x^1, p_0) = \begin{cases} 1 & \text{if } \mu_i^2(s_j^1, x^2, x^1, p_0) \geq 1/2 \\ 0 & \text{otherwise} \end{cases}$$

and  $\mu_i^2(s_j^1, x^2, x^1, p_0)$  is determined by Bayes's Rule whenever possible. If a hard signal ever reveals  $\beta$ , the appellate judge believes the hard signal regardless of an appeal.

*Proof:* See appendix B.

A striking feature of this equilibrium is that every case is adjudicated correctly: the expected number of errors is zero! But, the caseload for the appellate judge is just its share of the cases incorrectly decided at the first level, which on average is  $(1 - \pi^1(c/n^1))\min\{p_0, 1 - p_0\}C^1$ . If the trial judges are overburdened, their accuracy ( $\pi^1$ ) will be low. Then, if prior beliefs are not very useful in sorting litigants correctly at trial, the appellate caseload may be large. And in turn, if the appellate caseload is heavy, the critical condition on appeals accuracy ( $\pi^2 \geq 1 - c/d$ ) may fail and the separating equilibrium will not be possible.

**SEMISEPARATING (PARTIAL POOLING) EQUILIBRIUM.** Suppose the probability of a hard signal on appeal isn't high enough to ensure separation of types. Then the following is an equilibrium: the trial judge acts as above; absent a hard signal at trial, a losing litigant definitely appeals an incorrect judgment and may appeal a correct one. Thus, no appeal is a sign of guilt by the losing litigant (but a payoff irrelevant one because this ends the game). An appeal, however, has an ambiguous meaning. Absent a hard signal on appeal, the appellate court reverses some judgments but affirms others. This equilibrium involves rather delicate mixing.

*Proposition 2b:* If  $\pi^2 < 1 - (c/d)$ , the following is an equilibrium: The trial judge plays the strategy

$$\rho_j^1(x^1, p_0) = \begin{cases} 1 & \text{if } x^1 = l, \text{ or if } x^1 = 0 \text{ and } p_0 > \frac{1}{2} \\ \alpha \in [0,1] & \text{if } x^1 = 0 \text{ and } p_0 = \frac{1}{2} \\ 0 & \text{if } x^1 = \bar{l}, \text{ or if } x^1 = 0 \text{ and } p_0 < \frac{1}{2} \end{cases}$$

An incorrectly losing plaintiff plays the strategy

$$\sigma_j^1(v^1 = \bar{l}, x^1, \beta = l, \rho_i^1(\cdot)) = \begin{cases} 1 & \text{if } x^1 = \beta = l, \text{ or if } x^1 = 0, \beta = l \\ & \text{and } \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) \geq \frac{c - d\pi^2}{d(1 - \pi^2)} \\ 0 & \text{if } x^1 = 0, \beta = l \\ & \text{and } \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) < \frac{c - d\pi^2}{d(1 - \pi^2)} \end{cases}$$

A correctly losing plaintiff plays the strategy

$$\sigma_p^1(v^1 = \bar{l}, x^1, \beta = \bar{l}, \rho_i^1(\cdot)) = \begin{cases} 1 & \text{if } x^1 = 0, \beta = \bar{l} \\ & \text{and } \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) > \frac{c}{d(1 - \pi^2)} \\ \frac{p_0}{1 - p_0} & \text{if } x^1 = 0, \beta = \bar{l}, \\ & \text{and } \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) = \frac{c}{d(1 - \pi^2)} \\ 0 & \text{if } x^1 = \beta = \bar{l}, \text{ or if } x^1 = 0, \beta = \bar{l} \\ & \text{and } \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) < \frac{c}{d(1 - \pi^2)} \end{cases}$$

An incorrectly losing defendant plays the strategy

$$\sigma_{D=i}^1(v^1 = l, x^1, \rho_i^1(\cdot)) = \begin{cases} 1 & \text{if } x^1 = \bar{l}, \text{ or if } x^1 = 0 \\ & \text{and } \rho_i^2(v^1 = l, x^1 = x^2 = 0) \leq \frac{d - c}{d(1 - \pi^2)} \\ 0 & \text{if } x^1 = 0 \\ & \text{and } \rho_i^2(v^1 = l, x^1 = x^2 = 0) > \frac{d - c}{d(1 - \pi^2)} \end{cases}$$



A correctly losing defendant plays the strategy

$$\sigma_{i \neq j}^1(v^i = l, x^1, \rho_i^1(\cdot)) = \begin{cases} 1 \text{ if } x^1 = 0 \\ \text{and } \rho_i^2(v^i = l, x^1 = x^2 = 0) < 1 - \frac{c}{d(1 - \pi^2)} \\ \frac{1 - p_0}{p_0} \text{ if } x^1 = 0 \\ \text{and } \rho_i^2(v^i = l, x^1 = x^2 = 0) = 1 - \frac{c}{d(1 - \pi^2)} \\ 0 \text{ if } x^1 = l \text{ or if } x^1 = 0 \\ \text{and } \rho_i^2(v^i = l, x^1 = x^2 = 0) > 1 - \frac{c}{d(1 - \pi^2)} \end{cases}$$

Finally, the appellate judge plays the strategy

$$\rho_i^2(v^i = \bar{l}) = \begin{cases} 1 \text{ if } \mu_i^2 = 1, \text{ or if } x^1 = x^2 = 0 \\ \text{and } \sigma_{i \neq j}^1(x^1 = 0, \beta = \bar{l}) > \frac{p_0}{1 - p_0} \\ \frac{c}{d(1 - \pi^2)} \text{ if } x^1 = x^2 = 0 \\ \text{and } \sigma_{i \neq j}^1(x^1 = 0, \beta = \bar{l}) = \frac{p_0}{1 - p_0} \\ 0 \text{ if } \mu_i^2 = 0, \text{ or if } x^1 = x^2 = 0 \\ \text{and } \sigma_{i \neq j}^1(x^1 = 0, \beta = \bar{l}) < \frac{p_0}{1 - p_0} \end{cases}$$

$$\rho_i^2(v^i = l) = \begin{cases} 1 \text{ if } \mu_i^2 = 1, \text{ or if } x^1 = x^2 = 0 \\ \text{and } \sigma_{i \neq j}^1(x^1 = 0, \beta = l) > \frac{1 - p_0}{p_0} \\ 1 - \frac{c}{d(1 - \pi^2)} \text{ if } x^1 = x^2 = 0 \\ \text{and } \sigma_{i \neq j}^1(x^1 = 0, \beta = l) = \frac{1 - p_0}{p_0} \\ 0 \text{ if } \mu_i^2 = 0, \text{ or if } x^1 = x^2 = 0 \\ \text{and } \sigma_{i \neq j}^1(x^1 = 0, \beta = l) < \frac{1 - p_0}{p_0} \end{cases}$$

Beliefs are determined by Bayes's Rule whenever possible. If a hard signal ever reveals  $\beta$ , the appellate judge believes the hard signal regardless of an appeal.

*Proof:* See appendix B.

These strategies may look forbiddingly complex, but the play of the game is actually quite simple. If a hard signal occurs at trial, the trial judge rules accordingly, and there is no appeal. If no hard signal emerges at trial, what happens depends on the initial probability that the defendant is liable.

If  $p_0 > .5$  and the trial judge's signal is uninformative, the trial judge holds the defendant liable. An incorrectly losing defendant appeals with certainty. A correctly losing Defendant appeals with probability  $(1 - p_0/p_0)$  (which lies between 0 and 1). Given an appeal by a defendant but no hard signal on appeal, the appellate judge sustains with probability  $1 - (c/d(1 - \pi^2))$  and reverses with the reciprocal probability. An incorrectly losing plaintiff appeals with certainty; and the appellate court reverses. (When the trial signal is uninformative, there are no correctly losing plaintiffs given the strategy of the trial judge.)

If  $p_0 < .5$ , absent a hard signal at trial, the trial judge holds the Defendant not liable. An incorrectly losing plaintiff appeals with certainty. A correctly losing plaintiff appeals with probability  $(p_0/1 - p_0)$  (which lies between 0 and 1). Given an appeal by a plaintiff but no hard signal on appeal, the appellate judge again sustains with probability  $1 - (c/d(1 - \pi^2))$  and reverses with the reciprocal probability. An incorrectly losing defendant appeals with certainty and the appellate judge reverses. (When the trial signal is uninformative, there are no correctly losing defendants given the strategy of the trial judge.)

**POOLING EQUILIBRIUM.** There are two possible pooling equilibria: an "everyone appeals" equilibrium, and a "no-one appeals" equilibrium.

A moment's reflection shows that the former cannot be an equilibrium. Because litigants pool, the appellate judge's beliefs must be the same as those of the trial judge in cases lacking a hard signal at either level. Consequently, in those cases, the appellate judge will uphold the trial judge's verdict. But this means that an appeal by a losing litigant in a correctly decided case can never be profitable (since either a hard signal is received, resulting in an affirmance, or one is not received, again resulting in affirmance).

A “no-appeals” equilibrium can hold if the probability of a hard signal is rather low. But this equilibrium is vulnerable to forward induction-style refinements. In this equilibrium, no litigants appeal. But if one did, the appellate judge (absent a hard signal) must affirm the lower court—in the equilibrium, an appeal (an out-of-equilibrium event) is taken as a signal of guilt. (Otherwise, losing litigants will appeal, breaking the equilibrium). Given this interpretation of an appeal, no one appeals. Although this is an equilibrium, an obvious issue is, is it reasonable for the appellate judge to believe that an appeal signals guilt rather than innocence? Note that for correctly losing litigants, a no-appeal equilibrium dominates no appeal (since if there is a hard signal, his guilt emerges, and if not, the judge believes he is guilty and affirms anyway). By construction, no appeal must also be more appealing than appeal for an incorrectly losing litigant. But it must be the case that, for any response by the appellate judge to an appeal, appeal is more attractive for an incorrectly losing litigant than for a correctly losing litigant (because of the possibility of a hard signal, which would vindicate the incorrectly losing litigant). Hence, if the appellate judge sees an appeal, she should put no weight on an appeal from an incorrectly losing litigant.<sup>11</sup> But given this reasoning, an appeal could only come from an incorrectly losing litigant, which should compel the judge to reverse, even absent a hard signal. Hence, this equilibrium appears implausible.

The reasoning establishes:

*Proposition 2c:* No universally divine pooling equilibrium exists in the two-tier game.

## Discussion

The two-tier separating equilibrium performs well, but it is fragile. It can exist only if judicial accuracy in the upper tier is quite high. But if judicial accuracy in the lower tier is poor, then the upper tier’s caseload can be very high, and its accuracy correspondingly low. In that case, the partial pooling equilibrium—which is far less desirable—looks more probable. The equilibrium in the following section seems to point to a easy structural “fix” for the two-tier system’s vulnerability.

We now consider a three-tier hierarchy and focus on a truly remarkable separating equilibrium. In this equilibrium, litigants separate at the first level of appeals, so that only incorrectly decided cases appeal. In the intermediate court, all appealed initial judgments are reversed (in equi-

librium). Then, losing litigants make no appeals to the “Supreme Court.” Hence, all cases are resolved correctly! Unlike the separating equilibrium in the two-layer hierarchy, separation does not depend on high levels of accuracy at the second level. Rather, the key in constructing the equilibrium is a high level of accuracy at the *highest* level—even though this court hears no cases. If a correctly losing litigant makes a “bogus” appeal to the intermediate court and the case is reversed (absent a hard signal), then the litigant who initially (correctly) won has an incentive to appeal to the Supreme Court, which is very likely to receive a hard signal and decide the case correctly. Conversely, in a case that is initially incorrectly decided but corrected upon appeal, the correctly losing litigant will have no incentive to appeal to the high court, if the accuracy level at that level is sufficiently high. This high level of accuracy is quite reasonable, given the Supreme Court’s low caseload (actually, zero in equilibrium).

*Proposition 3:* If  $\pi^3 > 1 - (c/d)$ , the following is an equilibrium in the three-tier hierarchy game:

$$\rho_i^3(p_0, x^1, v^1, s^1, x^2, v^2, s^2, x^3) = \begin{cases} 1 & \text{if } \mu_i^3(\cdot) \geq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

$$\sigma_j^2(p_0, x^1, v^1, s^1, x^2, v^2) = \begin{cases} 1 & \text{if } v^2 \neq \beta \\ 0 & \text{otherwise} \end{cases}$$

$$\rho_i^2(p_0, x^1, v^1, s^1, x^2) = \begin{cases} 1 & \text{if } \mu_i^2(\cdot) \geq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

$$\sigma_j^1(p_0, x^1, v^1) = \begin{cases} 1 & \text{if } v^1 \neq \beta \\ 0 & \text{otherwise} \end{cases}$$

$$\rho_i^1(p_0, x^1) = \begin{cases} 1 & \text{if } \mu_i^1(\cdot) \geq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

Beliefs are determined wherever possible by Bayes’s Rule. If a hard signal ever reveals the defendant’s type, the beliefs of subsequently acting judges are fixed accordingly. Following an appeal of the intermediate court’s judgment, in the absence of any hard signals the high court believes an error occurred at the intermediate court.

*Proof:* See appendix B.

Conceptually, the conditional error introduction rate at the Supreme Court is small enough so that it dissuades bogus appeals. This, in turn, allows the conditional error correction rate at the penultimate stage to be very high (even absent much judicial accuracy), thereby encouraging “legitimate” appeals.

The following Corollary is striking:

*Corollary 3:* There is never a need to have more than three tiers in a judicial hierarchy, regardless of the caseload.

*Proof:* The error rate in the equilibrium detailed in Proposition 3 is zero. Adding additional layers cannot improve upon this performance. And, the equilibrium is robust to the caseload.

This remarkable result obviously reflects the strong assumptions in the model. It is important to identify the critical assumption, which is: *after trial, both litigants know whether the defendant is truly liable*. The hierarchy then exploits their hidden but mutual knowledge, implicitly encouraging them to police the outcome via appeals.

It is also tempting to interpret Propositions 2 and 3 in the following way. If the caseload in a judicial system is sufficiently small, a two-tier hierarchy is “good enough” to exploit the power of litigant selection. But if the caseload becomes too burdensome, a three-tier hierarchy—and no more than three tiers—is needed to gain the full benefit of litigant selection. In fact, the historical development of state judicial systems in the United States follows this stylized script (see Kagan et al. 1978). And broadly speaking, so does the history of the federal judiciary (Frankfurter and Landis 1928). Of course, functionalist “just-so” stories can never adequately account for actual historical developments. But the models may suggest some fruitful directions for historical research on judicial hierarchies.

One implication of Proposition 3, however, appears quite counterfactual. Proposition 3 seems to imply that few judges are needed at the trial level (since judicial accuracy there is immaterial), and few judges are needed at tier two (because accuracy there is also, in equilibrium, immaterial). But, on the assumption that a larger panel implies a more accurate judgment on the law, several judges are probably needed at the highest level—even though these judges hear no cases.

Reflection suggests, however, that many judges are surely required, at least at the trial level. Recall the key assumption in the model: both litigants learn all private information during trial, even if the judges do not. For trials to proceed in a sufficiently informative way to assure mutual

revelation, surely considerable judicial resources are required.<sup>12</sup> (In addition, of course, trial courts make judgments of law, an activity not addressed by the model in this essay). Thus, tier one may require many judges. A relatively sparse layer of intermediate appellate judges is then necessary to process the large number of appeals from an inaccurate trial court; each intermediate appellate judge need not have great accuracy. Accuracy at the top, however, is essential, so we should expect a reasonably sized panel of judges prepared to act en banc on any given appeal. The current structure of trial courts in the United States arguably fits that description. There are many trial judges, but they do not sit in panels; they serve only to guide fact-finding. Supreme courts in the United States generally have large panels of five to nine judges. There are many intermediate appellate courts that usually sit in panels of three.

### The Structural Consequences of Litigant Selection

Our simple models of litigant selection illustrate the complexity of the structural relations within a judicial hierarchy. Here we briefly illustrate the use of the theorems in the first part of the essay—the structural analysis—by applying them to the game theoretic models in the second part of the essay. We focus on the move from a flat organization to a two-tiered hierarchy, first in the separating equilibrium and then in the partial pooling equilibrium. Application of the theorems to the move to a three-tiered hierarchy is similar, and is omitted for brevity. Within our framework, of course, theorem three shows that adding a fourth or higher tier cannot improve the performance of the hierarchy.

#### ADDING A SECOND TIER: SEPARATING EQUILIBRIUM

Our propositions 2a to 2c indicate the possible consequences of moving from a flat organization of a judicial system to a two-tiered hierarchical organization. Consider first the separating equilibrium described in proposition 2a. When the second tier is sufficiently accurate, the two-tiered hierarchy correctly decides every case. This occurs because the appeals process is highly—indeed, infinitely—selective as  $q_R^{1|2}$ , the probability that a case rightly decided at level 1 will be appealed to level 2, is simply 0, while  $q_W^{1|2}$ , the probability that a wrongly decided case at level 1 will be appealed to level 2, equals 1. The second tier thus introduces no errors and, as every wrongly decided case is appealed, it corrects all errors from below.

This analysis is consistent with an application of our theorem 1 from section 2 concerning add-on hierarchies. Perhaps the simplest way to see

this is to use (H3). In the case of a two-tier hierarchy, the two-tier hierarchy will be superior to the flat organization, if

$$\frac{1 - p_w^{2|2}}{1 - p^{1|2}} > \frac{q_R^{1|2} p_R^{2|2}}{q_w^{1|2} p^{1|2}} - \frac{\Delta^2}{q_w^{1|2} C_w^{1|2} (1 - p^{1|2})} \quad (\text{H3.a})$$

In the separating equilibrium,  $q_R^{1|2} = 0$ ,  $q_w^{1|2} = 1$ , and  $1 - p_w^{2|2} = 1$ , so (with a little algebra) H3.a reduces to  $C_w^{1|2} > -\Delta^2$ . However, using the earlier definitions,  $\Delta^2 = C_w^{1|1} - C_w^{1|2}$ . Substituting and rearranging reduces H3.a to:  $C_w^{1|1} > 0$ . This inequality is met in any reallocation hierarchy that simply reallocates some of the trial judges in the flat organization to an appellate tier. Thus, though the addition of an appellate tier degrades the quality of justice at the trial level, it may improve the performance of the system as a whole.<sup>13</sup>

#### ADDING A SECOND TIER: PARTIAL POOLING EQUILIBRIUM

We now consider the partial pooling equilibrium, which arises when the appellate tier is not sufficiently accurate to enforce the separating equilibrium. We begin with a “direct” comparison of the two organizational forms, by calculating the expected errors in the two-tier hierarchy with a partial pooling equilibrium. In the appendix, we prove

*Proposition 2d:* In the semiseparating equilibrium of the two-tier game, the expected proportion of correctly decided cases of the court system is  $1 - (1 - \pi^{1|2})(1 - \pi^{2|2})(1 - p_0)$ , if  $p_0 > .5$ , and  $1 - (1 - \pi^{1|2})(1 - \pi^{2|2})p_0$  if  $p_0 \leq .5$ .

Proposition 2d states that, in the semiseparating equilibrium, if a hard signal is likely at either the trial or appellate level, correct adjudication will be likely as well. Similarly, if the defendant is probably liable ( $p_0 \approx 1$ ) or probably not liable ( $p_0 \approx 0$ ), correct adjudication is probable. But if a hard signal is unlikely both at trial and on appeal and the Defendant is approximately equally likely to be liable as not liable, the probability of correct adjudication approaches one-half—a coin toss. This is far from the perfect adjudication of the separating equilibrium.

The following corollary follows immediately from proposition 2d and the observation that in the flat organization the total number of errors is  $(1 - \pi^1(c^{1|1}/n^{1|1}))\min\{p_0, 1 - p_0\}$ .

*Corollary:* A two-tiered hierarchy under the partial pooling equilibrium is an improvement over a flat organization if and only if  $(1 - \pi^{1|1}) > (1 - \pi^{1|2})(1 - \pi^{2|2})$ .

This corollary states that in the partial pooling case, the two-tiered hierarchy *reduces error if it reduces the probability that the judicial system receives no hard signal*. We have already noted in our discussion of the separating equilibrium that creating a second tier by reallocating judges from a trial bench to an appellate bench will degrade the quality of the trial courts; that is,  $(1 - \pi^{1|1}) > (1 - \pi^{1|2})$ . The accuracy of the appellate process must thus be sufficiently high to overcome this deficit.

Theorem 1 yields the same conclusion, in a somewhat less direct manner. Perhaps the simplest approach is to use (H0). In this case, (H0) indicates that the two-tier hierarchy will be superior to the flat organization, if

$$C_w^{1|1} > C_w^{1|2}(1 - q_w^{1|2}) + C_w^{2|2}$$

(recalling that  $A^{1|7} = 0$ ). Using the definitions for caseloads of wrong cases, this becomes

$$p^{1|1} C^{1|1} > C_w^{1|2}(1 - q_w^{1|2}) + C_R^{1|2} q_R^{1|2} p_R^{2|2} + C_w^{1|2} q_w^{1|2} p_w^{2|2} \quad (\text{H0.a})$$

To evaluate this expression, we need to specify the values of the key systemic variables that characterize the appellate tier in the partial pooling equilibrium. We employ the “aggregate” approach discussed in note 12, focusing on the case of  $p_0 > 1/2$ .<sup>14</sup> First,  $p^{1|1}$ , the error rate at level 1 in the flat organization, equals  $(1 - \pi^{1|1})(1 - p_0)$ . Second,  $q_R^{1|2}$ , the probability that a case rightly decided at level 1 will be appealed to level 2, is  $(1 - \pi^{1|2})(1 - p_0)/\pi^{1|2} + (1 - \pi^{1|2})p_0$ .<sup>15</sup> Third,  $q_w^{1|2} = (1 - \pi^{1|2})(1 - p_0)/(1 - \pi^{1|2})(1 - p_0) = 1$ , that is, the probability that a wrongly decided case at level 1 will be appealed to level 2 is equal to 1. Fourth,  $1 - p_R^{2|2}$ , the probability that, on appeal, a case correctly decided at level 1 will be upheld, is  $\pi^{2|2} + (1 - \pi^{2|2})(1 - (c/d)(1 - \pi^{2|2}))$ . Fifth,  $1 - p_w^{2|2}$ , the probability that, on appeal, a wrongly decided case will be reversed, equals  $\pi^{2|2} + (1 - \pi^{2|2})(c/d)(1 - \pi^{2|2})$ . Finally, note that  $C_w^{1|2} = (1 - \pi^{1|2})(1 - p_0)C^{1|2}$  and  $C_R^{1|2} = (\pi^{1|2} + (1 - \pi^{1|2})p_0)C^{1|2}$ .

Employing  $p^{1|1}$  and  $q_w^{1|2}$ , (H0.a) reduces to

$$(1 - \pi^{1|1})(1 - p_0)C^{1|1} > C_R^{1|2} q_R^{1|2} p_R^{2|2} + C_w^{1|2} q_w^{1|2} p_w^{2|2}$$

Substitution of the additional key values and some algebra yields

$$(1 - \pi^{1|1})(1 - p_0)C^{1|1} > (1 - p_0)(1 - \pi^{1|2})(1 - \pi^{2|2})C^{1|2},$$

so that if the initial caseload in the two organizational forms is the same, the conclusion in the corollary is immediate. Of course, the number of disputes brought to a court system will often depend on the quality of justice provided by the system.

## Conclusion

In this essay, we first conducted a structural analysis of appeals procedures, the first complete analysis of this kind that we know. We then discussed possible information structures that might arise in judicial hierarchies. Finally, we examined one of these in a rather stylized, game-theoretic setting. The purpose of the game-theoretic analysis was less to propose models of courts that one could take to data, than to conduct a series of thought experiments aimed at sharpening our theoretical intuitions about the logic and incentives in appeals system. The game-theoretic models point to the value of litigant selection, particularly in the situation of judicial implementation. The intuition is simple but illuminating: *litigant selection places the burden of appeals on informationally advantaged actors*. The separating equilibrium in the three-tier hierarchy suggests the remarkable power of this mechanism, even in the face of relatively weak technologies of judging. We suspect litigant selection will be less potent in situations of law creation, because the litigants are not likely to be informationally advantaged relative to judges.<sup>16</sup> Presumably, the absence of information asymmetries precludes a striking structural result, like the three-tier result in the implementation models (Corollary 3). Nonetheless, if the initial trial provides the litigants with good information about the prospects of reversal (that is, the probability a higher court will find a better rule), litigants will still screen out the least promising cases for law creation. In short, the logic of litigant selection seems compelling in systems of appeals.

## Appendix A: Comparing a T-tier hierarchy and a T+1 tier Hierarchy

We employ the following notation:

- $C \equiv C^1$ , the total number of cases to be decided in the system
- $n \equiv$  total number of judges in the system
- $T \equiv$  the total number of tiers (indexed by  $t$ )
- $n^t \equiv$  the total number of judges at tier  $t$
- $C^t \equiv$  the total number of cases at tier  $t$
- $q_w^t \equiv$  the proportion of wrongly decided cases at level  $t$  that are appealed to level  $t + 1$
- $q_R^t \equiv$  the proportion of rightly decided cases at level  $t$  that are appealed to level  $t + 1$

- $p^t \equiv$  the error rate at  $t$  with  $(1 - p^t)$  the nonerror rate at  $t$ .
- $p_R^t \equiv$  probability that a rightly decided case at level  $t - 1$  will be wrongly decided at  $t$ , conditional on its being appealed to level  $t$ .
- $p_w^t \equiv$  probability that a wrongly decided case at level  $t - 1$  will be wrongly decided at  $t$ , conditional on its being appealed to level  $t$ .

We now prove a theorem about the desirability of an *add-on appeals process* (which yields an *add-on hierarchy*). In an add-on hierarchy, we append an appeals process to an existing court structure. Theorem 1 identifies conditions under which an add-on hierarchy is superior to the underlying court structure standing alone.

First, define the *caseload of correctly decided cases* and the *caseload of incorrectly decided cases* at tier  $t$  in a  $T$ -level hierarchy. At tier 1, these are:  $C_R^{1|T} = (1 - p^{1|T})C^{1|T}$  and  $C_W^{1|T} = p^{1|T}C^{1|T}$  (note the superscript indicating the "height" of the hierarchy). At the higher levels,  $t > 1$ , they are:  $C_R^{t|T} = C_R^{t-1|T}q_R^{t-1|T}(1 - p_R^{t|T}) + C_W^{t-1|T}q_W^{t-1|T}(1 - p_W^{t|T})$  and  $C_W^{t|T} = C_R^{t-1|T}q_R^{t-1|T}p_R^{t|T} + C_W^{t-1|T}q_W^{t-1|T}p_W^{t|T}$ , respectively.

Our metric for evaluating a judicial hierarchy is the total number of errors remaining after the operation of the hierarchy. The total number of errors in a  $T$ -tier hierarchy is simply the unappealed errors at each of the lower levels, plus the errors at tier  $T$ . Define the sum of unappealed errors in the levels lower than  $t'$  in a  $T$ -tier hierarchy as  $A^{t'|T} = \sum_{t=1}^{t'-1} C_W^{t|T}(1 - q_W^{t|T})$ . (We define  $A^{1|T} \equiv 0$ ). The total number of errors in a  $T$ -tier hierarchy is then  $E^{T|T} = A^{T|T} + C_W^{T|T}$ . Finally, define  $\Delta^{T+1} \equiv A^{T|T} - A^{T|T+1} + C_W^{T|T} - C_W^{T|T+1}$ . Note that this is the *errors reduced in tiers 1 to  $T$  by adding tier  $T + 1$* . (If the addition of the  $T + 1$  tier increases errors in tiers 1 to  $T$ , then  $\Delta^{T+1}$  is negative.) We can now state Theorem 1:

**Theorem 1:** An add-on hierarchy is superior to the underlying court structure if and only if (1) the add-on hierarchy is sufficiently selective at  $T$ , or (2) the add-on hierarchy is sufficiently error correcting at  $T + 1$ , or (3) the addition of the  $T + 1$  level sufficiently improves the performance in the underlying court structure.

*Proof:* Moving from the  $T$ -level hierarchy to the  $T + 1$  hierarchy reduces errors, if

$$A^{T|T} + C_W^{T|T} > A^{T|T+1} + C_W^{T|T+1}(1 - q_W^{T|T+1}) + C_W^{T+1|T+1} \quad (H0)$$

$$\Leftrightarrow A^{T|T} - A^{T|T+1} + C_W^{T|T} - C_W^{T|T+1} > C_R^{T|T+1}q_R^{T|T+1}p_R^{T+1|T+1} - C_W^{T|T+1}q_W^{T|T+1}(1 - p_W^{T+1|T+1})$$

$$\Leftrightarrow \Delta^{T+1} + C_W^{T|T+1}q_W^{T|T+1}(1 - p_W^{T+1|T+1}) > C_R^{T|T+1}q_R^{T|T+1}p_R^{T+1|T+1} \quad (H1)$$

that is, the errors reduced in tiers 1 to  $T$  by the addition of tier  $T + 1$ , plus the errors corrected directly by tier  $T + 1$ , are greater than the errors introduced directly by tier  $T + 1$ .

Rearranging (H1) yields

$$\frac{q_W^{T|T+1}}{q_R^{T|T+1}} > \frac{p_R^{T+1|T+1}}{C_W^{T|T+1}} \frac{C_R^{T|T+1}}{1 - p_W^{T+1|T+1}} - \frac{\Delta^{T+1}}{q_R^{T|T+1} C_W^{T|T+1} (1 - p_W^{T+1|T+1})}$$

Multiplying the first term on the right-hand side of this expression by  $(C^{T|T+1}/C^{T|T+1})$  yields

$$\frac{q_W^{T|T+1}}{q_R^{T|T+1}} > p_R^{T+1|T+1} \frac{C^{T|T+1}}{C_W^{T|T+1}} \frac{C_R^{T|T+1}}{C^{T|T+1}} \frac{1}{1 - p_W^{T+1|T+1}} - \frac{\Delta^{T+1}}{C_W^{T|T+1} (1 - p_W^{T+1|T+1})}$$

Noting that  $(C^{T|T+1}/C_W^{T|T+1}) = (1/p^{T|T+1})$  and  $(C_R^{T|T+1}/C^{T|T+1}) = 1 - p^{T|T+1}$ , we then have

$$\frac{q_W^{T|T+1}}{q_R^{T|T+1}} > \frac{p_R^{T+1|T+1}}{p^{T|T+1}} \frac{1 - p^{T|T+1}}{1 - p_W^{T+1|T+1}} - \frac{\Delta^{T+1}}{q_R^{T|T+1} C_W^{T|T+1} (1 - p_W^{T+1|T+1})} \tag{H2}$$

(H2) establishes the first part of the theorem because it states that (H1) holds if and only if the selection ratio at level  $T$  is greater than the error introduction ratio at  $T + 1$  times the inverse error correction ratio at  $T + 1$ , minus a weighted version of the errors reduced in tiers 1 to  $T$  by adding tier  $T + 1$ .

To establish part (2) we simply rewrite (H2) as

$$\frac{1 - p_W^{T+1|T+1}}{1 - p^{T|T+1}} > \frac{q_R^{T|T+1}}{q_W^{T|T+1}} \frac{p_R^{T+1|T+1}}{p^{T|T+1}} - \frac{\Delta^{T+1}}{q_W^{T|T+1} C_W^{T|T+1} (1 - p^{T|T+1})} \tag{H3}$$

(H3) states that the add-on hierarchy is superior if the error-correction ratio at  $T + 1$  is greater than the inverse selection ratio at  $T$  times the error-introduction ratio at  $T + 1$ , minus a weighted version of the errors reduced in tiers 1 to  $T$  by adding tier  $T + 1$ .

Part (3) of the theorem follows immediately from (H1), by rewriting it as

$$\Delta^{T+1} > C_R^{T|T+1} q_R^{T|T+1} p_R^{T+1|T+1} - C_W^{T|T+1} q_W^{T|T+1} (1 - p_W^{T+1|T+1}) \tag{H4}$$

This completes the proof.

We now consider the corollaries to the theorem.

*Corollary 1.1:* An add-on hierarchy that is error-introduction neutral and error correction neutral is superior to the underlying court structure, if and only if the added appeals process is sufficiently selective.

*Proof:* We rewrite (H2) on the assumption that the appeals process is error-introduction and error-correction neutral. The reallocation hierarchy will be superior to a flat organization, if and only if

$$\frac{q_W^{T|T+1}}{q_R^{T|T+1}} > 1 - \frac{\Delta^{T+1}}{q_R^{T|T+1} C_W^{T|T+1} (1 - p_W^{T+1|T+1})} \tag{H2.1}$$

If  $\Delta^{T+1} \leq 0$ , so that addition of the  $T + 1$  tier either degrades performance in the lower tiers or leaves them unaffected, a necessary condition for (H2.1) to hold is that  $q_W^T/q_R^T > 1$ : the appeals process from  $T$  to  $T + 1$  must be selective. But if  $\Delta^{T+1} > 0$ , so that addition of the  $T + 1$  tier improves performance below, (H2.1) can hold even if the appeals process from  $T$  to  $T + 1$  is antiselective.

*Corollary 1.2:* Consider an add-on appeals process that is selection neutral. If  $\Delta^{T+1} \leq 0$ , the add-on hierarchy is superior to the underlying court structure if and only if the error correction ratio exceeds the error introduction ratio for the add-on appeals process.

*Proof:* We rewrite (H3) on the assumption that the add-on appeals process is selection neutral:

$$\frac{1 - p_W^{T+1|T+1}}{1 - p^{T|T+1}} > \frac{p_R^{T+1|T+1}}{p^{T|T+1}} - \frac{\Delta^{T+1}}{q_W^{T|T+1} C_W^{T|T+1} (1 - p^{T|T+1})} \tag{H3.1}$$

The LHS of (H3.1) is the error-correction ratio. The first term in the RHS of (H3.1) is the error-introduction ratio. Under the maintained assumptions, the RHS of (H3.1) is greater than the error-introduction ratio. Hence, if (H3.1) is to hold, it is necessary that the error-correction ratio exceed the error-introduction ratio.

Up to this point, we have not drawn a sharp distinction between adding a new tier using additional judicial resources, and adding a new tier by reallocating a fixed number of judges. But suppose judicial performance degrades when per-judge caseload increases (that is, resources-per-case decreases). Now consider some court structure with  $T$  tiers and  $n$  judges. Should an institutional designer reallocate judges to create tier  $T + 1$ ? Doing so adds a new layer of appeals, which might seem to be a good thing, if, say, the error-correction rate exceeds the error-introduction rate at  $T + 1$ . But it will also increase per-judge caseloads below, and thus degrade judicial performance in the lower levels.

Theorem 1 indicates that the direction and magnitude of  $\Delta^{T+1}$  and the relationship among the selection, error-correction, and error-introduction ratios at  $T$  and  $T + 1$  will be critical in evaluating a reallocation hierarchy.

Suppose judicial performance is very sensitive to caseload, so removing judges from lower tiers decreases substantially the conditional error correction rate and increases substantially the conditional error-introduction rates and the error rates. Then  $\Delta^{T+1}$  may be negative in sign and large in magnitude. Given this, the performance of the new tier will have to be very good indeed to offset the degradation in judicial performance below. Conversely, if judicial performance is relatively insensitive to per-judge caseloads, so  $\Delta^{T+1}$  is negative but modest in size, the new tier need not perform so well. Finally, if the addition of the new tier dramatically improves litigant selection of cases in the lower tiers, so  $\Delta^{T+1}$  is positive in sign despite the degradation in judicial performance, the new hierarchy can be superior even if the new tier itself performs rather poorly. This somewhat counterintuitive result underscores the need to examine behavioral models of hierarchy, moving beyond a purely the structural analysis.

## Appendix B: Proof of Propositions

### PROOF OF PROPOSITION 2A

We proceed to show, by backward induction, that each player's strategy is in equilibrium, given the strategies of other players.

**APPELLATE JUDGE** There are four possibilities to consider: (1) an incorrectly losing litigant appeals from a judgment based on an uninformative signal at trial, but the appellate judge receives an informative signal; (2) a (correctly or incorrectly) losing litigant appeals from judgment based on an uninformative signal to the trial judge, *and* the appellate judge also receives an uninformative signal; (3) an incorrectly losing litigant appeals from a judgment based on an informative signal at trial; and (4) a correctly losing litigant appeals, and either the trial judge or the appellate judge receives an informative signal. (Recall that an informative signal to a court reveals the defendant's type with complete accuracy and becomes common knowledge to the judiciary.)

In case (1), the hard signal on appeal fixes the appellate judge's beliefs at 0 or 1, and the judgment follows from the logic of Proposition 1. In case (2), the litigants' strategy and Bayes's Rule fix the appellate judge's beliefs at either 0 or 1 (the former when the appellant is the defendant, the latter when the appellant is the plaintiff). Given these beliefs, the judgment again follows immediately. Case (3) is an out-of-equilibrium event, so Bayes's Rule has no bite. But the beliefs indicated in the proposition fix the appellant judge's beliefs according to the hard signal, and again the in-

dicated judgment follows. Now consider case (4), which occurs only off the equilibrium path, as the trial judgment is improperly appealed. Again, Bayes's Rule has no bite, but the specified beliefs require the appellate judge to believe the informative signal. The appellate judge thus upholds the judgment of the trial court.

**LOSING LITIGANT** There are two cases. (1) The trial court received an informative signal, and (2) the trial court received an uninformative signal.

1. Suppose no informative signal at trial ( $x^1 = 0$ ). We consider the optimal responses of an incorrectly and correctly losing litigant in turn.

A. An incorrectly losing litigant will definitely appeal, because doing so will result in either (1) an informative hard signal on appeal ( $x^2 = \beta$ ) leading to reversal, or (2) a believed signal of innocence in the absence of a hard signal ( $x^2 = 0$ ), from Bayes's Rule, again leading to reversal.

B. Given  $x^1 = 0$ , a correctly losing litigant will not appeal if the expected value from appeal is less than the sure value from not appealing, that is, for a correctly losing plaintiff  $(1 - \pi^2)d + \pi^2 0 - c \leq 0 \Rightarrow \pi^2 \geq 1 - (c/d)$  and for a correctly losing defendant  $(1 - \pi^2)0 + \pi^2(-d) - c \leq -d \Rightarrow \pi^2 \geq 1 - (c/d)$ . This is the condition indicated in the Proposition.

2. Suppose an informative signal at trial ( $x^1 = \beta$ ). Again we consider the optimal responses of a correctly losing and incorrectly losing litigant in turn.

A. A correctly losing litigant will not appeal, given the specified off-the-equilibrium path beliefs (the appellate judge believes the hard signal and thus will rule the same way as the trial judge, gaining the correctly losing litigant nothing but costing him an additional  $c$ ).

B. An incorrectly losing litigant will definitely appeal, because the appellate judge's (off the equilibrium path) belief is that the hard signal was correct, and so he reverses.

**TRIAL JUDGE** Given separation by the litigants and the appellate judge's strategy, the trial judge knows that a correct outcome will occur regardless of his judgment. But, the epsilon loss from reversal means the trial judge is not indifferent between his actions—he wants to judge correctly. Accordingly, he follows the utility-maximizing strategy indicated in Proposition 1.

### PROOF OF PROPOSITION 2B

We proceed, as in the proof of proposition 2a, by backward induction and show that each player's strategy is in equilibrium given the strategies of other players.

**APPELLATE JUDGE** If a hard signal ever reveals the defendant's type, the appellate judge rules accordingly. Suppose there is no hard signal at trial or on appeal.

If the plaintiff is the losing party at trial, then the expected value to the appellate judge of reversing is  $p_0\sigma_p^1(v^1 = \bar{l}, x^1 = 0, \beta = 1, \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0))(1) + (1 - p_0)\sigma_p^1(v^1 = \bar{l}, x^1 = 0, \beta = \bar{l}, \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0))(0)$ , which is equal to  $p_0$ , when  $\sigma_p^1(\cdot) = 1$ , which it will be if  $\rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) \geq (c - d\pi^2/d(1 - \pi^2))$  (from above).

The judge's expected value of sustaining a judgment for a defendant on appeal is  $p_0\sigma_p^1(v^1 = \bar{l}, x^1 = 0, \beta = l, \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0))(0) + (1 - p_0)\sigma_p^1(\cdot)(1)$ . The expected values of the two are equal when  $\sigma_p^1(v^1 = \bar{l}, x^1 = 0, \beta = l, \rho_i^2(\cdot)) = (p_0/1 - p_0)$ . Clearly, if  $\sigma_p^1(\cdot) > (p_0/1 - p_0)$ , the appellate judge will hold the defendant liable with certainty, and hold the defendant not liable with certainty if the strict inequality is reversed.

**LOSING LITIGANT** Recall that there are three types of losing litigants, losing plaintiffs, losing defendants of type  $l$  and losing defendants of type  $\bar{l}$ . We treat them simultaneously. If the defendant's type is revealed at trial ( $x^1 = \beta$ ), then  $v^1 = \beta$ . In this case, the (correctly) losing litigant will not appeal, given the specified off-the-equilibrium path beliefs of the appellate judge (the appellate judge believes the hard signal and thus rules the same way as the trial judge, gaining the correctly losing litigant nothing but costing him an additional  $c$ ).

If the defendant's type is not revealed at trial ( $x^1 = 0$ ), an incorrectly losing plaintiff will appeal if the expected value of the appellate lottery is greater than submitting to the incorrect trial verdict, that is, appeal if  $\pi^2d + (1 - \pi^2)(d\rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) + (1 - \rho_i^2(\cdot))0) - c > 0 \Leftrightarrow \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) > (c - \pi^2d)/(1 - \pi^2)d$  and not if the inequality goes the other way. (It will be shown momentarily that the former inequality holds.) Similarly, an incorrectly losing defendant should appeal if  $\pi^20 + (1 - \pi^2)(-d\rho_i^2(v^1 = l, x^1 = x^2 = 0) + (1 - \rho_i^2(\cdot))0) - c > -d \Leftrightarrow \rho_i^2(v^1 = l, x^1 = x^2 = 0) < (d - c)/(1 - \pi^2)d$ , and not if the inequality goes the other way. (It will be shown momentarily that the former inequality holds.)

If the defendant's type is not revealed at trial ( $x^1 = 0$ ), a correctly losing plaintiff may appeal. Recall, however, that, given the strategy of the trial judge, the value of  $p_0$  determines whether there are correctly losing plaintiffs or correctly losing defendants. Thus, there are correctly losing defendants if and only if  $p_0 \geq 1/2$  while there are correctly losing plaintiffs if and only if  $p_0 \leq 1/2$ .

A correctly losing plaintiff will appeal, if  $\pi^20 + (1 - \pi^2)(d\rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) + (1 - \rho_i^2(\cdot))0) - c \geq 0 \Leftrightarrow \rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) > (c/(1 - \pi^2)d)$ ,

and not if the inequality goes the other way. If this relation holds with equality (and the trial signal is uninformative), a correctly losing plaintiff will be indifferent between appealing and not appealing, and so can mix. Similarly, if the defendant's type is not revealed at trial, a correctly losing defendant will appeal, if  $-d\pi^2 + (1 - \pi^2)(-d\rho_i^2(v^1 = l, x^1 = x^2 = 0) + 1 - \rho_i^2(\cdot))0 - c > -d \Leftrightarrow \rho_i^2(v^1 = l, x^1 = x^2 = 0) \leq 1 - (c/(1 - \pi^2)d)$ , and not if the inequality goes the other way. Again, if this relation holds with equality, a correctly losing defendant will be indifferent between appealing and not appealing and so can mix (absent a hard signal at trial). A similar exercise for an appealing defendant leads to the strategy indicated by  $\rho_i^2(v^1 = l)$ .

It only remains to show that  $\rho_i^2(v^1 = \bar{l}, x^1 = x^2 = 0) = (c/d(1 - \pi^2))$  compels an incorrectly losing plaintiff to appeal, while  $\rho_i^2(v^1 = l, x^1 = x^2 = 0) = 1 - (c/d(1 - \pi^2))$  compels an incorrectly losing defendant to appeal. To establish the first part, note that  $(c/d(1 - \pi^2)) > (c - d\pi^2/d(1 - \pi^2)) \forall \pi^2 > 0$ ; the result then follows from plaintiff's appeal strategy, above. To establish the second part, note that  $1 - (c/d(1 - \pi^2)) = (d(1 - \pi^2) - c/d(1 - \pi^2)) < (d - c/d(1 - \pi^2)) \forall \pi^2 > 0$ . The result then follows from defendant's appeal strategy, above.

**TRIAL JUDGE** The trial judge behaves in the by-now familiar fashion, using hard signals if available and otherwise efficiently using prior beliefs.

#### PROOF OF PROPOSITION 2D

Using the above strategies, if  $p_0 > .5$ , the probability of a correct outcome is  $\pi^1 + (1 - \pi^1)\{\pi^2 + 1 - \pi^2[(1 - p_0)(1)(c/d(1 - \pi^2)) + p_0[(1 - p_0/p_0)(1 - (c/d(1 - \pi^2)))] + (2p_0 - 1/p_0)]\}$  which simplifies to the expression in the Proposition. Similarly, if  $p_0 \leq .5$  the probability of a correct outcome is  $\pi^1 + (1 - \pi^1)\{\pi^2 + (1 - \pi^2)[(1 - p_0)\{(p_0/1 - p_0)(1 - (c/d(1 - \pi^2))) + (1 - 2p_0/1 - p_0)\} + p_0(1)(c/d(1 - \pi^2))]\}$  which again simplifies as indicated.

#### PROOF OF PROPOSITION 3

The proof proceeds via backward induction.

**HIGH COURT (TIER 3) JUDGE** Appeals to the high court are out-of-equilibrium events, so Bayes's Rule has no bite. However, we require the high court judge's beliefs to be fixed in the natural way if any  $x^t \neq 0$ , ( $t = 1, 2, 3$ ). In that case, the indicated judgments follow from Proposition 1. Absent a hard signal, the most-favorable belief to appeals (and difficult for the equilibrium) is that an appeal of the intermediate court's judgment



signals  $v^2 \neq \beta$ . We assume this belief. But again, given this belief, the indicated judgment follows immediately.

**LOSING DEFENDANT AT LEVEL 2** If  $x^1$  or  $x^2 = \beta$ , the defendant surely appeals adverse  $v^2 = l \neq \beta$  because in this case, following appeal,  $\mu_i^3 = 0$  (from the specified out-of-equilibrium beliefs) and the defendant prevails. Conversely, if  $x^1$  or  $x^2 = \beta$  and  $v^2 = l = \beta$ , Defendant definitely does not appeal, because in this case,  $\mu_i^3 = 1$  (from the specified out-of-equilibrium beliefs) and the defendant loses at additional cost of  $c$ . If  $x^1 = x^2 = 0$ , incorrectly losing defendant surely appeals, because either  $x^3 = \beta$  and thus  $\mu_i^3 = 0$  and high court reverses, or  $x^3 = 0$  and thus  $\mu_i^3 = 0$  (from the specified out-of-equilibrium beliefs) and high court again reverses. If  $x^1 = x^2 = 0$ , the correctly losing defendant appeals if and only if the expected value of appealing is greater than or equal to the expected value of not appealing, to wit,  $(1 - \pi^3)0 + \pi^3(-d) - c \geq -d \Rightarrow \pi^3 \leq 1 - (c/d)$ . But this contradicts the condition on high court accuracy assumed in the equilibrium.

**LOSING PLAINTIFF AT LEVEL 2** The argument for the plaintiff is analogous and omitted for the sake of brevity.

**INTERMEDIATE APPELLATE (TIER 2) JUDGE** There are three cases. Case (1)  $x^1 = \beta$ . An appeal following a hard signal at trial is an out-of-equilibrium action so Bayes's Rule has no bite. We specify that the appellate judge believes the hard signal (so that  $\mu_i^2 = 0$  or 1, as  $x^2 = \bar{l}$  or  $l$ , respectively), and the indicated judgments follow immediately from Proposition 1.

Case (2)  $x^1 = 0$ ,  $x^2 = \beta$ . In this case,  $\mu_i^2 = 0$  or 1, as  $x^2 = \bar{l}$  or  $l$ , respectively, and the indicated judgment follows from Proposition 1.

Case (3)  $x^1 = 0$ ,  $x^2 = 0$ . Given the appellate strategies of the litigants and Bayes's Rule,  $\mu_i^2 = 0$  if the losing defendant appeals, and  $\mu_i^2 = 1$  if the losing plaintiff appeals. Again, the indicated judgment follows from Proposition 1.

**LITIGANT LOSING AT TRIAL** There are two cases. Case (1) Incorrectly losing litigant. An appeal definitely is reversed by the court at tier 2, with or without a hard signal. Following the reversal, at tier 2, the correctly losing litigant does not appeal so that the correct judgment stands. Hence, an incorrectly losing litigant at trial definitely appeals.

Case (2) Correctly losing litigant. If there is a hard signal at trial, an appeal gains nothing and costs  $c$ , so is not undertaken. Suppose no hard signal at trial. If a hard signal emerges at appeal, the appellant loses again, and further appeal is hopeless. If a hard signal does not emerge at appeal at tier 2, the appellant definitely wins at tier 2. But (from above) the (now

incorrectly) losing litigant, after appeal, will definitely appeal to the high court. Hence a correctly losing defendant should appeal (absent a hard signal at trial) if and only if  $\pi_i^2(-d) + (1 - \pi_i^2)(\pi_i^3(-d) + (1 - \pi_i^3)0) - c \geq -d \Rightarrow \pi_i^2 < 1 - (c/d(1 - \pi_i^3))$ . But this contradicts the condition assumed in the equilibrium (i.e., even if  $\pi_i^2 = 0$ ). The condition for a correctly losing plaintiff is identical and a derivation is omitted for the sake of brevity.

## Notes

Cameron gratefully acknowledges support from NSF Grant 0079952 and the Center for the Study of Democratic Politics, Woodrow Wilson School of Public and International Affairs, Princeton University. Kornhauser acknowledges support from the Max E. Greenberg and Filomen D'Agostino Research Fund of NYU School of Law.

1. Schwartz 1995 criticizes Shavell 1995 on this point, suggesting that perfect separation of "legitimate" from "illegitimate" appeals (or initial cases) may not be compatible with individually rational behavior.

2. Cameron, Segal, and Songer 2000 consider strategic auditing, as does Spitzer and Talley 1998. Cameron 1993 sketches a model of judicial tournaments (see also Kornhauser 1995). Judicial tournaments are then explored in more detail in McNollgast 1995. Shavell 1995, footnote 2, provides citations to the literatures on appeals by employers and in administrative agencies.

3. Note that Shavell 1995 does not fully characterize an appeals process because the paper does not define  $q'_n$  and  $q''_n$ .

4. Appendix A to this essay provides the proof for a theorem about the desirability of an *add-on appeal process*.

5. Shavell's model compares the "optimal" flat organization to the "optimal" two-tiered hierarchy.

6. Parts of this section draw heavily on Kornhauser 1995.

7. As is well known, an initial informational asymmetry across the litigants is essential; otherwise, they will settle (if possible).

8. We assume the defendant pays the court costs for the trial.

9. We assume that cases are divided equally over judges and that the judges divide their time equally over cases. It is straightforward to show that error-minimizing judges should allocate their time equally over objectively indistinguishable cases.

10. Obviously, this is a rather special judicial technology. In Cameron and Kornhauser 2003, we consider a technology in which trials and appeals are always somewhat informative but never perfectly so.

11. This is refinement D1 in Cho and Kreps 1987, and is equivalent to Banks and Sobel's universal divinity.

12. Phrased differently, as the resources devoted to "fact-finding" at the trial level increase, the probability that there is mutual revelation of type to the parties approaches 1.

13. We note a rather subtle issue in defining the key structural variables of section 2. They may be calculated in several different ways. They might be defined in terms of the behavior required by the equilibrium strategy or in terms of actual equilibrium behavior. These variables might also be defined at each information set—when the trial court re-

ceived a hard signal and when it did not. Or, the variables might be defined simply in aggregate terms. As the variables may have different values at different information sets, we choose to define a "aggregate" variable specified in terms of actual equilibrium behavior. The definition permits us to invoke Theorem 1 and its corollaries. For example, consider  $1 - p_R^{2/2}$ , the probability that, on appeal, a case correctly decided at level 1 will be upheld. This event has a probability 0, because, in equilibrium, no correctly decided cases are appealed. On an actual, aggregate-behavior definition, this variable is thus undefined. Using a definition based on the equilibrium strategy at different information sets, the value of  $1 - p_R^{2/2} = 1$ , if there was a hard signal. If there was no hard signal, then the probability  $1 - p_R^{2/2}$  equals  $\pi^{2/2}((C^{2/2}/n^{2/2}))$ , the probability that the appellate court receives a hard signal that reveals the appellant's type. By contrast  $1 - p_W^{2/2}$ , the probability that, on appeal, a wrongly decided case will be reversed, does occur in equilibrium, and it equals 1 under both an aggregate behavioral definition and on a definition that relies on equilibrium strategies at each information set.

14. The case of  $p_0 < 1/2$  is virtually identical in reasoning and is omitted for the sake of brevity.

15. This is the rightly decided appealed cases, over all rightly decided cases.

16. But, see Daughety and Reinganum 2000 for a contrary view.