RAND Journal of Economics
Vol. 31, No. 3, Autumn 2000
pp. 502–525

Appealing judgments

Andrew F. Daughety*

and

Jennifer F. Reinganum*

We use axiomatic and Bayesian methods to model information and decisions in a hierarchical judicial system. Axioms represent constraints that rules of evidence, procedure, and higher court review impose at the trial level; a one-parameter family of functions provides their unique continuous solution. This generates a type space for the appeals stage wherein the appellant and the appeals court each receive private signals of a yet superior court's value of the parameter, reflecting its interpretation of the law. The appeals court uses the defendant's choice about appeal to improve its estimate of the superior court's preferred interpretation of the law.

1. Introduction

■ In this article we model judicial decision making in a hierarchy of courts, in particular, a trial court and an appeals court. The first stage involves a trial court that aggregates credible evidence provided by the parties, subject to a set of axioms representing the constraints that rules of evidence, procedure, and higher court review impose. The second stage involves a Bayesian appeals court that uses information gleaned from the past behavior of a yet higher court and, in part, from the actions of the appellant.

Trial courts and appeals courts fulfill different roles in the U.S. legal system. In trial courts, evidence about a case is presented and the facts are determined. As we shall discuss in Section 2, aside from assessing the credibility of evidence, the law goes out of its way to limit and/or direct the use of (Bayesian) inference by a trial court. This suggests that an axiomatic approach may be useful here; we employ such an approach in Section 3 and find a parametrized family of functions that is the unique continuous solution to the axioms provided. This parameter can be viewed as a measure of the breadth of a law's interpretation, which leads naturally to our model of the appeals process. Although a preliminary interpretation by the trial court is required to generate a decision, ultimately, legal interpretation is the province of appeals courts.

Although they inherit the facts found by the trial court, appeals courts exercise substantial discretion in interpreting the law, subject only to risking reversal because their interpretation differs from that of a yet superior court. Thus, a Bayesian appeals

* Vanderbilt University; andrew.f.daughety@vanderbilt.edu, jennifer.f.reinganum@vanderbilt.edu.

We thank Larry Ausubel, James Foster, Tracy Lewis, R. Preston McAfee, Richard A. Posner, Nicholas Zeppos, and the referees for helpful conversations and suggestions.

court, which tries to estimate the interpretation of the supreme court, is of interest. Since courts are responders (in the case of appeals courts, there must first be an appellant who brings a case), they can draw inferences both from the appellant's decision to pursue an appeal and from other sources (e.g., supreme court decisions). Of course, strategic appellants understand this and account for it when they evaluate the value of pursuing an appeal. We provide such a model in Section 4. In Section 5 we consider some of the implications of the perfect Bayesian equilibrium. In particular, we find that if reduced resources for such courts increase their reliance on the information generated by an appellant's actions, this can result in reduced reliability of that signal, leading to less well-informed appeals court decisions. An increase in the award or the evidentiary standard (an exogenously determined level of certainty on the part of the court necessary to find liability), or a decrease in the cost of appeal, increases the equilibrium likelihood of appeal. The model also provides a precise characterization of the legal concept of ''harmless error,'' under which errors by trial courts are noted by an appeals court but are deemed not to provide a sufficient basis for reversal of the trial court's verdict. Finally, Section 6 provides a summary. All proofs are in the Appendix.

2. Background, modelling issues, and related literature

■ In this section we provide some background on trial and appeals courts, discuss some modelling issues and how we plan to deal with them, and very briefly describe how others have modelled court decision making; for a more extensive discussion of the related literature in economics, law, and political science, see Daughety and Reinganum (1999).

□ **Background on what trial and appeals courts do.** Trial and appeals courts perform substantially different functions. Trial courts determine the ''facts'' in a case and interpret and apply the law to those facts to come to a judgment (e.g., about liability), while appeals courts take the facts as determined by the trial court and engage in *de novo* review of issues of law. For example, in a case alleging manufacture and distribution of a defectively designed product,[1] such as cases pending against cigarette manufacturers, facts might include whether the plaintiff was exposed to the defendant's product, the nature and extent of the exposure (e.g., first-hand, second-hand, *in utero*), and information about the composition of the inputs (e.g., tobacco leaves, chemical additives, and filter material). In making a judgment about whether or not the manufacturer should be found liable, a legal issue (requiring an interpretation of the law) is: To whom does the firm owe a duty to take care? Possible answers include no one; only smokers; smokers and those others who directly inhale the smoke; or smokers, those others who directly inhale the smoke, and those others who are affected by smoking through yet more indirect means such as *in utero* exposure. A given level of care taken by the firm may be adequate (resulting in a finding of no liability) if the firm's duty is narrowly construed, but the same level of care may be inadequate (resulting in a finding of liability) if the firm's duty is broadly construed.[2] The trial court must employ some interpretation of the law in order to reach a decision; however, this interpretation is subject to appellate review. ''Appellate courts give plenary review to pure issues of law; that is, they do not give any deference to the trial judge's view on such issues

[1] A product is defectively designed if a safer alternative design, which was ''cost effective,'' was available but not chosen.

[2] Whether a defendant exercised due care is often determined via the ''Hand Formula,'' which specifies that optimal care equates the marginal cost of care to the expected marginal reduction in harm. Clearly, this level depends on the magnitude of the included harms and, thus, on how broadly duty is construed.

. . . . But appellate courts do defer quite broadly to the trial judge's or jury's findings of fact'' (Posner, 1998, p. 643). The aforementioned division of labor is reflected in the resources devoted to fact finding versus legal research at each level. Trial courts do not receive significant resources for the purpose of extensive legal research, but they devote considerable time and expense to documenting evidence. On the other hand, appeals courts take the facts as determined by the trial court and focus their time, attention, and resources (e.g., law clerks) on research and evaluation of the legal issues.

Appeals courts consider two types of appeals, both of which involve potential errors made in the trial. One type of appeal concerns alleged violations of evidentiary or procedural rules. For example, an appellant might claim that evidence was excluded or included erroneously in the trial. We do not consider appeals of this type, but it should become clear that the model could be directly extended to address them. The other type of appeal concerns the interpretation of law. For example, a cigarette manufacturer found liable for injuries due to second-hand smoke might appeal on the basis that the trial court construed the firm's duty too broadly. That is, the firm wants the appeals court to substitute a different (and narrower) interpretation for that of the trial court. The breadth of the interpretation of a law is the main focus of our model of an appeals court.

□ **Modelling issues.** As economists, we are naturally tempted to model the trial court as a Bayesian decision maker; that is, the trial court posits a joint distribution over the defendant's true liability, the evidence provided by the plaintiff, and the evidence provided by the defendant, and uses the evidence to estimate the defendant's true liability. However, inasmuch as the parties' evidence is properly regarded as the outcome of strategic search processes in which the trial court observes only the evidence the parties actually present, there is a huge amount of ''missing data,'' which would tend to make Bayesian inference highly prior-dependent. For instance, there may be relevant evidence that is available but not presented; this might occur because a party chooses to suppress it. Relevant evidence may also be ruled inadmissible with respect to a determination of liability; when such evidence is admitted inadvertently, jurors are admonished to ignore it and a judge is also expected to ignore such evidence if he is the fact finder.[3] Another aspect of missing data is the extent of a party's ''effort'' (e.g., how much the party spent on the search for evidence). Certain characteristics that might help one to infer the extent of search, such as the parties' wealth and their costs of search, are also unobservable. Finally, the underlying true liability of the defendant, which presumably affects the defendant's ability to find (and the cost of) exculpatory evidence, is unobservable to the court (though it may be known by one or both parties). To use a Bayesian model of the trial court, one must substitute a subjective prior distribution for all of this ''missing data.'' While this is certainly a theoretical possibility, as will be discussed below, it does not seem consistent with many aspects of the rules of evidence and civil procedure.

The legal literature definitely recognizes some legitimate uses of statistical inference in trials; one example can be found in cases involving DNA evidence, where the probability of misclassification can be clearly quantified. In addition, there are situations in which the trial court is specifically allowed to exercise discretion (thus permitting Bayesian decision making in those instances), such as assessing the credibility of witnesses or interpreting the law; we address the former issue briefly below and return to the latter issue at the beginning of Section 4.

[3] This is the case with, e.g., failed settlement negotiations, character evidence, and the insurance status of the defendant; see Federal Rules of Evidence 408, 404, and 411, respectively, in Graham (1992).

However, there are also indications that the trial court process of fact finding and aggregation is *not* purely Bayesian but is constrained by rules of evidence and procedure, some of which conflict directly with Bayesian decision making but may promote broader values than accuracy in the case at hand.[4] For instance, in some situations the trial court is admonished to draw *no* conclusion from a party's decision not to present certain evidence (the foremost example of this is the privilege against self-incrimination). As another example, if the plaintiff submits *no* evidence it is generally conceded that he loses. Moreover, the plaintiff loses if *only* statistical evidence is provided, even if it would appear to suggest that the defendant's liability is likely.[5] This is because the plaintiff is allocated the ''burden of production'' of evidence, to discourage frivolous suits. Finally, in some instances the law requires a specific inference through the use of presumptions. For example, in employment discrimination suits, the *McDonnell Douglas* rule ''permits a plaintiff . . . to establish his prima facie case . . . with evidence merely that he was qualified for the job but was passed over in favor of someone of another race. . . . Satisfying the just-described burden of production creates a presumption of discrimination, meaning that if the defendant puts in no evidence the plaintiff is entitled to summary judgment,'' even though ''the probability that he lost the job opportunity *because* he was discriminated against might not seem to be very high if the only evidence is as described'' (Posner, 1999, p. 1503).

In addition, there are rules that allow a trial court judge to preempt or overrule a jury. For instance, if the judge determines that the evidence is insufficient to support (that is, it cannot be construed as supporting) a verdict of liable, he may dismiss the case or go further and enter a directed verdict in favor of the defendant. Similarly, after a verdict, a judge can overrule the jury by entering a ''judgment notwithstanding the verdict,'' if he determines that the evidence was insufficient to support the verdict. James and Hazard (1985) discuss in detail the mechanisms available to a judge for focusing a jury's attention on the evidence presented at trial.

All the aforementioned rules and conventions are inconsistent with a purely Bayesian model, since one can construct reasonable subjective priors which, given the evidence, would result in a different decision than that dictated by the rule or convention. Rather, it is plausible to interpret such rules as focusing decision making on the evidence provided at trial and discouraging the substitution of the fact finder's subjective prior for evidence. To the extent that a decision relies on a (possibly strong) subjective prior, this reliance reduces the incentives for the parties to provide evidence (Posner, 1999). Presumably, exculpatory evidence is easier to produce if the defendant really has been careful, so reducing the value of exculpatory evidence also reduces the defendant's incentive to take care. To support the provision of care and evidence, it is reasonable for the legal system to try to restrict the fact finder's reliance on subjective priors and to focus instead on the evidence presented at trial.

Thus, one approach to modelling trial court decision making is to incorporate Bayesian inference where it would appear to be permitted (i.e., in assessing the credibility of witnesses or in the interpretation of law), and to employ other methods of decision making where it would appear to be constrained (i.e., in the trial court's assessment of the defendant's liability). We pursue this strategy in Section 3 by posing

[4] See Posner (1999) for efficiency-based rationales for many rules of evidence and procedure, and Lewis and Poitevin (1997) and Sanchirico (1997c) for models wherein a sophisticated Bayesian decision maker commits *ex ante* to a decision rule that does not make optimal use of information *ex post.*

[5] If ''the only evidence the victim of a bus accident had linking the accident to the defendant bus company was that the defendant operated 80 percent of the buses on the route where the accident occurred, the victim could not win without additional evidence of the defendant's liability'' (Posner, 1998, p. 604).

a collection of axioms for trial court behavior that we believe is broadly descriptive of (rule-constrained) decision making.

□ **Related literature.** In the remainder of this section, we briefly examine how others have modelled court decision making. A variety of other models of judicial decision making have been suggested.[6] For example, one branch of the literature specifies that the plaintiff's probability of prevailing is given by an exogenous function of the litigants' levels of effort or expenditure.[7] (Danzon, 1983; Braeutigam, Owen, and Panzar, 1984; Hause, 1989; Katz, 1987, 1988; Plott, 1987; and Landes, 1993).

A second branch of the literature employs a Bayesian model of the court in which effort (or expenditure) signals the defendant's type (Rubinfeld and Sappington, 1987; and Sanchirico 1997a, 1997b), sometimes allowing the decision maker to reallocate the burden of proof (Sobel, 1985; and Shin, 1994, 1998). In Daughety and Reinganum (1995), we examined the impact of making (currently inadmissible) pretrial negotiations admissible as evidence at trial before a Bayesian court.

A third branch of the literature assumes that information that is common knowledge to the parties is initially unknown to (or not observed by) the court (Milgrom and Roberts, 1986; Lipman and Seppi, 1995; and Seidmann and Winter, 1997). A further twist on this branch assumes that the parties have private information relative to the court and relative to each other (Froeb and Kobayashi, 1993, 1996; Farmer and Pecorino, forthcoming; and Daughety and Reinganum, forthcoming).

There are relatively few combined models of trial and appeal. Shavell (1995) shows that by making an appeals court more accurate than a trial court and imposing an appropriate fee or subsidy on appeals, only cases of error will be appealed. Since the cases of error are a subset of all cases, this rationalizes investing disproportionately in the accuracy of appeals courts. Our model differs from Shavell's in many respects, but in particular his model does not account for information contained in the decision to appeal. He notes that accounting for this information would unravel the result that only cases of error will be appealed, and therefore he assumes that the appeals court is committed not to make any inference from the decision to appeal (at the same time noting that there does not seem to be any doctrine forbidding such inference). In Section 4, we show how such inference can be incorporated in a Bayesian model of the appeals process.

Spitzer and Talley (1998) use an incomplete-information model to examine trial and appeal. Their model is similar in some respects to our model in Section 4, but there are several important differences as well. First, trial and appeal pertain to the same issue in their model, whereas we distinguish between the trial court's focus on facts and the appeals court's focus on issues of law. Second, there is no difference in ideology between the trial and appeals courts in our model, whereas ideological conflict between trial and appeals courts is an important element in their model. Finally, there are several differences between the models in terms of ‘‘who knows what when.’’

This article has both a prequel and a sequel. Daughety and Reinganum (forthcoming) model evidence generation (which occurs prior to its presentation and aggregation as modelled in this article) as strategic sequential search in a two-stage analysis of the determination of liability, followed by damages, using the liability assessment function derived below in Section 3. Daughety and Reinganum (1999) use a nonstrategic version

[6] There is an enormous literature in political science, law, and psychology on modelling decision making by jurors and/or judges, voting within courts, the objectives of judges, and judicial independence. For a partial bibliography, see Daughety and Reinganum (1999).

[7] Skaperdas (1996) has recently provided an axiomatic basis for the relative-effort model. We thank Tracy Lewis for making us aware of this related article.

of the appeals court model from Section 4 in the context of a collection of horizontally related courts (such as the U.S. Circuit Courts of Appeals). Decisions of such courts are not binding upon each other, but they may be influential. We show that this ''persuasive influence'' can result in herding behavior by such courts, in which each subsequent court to take up a particular issue relies increasingly less on its own legal research and increasingly more on the previous decisions of other courts. Unlike herding in markets (where an announcement may eliminate inefficient herding), the fact that courts must wait for cases to be brought means that herding on the wrong outcome may persist indefinitely.

3. An axiomatic model of evidence aggregation and liability assessment in a trial court

■ The purpose of this axiomatic analysis is twofold. First, it provides a positive model of a trial court's evidence assessment process, wherein submitted credible evidence is aggregated into an overall assessment that is responsive to the individual assessments provided (the evidence) and that respects basic notions of procedural fairness inherent in the law. The court's assessment is then compared to an exogenously imposed evidentiary standard, which yields a binary decision: the defendant is found to be either liable or not liable. Second, the analysis produces a one-parameter family of functions that aggregates the submitted evidence into the court's assessment. This parameter has a natural interpretation as the breadth of applicable legal concepts and its domain forms the type space for the model to be described in Section 4.

Our focus here will be on a civil suit and the issue will be liability, so the outcome of interest concerns a defendant being found liable or not liable. For modelling convenience, we assume the value of the award is common knowledge. This means that the purpose of the trial is to allow both sides (the plaintiff, denoted P, and the defendant, denoted D) to provide evidence about D's liability and to come to a judgment as to whether D will be held liable or not. In reality, a litigant's evidence (his ''case'') is made up of observations, purported facts, and assertions, some of which may be of uncertain credibility (e.g., unsupported, self-serving statements) and others more likely to be viewed as credible (e.g., the opinions of qualified expert witnesses). We abstract from the credibility assessment issue by assuming that both parties provide credible evidence (any noncredible evidence having been discounted or discredited).[8] Thus, each litigant's case is a bundle obtained by sequentially sampling such a (credible evidence) space, keeping items that support the case to be made and discarding those that do not.[9] We associate a liability assessment, which is a probability that D should be found liable, with a case. If we think of a litigant's case as a stack of pages,[10] then two observations follow. First, the stronger case need not be the case with the higher stack of pages (i.e., it is not merely effort or expenditure that determines the strength of the case). Second, one could assign assessments to a partition of a stack of papers for a case, and then each subset of the original stack would have an associated assessment. Note that while the pages in each substack sum to the total number of pages in the

[8] Alternatively, one could view the overall trial as having two stages; the first involves evaluating the credibility of evidence using a Bayesian approach, and the second is our focus here. From this perspective, the litigants provide credible evidence either directly or as the result of ''preprocessing'' for credibility.

[9] Thus, such items may not be presented at trial unless the other litigant also obtains them through his sequential sampling process, since the two litigants have access to the same evidence space.

[10] For example, in a products liability case, one expert may address the production process while another addresses labelling of the product. In the cigarette example, a second-hand smoke case will require various medical experts as well as ventilation experts (e.g., if the setting is an office or an airplane).

case's stack, the assessments of the substacks need not sum to the assessment for the stack: a case is, generally, multidimensional in what it addresses, and there may be weaker and stronger elements, all of which are combined to ''make a case.''

Courts aggregate the assessments of the litigant's cases to come to an overall assessment of D's liability. They then compare this composite assessment with an ''evidentiary standard,'' which is a hurdle prescribed by law (usually via a statute). Only if the court's assessment meets this hurdle is D found liable. Thus, we model the trial court's decision process as made up of three elements: (1) credible evidence presented by a plaintiff and a defendant; (2) a function that aggregates the evidence into the court's assessment of the defendant's likelihood of being liable; and (3) an evidentiary standard with which to compare the court's assessment. These three elements lead to an outcome indicating a winner and a loser.

More formally, let π^P and π^D denote the assessments of D's liability proffered by P and D, respectively, where $\pi^P \in [0, 1]$ and $\pi^D \in [0, 1]$. Let Π be the *assessment space*, that is, $\Pi \equiv [0, 1] \times [0, 1]$; any point in this space represents the pair of assessments summarizing the two cases provided by the litigants. Thus, a higher π^P reflects a stronger case for P, while a higher π^D reflects a weaker case for D. The court's liability assessment process is represented by the function $\ell(\pi^P, \pi^D)$, where ℓ: $\Pi \rightarrow [0, 1]$; details of this function will be provided below. Finally, let γ be the exogenously specified evidentiary standard, with $\gamma \in [0, 1]$. Three levels of γ are frequently employed: (1) ''preponderance of the evidence'' appears to translate to (approximately) .51 and applies in most civil cases; (2) ''clear and convincing'' is murkier, appearing to lie somewhere in the interval [.6, .8], and is used by the majority of states in punitive damages cases; and (3) ''beyond a reasonable doubt'' involves a very high level of γ (say, .99) and is used in criminal cases and by Colorado in punitive damages cases.

Thus, given a court's assessment $\ell(\pi^P, \pi^D)$, the defendant is found liable if $\ell(\pi^P, \pi^D) \geq \gamma$. We formalize this as the liability determination function, $L(\ell(\pi^P, \pi^D); \gamma)$, where

$$L(\ell(\pi^P, \pi^D); \gamma) = \begin{cases} 1 & \text{if } \ell(\pi^P, \pi^D) \geq \gamma \\ 0 & \text{otherwise.} \end{cases}$$

This function provides the binary outcome of the trial process, and it plays a basic role in the appeals model to be discussed in the next section.

Next, we consider the court's assessment function, $\ell(\pi^P, \pi^D)$. Based on the discussion in the previous section, the assessment function should embody two characteristics reflective of the rules of evidence and civil procedure. Thus, the assessment should (1) be responsive to the evidence provided and (2) reflect notions of procedural fairness, meaning that credible evidence provided by the parties should be used in an unbiased manner. In what follows we assume that ℓ is continuous for all $(\pi^P, \pi^D) \in \Pi$, and that the indicated properties are to hold for all $(\pi^P, \pi^D) \in \Pi$.

The first characteristic (*responsiveness*) is fairly straightforward to implement, which we do via the properties of *strict monotonicity* and *interiority*.

$$\pi^{P*} > \pi^P, \quad \pi^{D*} > \pi^D$$
$$\Rightarrow \ell(\pi^{P*}, \pi^D) > \ell(\pi^P, \pi^D) \quad \text{and} \quad \ell(\pi^P, \pi^{D*}) > \ell(\pi^P, \pi^D); \qquad \text{(Strict Monotonicity)}$$

$$\max\{\pi^P, \pi^D\} \geq \ell(\pi^P, \pi^D) \geq \min\{\pi^P, \pi^D\}. \qquad \text{(Interiority)}$$

Strict monotonicity indicates that $\ell(\pi^P, \pi^D)$ is strictly increasing; since π^P and π^D are assessments of the defendant's liability, an increase in either party's assessment should lead to an increase in the court's assessment. Interiority argues that the court's assessment should lie within the range of assessments provided by the litigants; we argued in Section 2 that judicial options such as involuntary dismissal, directed verdicts, and judgment notwithstanding the verdict can be viewed as trying to confine the court's assessment to this interval.

Note that an implication of interiority is the property of *reflexivity*: $\ell(z, z) = z$, that is, if both litigants submit evidence indicating the same level of liability (i.e., z), then the outcome of the court's assessment process would be that same level. The intuitively plausible nature of reflexivity is another reason why Bayesian models of a trial court, while well suited to the issue of inferring credibility, have difficulty capturing the restrictions that the legal system reasonably places on the court's assessment process once noncredible evidence has been excluded or discounted. For the class of Bayesian inference models wherein the joint distribution over the true liability (denoted ρ) and the evidence (π^P, π^D) is strictly positive on $[0, 1] \times \Pi$, reflexivity cannot hold (see the Appendix).

The second main characteristic, namely *procedural fairness,* requires that credible evidence provided by the litigants be treated in an unbiased manner. We implement this via three further axioms, *unbiasedness, burden of production,* and *independence of presentation.*

(Unbiasedness):

$$\text{(a)} \qquad \ell(\pi^P, \pi^D) = \ell(\pi^D, \pi^P); \qquad \text{(Symmetry)}$$

$$\text{(b)} \qquad \ell(\lambda\pi^P, \lambda\pi^D) = \lambda\ell(\pi^P, \pi^D) \quad \forall\lambda, \quad 0 < \lambda \le 1; \qquad \text{(Homogeneity)}$$

$$L(\ell(0, \pi^D); \gamma) = 0 \quad \forall\pi^D \in [0, 1], \quad \forall\gamma > .5; \qquad \text{(Burden of Production)}$$

$$\forall u, v, w, z \in [0, 1], \qquad \ell(\ell(u, w), \ell(v, z)) = \ell(\ell(u, z), \ell(v, w)).$$

(Independence of Presentation)

Unbiasedness is invoked via an absolute and a proportional property. The absolute property is symmetry: since the evidence is credible, it should not matter in the court's assessment who provided the evidence; the assessment based on the two evidence submissions should be the same if the two submissions were switched. This assumption is potentially controversial; we examine the implications of relaxing it in the Appendix.

The second unbiasedness property, homogeneity, is a relative statement requiring that proportional scaling alone of the evidence should not influence the outcome disproportionately toward one party or the other. Thus, for example, if we were to cut both estimates in half, the court's assessment should fall, but there is no obvious reason why it should fall disproportionately for P or for D. If the court's assessment were to fall by less than half, this suggests a pro-plaintiff bias, while a fall by greater than half suggests a pro-defendant bias. The seemingly natural assumption is that the overall assessment should be reduced to half of the original assessment. Notice that reflexivity implies that $\ell(\lambda z, \lambda z) = \lambda z = \lambda\ell(z, z)$; thus reflexivity implies that $\ell(\pi^P, \pi^D)$ is homogeneous of degree one along the ray $\pi^P = \pi^D$. Our assumption of homogeneity simply extends this property to the rays other than 45° (that is, to rays $\pi^P = \alpha\pi^D$, $0 < \alpha < \infty$).

The burden-of-production axiom expresses the notion that if P does not submit any evidence of D's liability (which we represent[11] as $\pi^P = 0$, since this is the assessment for the weakest possible case for P), then D should be found not liable ($L = 0$) for any evidentiary standard (γ) greater than .5. As discussed in Section 2, since it is P who has sued D, the initial burden of production falls on P. Thus, P must produce some evidence at trial, or else D will benefit from (at the least) an involuntary dismissal of P's case. While this property appears to favor D, it actually acts to redress the asymmetry created by P's suit, by requiring P to present some credible evidence. We have assumed this property for all relevant evidentiary standards, that is, for $\gamma > .5$.

Finally, independence of presentation guarantees that the court's assessment is independent of such extraneous factors as the nature, style, or sequence of presentation of the cases. To see this, recall the analogy of P's case to a stack of pages, the totality of which yields the assessment π^P; similarly, D's case consists of a stack of pages, the totality of which yields the assessment π^D. As suggested earlier, each of these stacks could be (arbitrarily) divided into two parts so that, say, for P the totality of the evidence in the first substack yielded assessment u while that for the remaining substack yielded assessment v (recall that this does not presume $\pi^P = u + v$, nor does it presume $\pi^P > u$ and v). Similarly, let the substack assessments for D be denoted w and z. This property asserts that the court's decision process will come to the same conclusion by comparison of the substacks, followed by comparisons of the assessments based on the substacks, independently of how the substacks are compared. Thus, note that on the left, u is compared with w and v with z, while on the right, w and z have been switched. Essentially, this axiom removes any role for psychology or ''style'' from the court's assessment procedure: the court's assessment is not influenced by how the cases are presented. Alternatively, we assume that both parties have access to equally qualified litigators, and that any such effects wash out of the court's assessment.

The above axioms are satisfied by a unique one-parameter class of continuous functions.

Theorem 1. The family of functions, indexed by the parameter q, given by $\ell(\pi^P, \pi^D; q) = \{((\pi^P)^q + (\pi^D)^q)/2\}^{1/q}$, $q \in (-\infty, 1]$, $q \neq 0$; and $\ell(\pi^P, \pi^D; 0) = (\pi^P\pi^D)^{.5}$ is the unique family of continuous functions satisfying strict monotonicity, interiority, symmetry, homogeneity, burden of production, and independence of presentation.

The parameter q has a natural interpretation as the breadth of the interpretation of the law relevant to the assessment of liability (such as, to whom does a defendant owe a duty?). A narrow interpretation means that (for any given pair (π^P, π^D)) the court's assessment will be lower than under a broad interpretation. In the Appendix we show that $\partial\ell(\pi^P, \pi^D; q)/\partial q > 0$ for $\pi^P \neq \pi^D$ ($\pi^P, \pi^D > 0$), $q \in (-\infty, 1]$. Thus, low values of q can be interpreted as reflecting a narrow interpretation, while higher values of q reflect progressively broader interpretations. This is why we attach significance to the parameter q: it reflects the breadth of interpretation of an issue of law. As will become clear in Section 4, its domain provides the type space for the incomplete-information model of appeals.

There are both technical interpretations and implications that also contribute to understanding $\ell(\pi^P, \pi^D; q)$ as a model of aggregation. The function $\ell(\pi^P, \pi^D; q)$ is a ''quasi-arithmetic mean,'' sometimes called the mean of order q.[12] In particular, $q = 1$

[11] If D provides no evidence, we assign $\pi^D = 1$ (as the assessment for the weakest possible case for D). Since the burden of production still lies with P, this satisfies $L(\ell(0, 1); \gamma) = 0$, $\forall\gamma > .5$.

[12] Using a different set of axioms in the context of inequality measurement, Foster and Shneyerov (2000) show how the mean of order q again arises. We thank James Foster for bringing this class of functions to our attention and for suggesting homogeneity. On quasi-arithmetic means, see Aczél (1966).

corresponds to the arithmetic mean of π^P and π^D, $q = 0$ yields the geometric mean, $q = -1$ yields the harmonic mean, and $q \to -\infty$ yields the minimum of π^P and π^D. An alternative interpretation of the resulting assessment function is that $\ell(\pi^P, \pi^D; q)$ is a CES production function that is symmetric in the inputs, with the parameter multiplying the inputs equal to ½. Thus, $q = 1$ corresponds to a linear production function, $q = 0$ corresponds to the Cobb-Douglas production function, and $q \to -\infty$ yields the Leontief production function.

Figure 1 illustrates three curves for different values of q in the evidence space Π for the evidentiary standard γ. Observe that γ partitions Π into three sets (these sets are demarcated in the figure by the heavy lines):[13]

(i) The Never Liable set $NL(\gamma) \equiv \{(\pi^P, \pi^D) \in \Pi \mid (\pi^P + \pi^D)/2 < \gamma\}$. In this set the evidence never supports a judgment (that is, a liability determination) against the defendant: $L(\ell(\pi^P, \pi^D; q), \gamma) = 0\ \forall q \in (-\infty, 1]$. In Figure 1, this is the region below the line labelled "$q = 1$."

(ii) The Always Liable set $AL(\gamma) \equiv \{(\pi^P, \pi^D) \in \Pi \mid \min\{\pi^P, \pi^D\} \geq \gamma\}$. Here the evidence always supports a judgment against the defendant:

$$L(\ell(\pi^P, \pi^D; q), \gamma) = 1\ \forall q \in (-\infty, 1].$$

In Figure 1 this is the square region in the upper right of the diagram.

(iii) The Potentially Appealable set

$$PA(\gamma) \equiv \{(\pi^P, \pi^D) \in \Pi \mid (\pi^P + \pi^D)/2 \geq \gamma > \min\{\pi^P, \pi^D\}\}.$$

In this set, the judgment is sensitive to the value of q chosen. This last set comprises the two disjoint triangular regions in Figure 1, since the point (γ, γ) belongs to $AL(\gamma)$.

The set $PA(\gamma)$ will be of greatest interest to us in the next section. Recall that $\partial\ell(\pi^P, \pi^D; q)/\partial q > 0$ for $\pi^P \neq \pi^D$ $(\pi^P, \pi^D > 0)$, $q \in (-\infty, 1]$. This means that for a given value of γ, liability determination is ordered by q: for each $(\pi^P, \pi^D) \in PA(\gamma)$ there exists a $q_{\min}(\pi^P, \pi^D, \gamma)$ such that $\ell(\pi^P, \pi^D; q_{\min}(\pi^P, \pi^D, \gamma)) = \gamma$, and therefore $L(\ell(\pi^P, \pi^D; q), \gamma) = 0\ \forall q < q_{\min}(\pi^P, \pi^D, \gamma)$ and

$$L(\ell(\pi^P, \pi^D; q), \gamma) = 1\ \forall q \geq q_{\min}(\pi^P, \pi^D, \gamma).$$

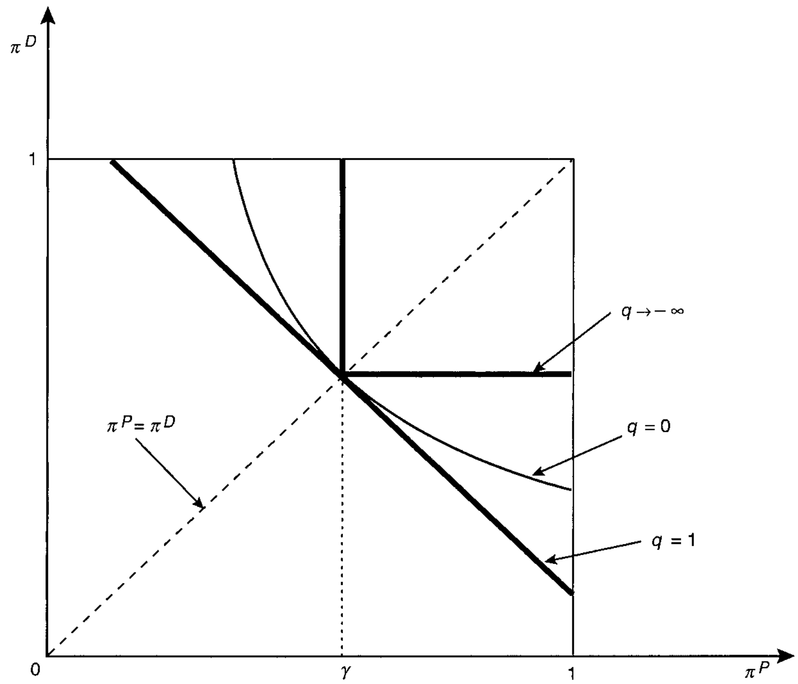
Thus, $q_{\min}(\pi^P, \pi^D, \gamma)$ induces a curve in $PA(\gamma)$ through the point (π^P, π^D). This curve partitions $PA(\gamma)$ into an $L = 1$ portion and an $L = 0$ portion. Moreover, this means that L is nondecreasing in π^P, π^D, and q for fixed γ.

Observe that when π^P is close in value to π^D (for $(\pi^P, \pi^D) \in PA(\gamma)$), the judgment is fairly insensitive to the value of q employed, while when, say, $\pi^P \gg \pi^D$, the judgment is sensitive to the level of q chosen. Alternatively put, when there is little disagreement about the case assessments implied by the submitted evidence, then the same judgment is likely to arise for a broad range of q-values, while substantial disagreement about the assessments (π^P and π^D quite different) leads to different judgments depending upon the q used. Intuitively, it is in the "outer" portions of $PA(\gamma)$ (that is, the (π^P, π^D) in $PA(\gamma)$ that lie furthest from the dotted $\pi^P = \pi^D$ line) where appeals are

[13] The dotted line indicates where $\pi^P = \pi^D$. While one might expect that all cases would involve $\pi^P \geq \pi^D$, this makes no difference to the analysis to follow. In Daughety and Reinganum (forthcoming) we discuss when π^D might exceed π^P in a simultaneous evidence submission model.

FIGURE 1

LIABILITY DETERMINATION IN THE ASSESSMENT SPACE



most likely to arise. Finally, since the parameter q represents the interpretation of law and is the source of the issue for appeal, we postpone discussion of the trial court's choice of q until Section 4.

4. A Bayesian model of appeals

■ Assume that there is a supreme court (that is, a court superior to the trial and appeals courts) and that it knows the correct value of q (denoted q_S) but the trial and appeals courts, and the litigants, do not. In this section we address the determination of q, first at the level of the trial court and then at the level of the appeals court, assuming that both the trial and appeals courts share the same objective of trying to estimate (and use) the supreme court's interpretation (q_S). Since the choice of q is within the discretion of both the trial and appeals courts, it is reasonable to model this aspect of the analysis as Bayesian. The model of the appeals court, which is the main focus of this section, posits a two-stage incomplete-information sequential game between D and the appeals court (denoted A). Both D's attorney[14] and A receive private signals about q_S, which are positively correlated (this will be made more precise later). We will assume that the private signals are identically distributed, conditional on q_S. One could, of course, assume that either A or D had more precise information than the other. This would mean that both would account for this difference in precision when updating their estimates, but the qualitative features of the equilibrium would be the same. Furthermore, it is not obvious which agent (A or D) would have more precise

[14] We ignore any agency problems in this article. Hereafter, we simply refer to D receiving a signal or making a decision.

information: D's attorney is likely to be as qualified, and have access to comparable resources, as A. For example, a judge must be a generalist to some degree, while an appellate attorney can specialize in a particular area of the law. D must decide whether or not to pursue an appeal, and thus he must use his signal to forecast the likely outcome of what will be a costly process.[15] Since this is a costly decision for D, A (who moves second) can improve his estimate of q_S by updating on the fact that D is pursuing the appeal (taking into account D's strategic incentives to appeal and thereby influence A). From A's perspective, D's action is a crude but potentially informative signal; since A also accounts for the costliness of D's signal, low-cost signals are appropriately discounted by A. We characterize A's best response to D's decision to appeal, D's optimal decision anticipating A's response, and the perfect Bayesian equilibrium of the game. The fact that D's decision to appeal is informative means that attributes of D's payoff (for example, the award made at trial and the cost of appeal) now affect the equilibrium probability of affirmance; this is discussed further in Section 5.

While it may seem that we possibly overemphasize the influence of D's decision to appeal, we view this as a simplification of a more complex, but more realistic, model (which is beyond the scope of the current article) wherein arguments about the law made by the litigants are informative to an appeals court, but the court also recognizes the strategic incentive for the litigants to manipulate what they submit. Certainly one would not argue that appeals courts disregard the content of such submissions, nor should one assume that they take such submissions at face value.

First, consider the problem of estimating q_S at the trial court level. This value is unknown to the litigants as well as the trial court, but we assume all three agents have a common prior density over q_S. Specifically, at the trial court level we assume that neither the court nor the litigants have any private information about q_S. This is a plausible assumption for a trial court, since it does not receive substantial resources to support legal research; its best guess about q_S is its unconditional mean. If we assume that the trial court is motivated by accuracy (in the sense of minimizing the mean squared error between its interpretation and q_S), then it will use the unconditional mean, denoted q_T, in its liability assessment and determination. While one could imagine the litigants observing private signals about q_S as a consequence of their own legal research, we will see in Section 5 that this private information will have no impact, in equilibrium, on a court's decision absent a private signal of its own. Thus, both P and D anticipate that the liability determination will be based on the family of assessment functions $\ell(\pi^P, \pi^D; q)$, the trial court's best estimate q_T, and the evidentiary standard. The record of the trial is the tuple $(\pi^P, \pi^D, \gamma, \mathcal{A})$, where the award is $\mathcal{A} > 0$ if $L = 1$ but otherwise zero; as we remarked earlier, we assume that the level of the award is undisputed.

Next, consider the appeals court. As observed earlier, an appeals court customarily takes π^P and π^D as given. Similarly, the appeals court will not choose a different evidentiary standard nor revise the award if it falls within the discretion of the trial court. However, appeals courts consider *de novo* the interpretation of the law, which we have identified with the parameter q. An appellant asserts that the trial court used the ''wrong'' value of q in making a liability assessment.

We consider the case of a defendant who was found liable at trial and is considering appealing this finding.[16] We assume that it is a dominant strategy for a plaintiff who

[15] The filing of a ''notice of appeal'' is cheap, but a number of cases drop by the wayside (''terminations without court action''). D's real costs arise from a full hearing (''terminations on the merits''), our interest here, which Posner indicates is approximately half of those filed (see Posner, 1996).

[16] Plaintiffs who have lost at trial, or feel that the award is insufficient, can also appeal. Thus, a similar analysis could be conducted for plaintiffs; for brevity, we consider here only the case of a defendant appealing a trial court finding of liability.

has won a judgment at trial to respond to an appeal, for several reasons. First, if the likelihood of reversal is sufficiently low or the award is sufficiently high, then the expected value of defending the judgment will exceed the cost of doing so.[17] Second, the plaintiff's attorney, who has typically entered into a partnership agreement with the plaintiff through a contingent fee arrangement, is bound by professional ethics and/or reputational considerations not to abandon a plaintiff who has won a judgment at the trial court. Finally, the court itself may order the judgment defended if it believes there is a socially relevant issue at stake.[18] Thus, there are only two active players in this stage, the defendant (who decides whether or not to appeal the trial court's decision) and the appeals court (which decides whether to affirm or reverse the trial court's decision).

Appeals courts have considerably more discretion in interpreting the law than does a trial court in interpreting the facts. In particular, appeals courts do not need to respect axioms such as interiority or reflexivity. Moreover, appeals courts are not restricted to make their decisions about a given legal issue based only on the cases and statutes raised by the litigants. This is because appeals courts have access to the same space of cases, laws, and rules as the litigants do, in contrast with the trial court setting wherein the court knows only those facts submitted by the litigants. As expressed recently (see *In re Rhone-Poulenc Rorer Inc.*, at 1299),

> The doctrine is not that if a party fails to offer a particular reason for its position, the court cannot consider that reason. . . . Were that the rule, the role of an appellate court would be confined to weighing the reasons, pro and con a particular ground, that the parties happened to proffer. Appellate consideration so truncated could not produce durable rules to guide decision in future cases; judicial opinions would be impoverished if all they did were call balls and strikes.

When an appeals court proceeds to construct reasons for its decision other than those provided by the litigants, it is forming its own estimate of q_S, which may differ from that of the trial court and those of the litigants.

To model the appeals court as a Bayesian decision maker, we assume that the appeals court and the appellant share a common prior regarding q_S. This prior may differ from that used by the parties and the court at the trial stage, if only because the actors at this level (litigators and appeals court judges and their clerks) specialize in reviewing matters of law. Moreover, we assume that the appeals court receives a private signal regarding q_S. This signal reflects the availability of resources to do specialized research in the relevant aspects of the law for the case at hand and any evolution of case law that has occurred between the trial and the appeal. We assume that D also receives a private signal regarding q_S and uses it to estimate the chances of winning on appeal.

Specifically, let q_i, $i = A, D$ represent the private signals received by A and D, respectively. Since we intend to use q_i to represent agent i's "type," we temporarily restrict q_i, $i = S, A, D$, to the compact interval $[p, 1]$, where $p > -\infty$; we will characterize equilibrium in the model for arbitrary finite p, and then allow $p \to -\infty$. Let the prior density on q_S be $h_S(q_S)$, a strictly positive and continuously differentiable function on $[p, 1]$. We assume that q_A and q_D are independent conditional on q_S. Further, let the conditional density of q_i be denoted $g(q_i | q_S)$, $i = A, D$.

[17] Since 1992, in approximately 90% of civil appeals, the lower court's decision was affirmed (Posner, 1996). The cost of defending a judgment may be low because the plaintiff can simply say "we believe the trial court was right," whereas the appellant defendant must argue why the trial court was wrong.

[18] An example of this was the U.S. Supreme Court's appointment of William T. Coleman to argue the government's position when the United States declined to support its own case in *Goldsboro Christian Schools, Inc.* v. *U.S.* and *Bob Jones University* v. *U.S.* in 1982.

Thus, the joint density for (q_S, q_A, q_D) is given by

$$f(q_S, q_A, q_D) = h_S(q_S)g(q_A|q_S)g(q_D|q_S).$$

We assume that all these density functions are strictly positive and continuously differentiable for $q_i \in [p, 1]$, $i = S, A, D$. Finally, let $h_i(\cdot)$, $i = A, D$, be the marginal densities for A and D, respectively. Note that $h_A(q) = h_D(q)$, $\forall q$ (that is, the appeals court and the appellant have the same marginal densities since the conditional distribution g is unsubscripted); we suppress the marginal density's subscript when $i = A$ or D. Finally, the prior density and the conditional densities are common knowledge to A and D.

Further, assume the densities $g(q_i|q_S)$ satisfy the monotone likelihood ratio property.

Assumption 1.

$$\forall q_i' > q_i \quad \text{and} \quad q_S' > q_S, \qquad [g(q_i|q_S)/g(q_i|q_S')] \geq [g(q_i'|q_S)/g(q_i'|q_S')], \qquad i = A, D.$$

Loosely, if i receives a high signal q_i, he should infer a high value for q_S. From Milgrom and Weber (1982), it follows that q_A, q_D, and q_S (or any subset thereof) are ''affiliated'' random variables. The practical implications of affiliation are that D can use his observation of q_D to obtain a posterior distribution regarding A's signal q_A, and A can use D's observable behavior (i.e., the decision to appeal) to update its distribution over q_S. Note that because D's decision to appeal is a crude signal of q_D, it will necessarily be less influential than A's own private signal, q_A. The value of affiliation is that certain conditional expectations of importance to this analysis have specific monotonicity properties that we will employ. However, affiliation itself is a somewhat weak condition (for instance, independent random variables are affiliated), so we will impose additional strengthening assumptions as needed.

The timing of the decisions is as follows. D observes his private signal and decides whether or not to pursue an appeal, based upon his posterior beliefs about q_A given q_D and his expectation about how A will respond to his decision to appeal. In the sequel we will assume that the possibility of further appeal does not influence D's objective function. This is because, independent of whether D wins or loses this appeal, his likelihood of any further hearing on his case is very small.[19] If D appeals, P automatically responds. Finally, A observes its private signal and decides whether to affirm or reverse the trial court's decision, based upon its posterior beliefs about q_S, given q_A and the fact that D appealed. In making this decision, we assume that A wants its interpretation of the law to conform as closely as possible to that of the supreme court were it to review the same case. That is, the appeals court simply wants to interpret the law correctly.

Thus A calculates its best estimate of q_S (given q_A and the fact that D appealed) and substitutes this value into the liability determination function $L(\ell(\pi^P, \pi^D; q), \gamma)$ to determine whether to affirm ($L = 1$) or reverse ($L = 0$) the decision. Note that we assume the appeals court uses the same machinery (liability assessment rule and liability determination rule) as did the trial court, but with a possibly different value of q. As former Chief Judge of the First U.S. Circuit Court of Appeals Frank M. Coffin

[19] Posner (1996) indicates that in the federal system, only 1.47% of cases that were appealed to the U.S. Supreme Court in 1993 were granted review (appeals of a first-level appeals court's decisions are generally heard at the discretion of the superior court).

(1994, p. 84) describes it, ‘‘our appellate courts step into the shoes of the trial judge and view the facts and issues as they were presented to him.’’

We employ perfect Bayesian equilibrium and thus we begin the analysis with A's decision, given that D has appealed. First, what should A infer from the fact that D has appealed? Since $\partial\ell(\pi^P, \pi^D; q)/\partial q > 0$ for $(\pi^P, \pi^D) \in PA(\gamma)$, anticipating a higher value of q (due to having observed a higher level of q_D) should lead D to expect a lower likelihood of success on appeal. Thus, if a D of type q_D would appeal, so would any D with a lower private signal (assume that D appeals when indifferent). Hence, a plausible inference on the part of A from the observation that D has appealed is that $q_D \in [p, t]$, for some $t \leq 1$. This value t cannot depend on q_D, which is unobservable to A, and it cannot depend on q_A because it is a conjecture about behavior by D, and D cannot observe q_A.

Let $\mu_A(q_S | q_A, q_D \in [p, t])$ denote A's posterior density function over q_S, given q_A and A's belief that $q_D \in [p, t]$. This posterior is

$$\mu_A(q_S | q_A, q_D \in [p, t]) = \int_p^t f(q_S, q_A, q_D)\, dq_D \Bigg/ \int_p^t \int f(q, q_A, r)\, dq\, dr$$

(where no limits of integration are indicated here and below, the integral is over the interval $[p, 1]$).

Formally, A's problem is to find

$$BR_A(t; q_A) \equiv \underset{q}{\operatorname{argmin}} \int (q - q_S)^2 \mu_A(q_S | q_A, q_D \in [p, t])\, dq_S,$$

which is readily seen to be $\mathrm{E}(q_S | q_A, q_D \in [p, t])$, A's posterior expected value of q_S. By Milgrom and Weber's Theorem 5 (1982), affiliation implies that this expectation (which provides A's best response to D's decision to appeal) is a nondecreasing function of t and q_A. Moreover, it can be shown that $BR_A(t; q_A)$ is continuous on $[p, 1]$ and is continuously differentiable in t on $(p, 1]$ and in q_A on $[p, 1]$.[20] We add the further assumption that the function $BR_A(t; q_A) = \mathrm{E}(q_S | q_A, q_D \in [p, t])$ has strictly positive first derivatives.

Assumption 2.

$$\partial \mathrm{E}(q_S | q_A, q_D \in [p, t])/\partial t > 0 \quad \text{and} \quad \partial \mathrm{E}(q_S | q_A, q_D \in [p, t])/\partial q_A > 0 \quad \text{for all } t, q_A.$$

Assumption 2 says that A's (posterior) expected value of the supreme court's q is increasing in two things: (1) A's belief about the maximal type of D who would appeal and (2) A's private signal. To determine whether to reverse or affirm,[21] A computes $L^*(t; q_A) \equiv L(\ell(\pi^P, \pi^D; BR_A(t; q_A)), \gamma)$; that is, A compares $\ell(\pi^P, \pi^D; BR_A(t; q_A))$ to the evidentiary standard γ and affirms if $\ell(\pi^P, \pi^D; BR_A(t; q_A)) \geq \gamma$. Recall that since $(\pi^P, \pi^D) \in PA(\gamma)$, there is a $q_{\min}$ defined by $\ell(\pi^P, \pi^D; q_{\min}) = \gamma$. Then, A's decision rule can alternatively be cast as comparing $BR_A(t; q_A)$ to $q_{\min}$ and affirming when $BR_A(t; q_A) \geq q_{\min}$. This yields A's decision rule, which can be expressed as $L^*(t; q_A) = 0$ if $BR_A(t; q_A) < q_{\min}$ and $L^*(t; q_A) = 1$ if $BR_A(t; q_A) \geq q_{\min}$.

Let $q^*(t) = \min\{q_A \in [p, 1] | BR_A(t; q_A) \geq q_{\min}\}$ when this set is nonempty; otherwise, let $q^*(t) = 1$. The expression $q^*(t)$ is the minimum value of q_A necessary

[20] The proofs here are straightforward and are available from the authors upon request.

[21] Since (π^P, π^D, γ) is fixed, we abuse notation and suppress these terms in what follows.

for A to affirm the trial court's decision, given the belief that $q_D \leq t$. The function $q^*(t)$ is continuous on $[p, 1]$ and is illustrated in Figure 2, which indicates A's decision for arbitrary values of A's belief t. Whenever $q^*(t) \in (p, 1)$, it is defined implicitly by $BR_A(t; q^*(t)) = q_{\min}$; since BR_A has continuous positive first derivatives, the implicit function $q^*(t)$ is continuously differentiable with $q^{*\prime}(t) < 0$. This means that the more convinced A is that D observed a low signal (that is, the lower is t), the higher is the signal q_A (that is, $q^*(t)$) necessary to cause A to affirm the trial court's decision.

Now consider D's decision to appeal. Although D doesn't know A's private signal q_A, D has his own private signal q_D. Since q_A and q_D are affiliated, D can form posterior beliefs about q_A given q_D, denoted $\mu_D(q_A | q_D)$. This density is given by

$$\mu_D(q_A | q_D) = \int [f(q_S, q_A, q_D)/h(q_D)]\, dq_S.$$

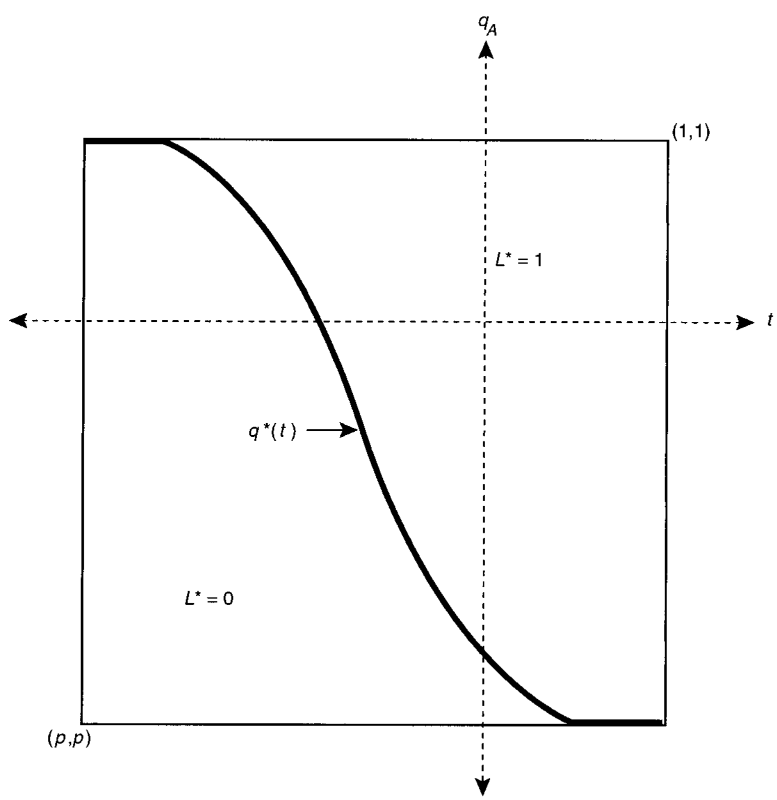
Let the cumulative distribution of q_A given q_D be denoted $\mathrm{M}_D(q_A | q_D)$, where

$$\mathrm{M}_D(q_A | q_D) = \int_Q \mu_D(q | q_D)\, dq,$$

with $Q = [p, q_A]$. Observe that $\mathrm{M}_D(q_A | q_D)$ is continuously differentiable in q_D and

FIGURE 2

APPEALS COURT DECISION TO AFFIRM OR REVERSE



affiliation of q_A and q_D implies first-order stochastic dominance (by Milgrom and Weber (1982) Theorem 5); that is, $\mathrm{M}_D(q_A|q_D)$ is nonincreasing in q_D for all $q_A \in [p, 1]$. We add a further strengthening assumption:

Assumption 3. $\partial \mathrm{M}_D(q_A|q_D)/\partial q_D < 0$ for all $q_A \in (p, 1)$.

If D chooses not to appeal, he acquiesces to the trial court's decision and pays the award $\mathcal{A}$. If D chooses to appeal, he pays the cost of appeal, denoted C ($C < \mathcal{A}$), and anticipates (for arbitrary beliefs t by A) a payment to the plaintiff of

$$\mathcal{A} \int L^*(t; q_A)\mu_D(q_A|q_D)\, dq_A.$$

Thus, D will appeal if and only if $V(t; q_D) \equiv \mathcal{A} - \mathcal{A} \int L^*(t; q_A)\mu_D(q_A|q_D)\, dq_A - C \geq 0$. Since L^* is zero over the interval $[p, q^*(t))$ and one over the interval $R(t) \equiv [q^*(t), 1]$, this means that D will appeal if and only if

$$V(t; q_D) = \mathcal{A} - \mathcal{A} \int_{R(t)} \mu_D(q_A|q_D)\, dq_A - C = \mathcal{A} - \mathcal{A}[1 - \mathrm{M}_D(q^*(t)|q_D)] - C \geq 0,$$

which characterizes D's appeal behavior as a function of an arbitrary belief t on the part of A. Again, Milgrom and Weber (1982) imply $V(t; q_D)$ is a nonincreasing function of q_D. Since $q^*(t)$ is continuous in t and $M_D(\cdot|q_D)$ is continuous in q_D, it follows that $V(t; q_D)$ is continuous in both t and q_D. Finally, notice that since $q^*(t)$ is nonincreasing in t, $V(t; q_D)$ is also nonincreasing in t. To summarize, the function $V(t; q_D)$ is continuous and nonincreasing in both its arguments.

A perfect Bayesian equilibrium is characterized by beliefs t^* and a marginal appellant type q_D^* such that (1) $V(t^*; q_D) \geq 0$ for all $q_D \leq q_D^*$, and $V(t^*; q_D) < 0$ for all $q_D > q_D^*$, and (2) $t^* = q_D^*$. That is, first, if A believes that the set of types that appealed is $[p, t^*]$, then the actual set of types that would appeal is $[p, q_D^*]$; second, A's beliefs are consistent with D's equilibrium decision, $t^* = q_D^*$. This leads to the following existence and uniqueness result.

Theorem 2. (i) If $V(p; p) < 0$, it is optimal for all D types not to appeal, regardless of A's beliefs.

(ii) If $V(1; 1) > 0$, it is optimal for all D types to appeal, regardless of A's beliefs.

(iii) If $V(p; p) \geq 0$, and $V(1; 1) \leq 0$, there exists $q_D^* \in [p, 1]$ such that $V(q_D^*; q_D^*) = 0$; q_D^* is unique and satisfies the equilibrium conditions.

Thus, the equilibrium typically involves some intermediate value of q_D^* that acts as a cutoff: D finds it worthwhile to appeal only if his signal q_D does not exceed this value. An appealed case provides this level of crude information to A about D's observation. Note that the role of p in the analysis allows us to pass to the limit ($p \to -\infty$), meaning that Theorem 2 actually characterizes the equilibria for the original type space $(-\infty, 1]$ with minor modifications.[22]

5. Implications of the equilibrium

■ **Comparative statics.** There are three parameters of interest: $\mathcal{A}$, C, and γ; we consider only interior equilibria (that is, $q_D^* \in (-\infty, 1)$). To analyze the impact of

[22] If $\lim_{p \to -\infty} V(p; p) \leq 0$, then no appeals occur (as in (i) in the theorem). If this limit is positive and $V(1; 1) \leq 0$, then there is a unique equilibrium $q_D^* \in (-\infty, 1]$ (as in (iii) in the theorem).

changes in $\mathcal{A}$ and C we use the equilibrium condition that $V(q_D^*; q_D^*) = 0$, which means that

$$\int L^*(q_D^*; q_A)\mu_D(q_A | q_D^*)\, dq_A = (\mathcal{A} - C)/\mathcal{A}.$$

On the left is the probability that the marginal case is affirmed. Since, at q_D^*, $V(z; z)$ is continuously differentiable in z, with $dV/dz < 0$ (as shown in the proof of Theorem 2 in the Appendix), an increase in C implies that q_D^* falls. Thus, some cases that previously would have been appealed will no longer be appealed. Similarly, if the award $\mathcal{A}$ increases then q_D^* increases, increasing the set of possible q_D values that will result in an appeal. Finally, recall that $q_{\min}$ depends upon π^P, π^D, and γ; an increase in the evidentiary standard γ raises $q_{\min}(\pi^P, \pi^D, \gamma)$. This holds because

$$\ell(\pi^P, \pi^D; q_{\min}(\pi^P, \pi^D, \gamma)) = \gamma \quad \text{and} \quad \partial\ell/\partial q > 0,$$

yielding $dq_{\min}(\pi^P, \pi^D, \gamma)/d\gamma > 0$. Note that this means a stronger case against D would be required to find D liable. This further implies that since $BR_A(t; q^*(t)) = q_{\min}(\pi^P, \pi^D, \gamma)$, the value $q^*(t)$ must also increase, which means that $\mathcal{A} \int_{R(t)} \mu_D(q_A | q_D)\, dq_A$ decreases (recall that $R(t) = [q^*(t), 1]$). This, in turn, implies that V increases, forcing q_D^* to increase to reestablish equilibrium. Thus, an increase in the evidentiary standard leads to a higher likelihood of appeal.

The probability that the trial court's judgment is affirmed is also affected by changes in $\mathcal{A}$, C, and γ. This probability is given by $\int L^*(q_D^*; q_A)\mu_D(q_A | q_D^*)\, dq_A = (\mathcal{A} - C)/\mathcal{A}$ for the marginal type q_D^* and by $\int L^*(q_D^*; q_A)\mu_D(q_A | q_D)\, dq_A$ for an inframarginal type q_D. It is clear that an increase in the award increases the probability that the marginal type q_D^* loses on appeal (that is, the probability increases that the judgment against q_D^* is affirmed). Moreover, an increase in C has the opposite effect, and an increase in γ has no effect, on this probability.

A more subtle result is that the likelihood that an *inframarginal* type q_D loses on appeal is also increased by an increase in the award because $\int L^*(q_D^*; q_A)\mu_D(q_A | q_D)\, dq_A$ is an increasing function of q_D^* (see the proof of Theorem 2 in the Appendix). This represents a negative externality imposed on q_D by the marginal type q_D^*. As the award increases, more D types (with weaker cases) appeal, and thus A draws a less favorable inference from the observation that D has appealed and is more likely to affirm the trial court's judgment. This adversely affects all D types, not just the marginal one, even though only the marginal type's behavior is affected by the increase in the award. The analysis of an increase in C is similar, with the implication that an increase in C results in fewer appeals (some weaker cases having been screened out by the higher cost of appeal), and thus A draws a more favorable inference from the observation that D has appealed and is less likely to affirm the trial court's judgment. This benefit accrues to all D types, not just the marginal one, though again only the marginal type's behavior is affected. The effect of increasing γ on the probability that the inframarginal case is affirmed is ambiguous; the direct effect of increasing γ is to decrease this probability, but the indirect effect (through increasing q_D^*) is to increase it.

Note that these comparative statics exercises hold the set of cases going to trial, and found liable at trial, constant. The overall impact on the caseload for the appeals court also depends upon how an increase in the evidentiary standard affects pretrial settlement bargaining and incentives for care, as well as how many cases are decided against D at trial. We have fixed $(\pi^P, \pi^D) \in PA(\gamma)$, so these other factors may influence the flow of cases to and through the trial court. It should be clear that the earlier statements are also partial equilibrium in the sense that changes in the award $\mathcal{A}$ and the cost C may affect the likelihood of harm (perhaps greater or lesser care will be taken) and the likelihood of cases settling via pretrial bargaining (since the threat of going to trial has changed).

□ **Harmless error.** An appeals court fails to reverse a trial court's decision if it finds that errors at the trial level provide insufficient reason for "granting a new trial, or for setting aside a verdict or for vacating, modifying or otherwise disturbing a judgment or order . . ." (Federal Rule of Civil Procedure 61, in Yeazell (1996)). This rule appears to promote a degree of collegial deference, in this case deference to the trial court on the value of q to use. This rule, however, and the decisions that are based on it, can simply reflect rational behavior. In our model, after observing q_A and the appeal, the appeals court computes its estimate of q_S, namely $BR_A(q_D^*; q_A)$. Clearly, the following possibilities can occur. One possibility is that $BR_A(q_D^*; q_A) < q_{\min}$, in which case the trial court has committed reversible error and the appeals court reverses the decision (possibly remanding for retrial). The remaining two alternatives[23] are that (i) $q_{\min} \leq BR_A(q_D^*; q_A) < q_T$ or (ii) $q_T \leq BR_A(q_D^*; q_A)$. In case (ii), the trial court's interpretation was too narrow, but the outcome was correct, so while there may be an impact on future cases, there is no impact on the case at hand. In case (i), the appeals court finds the trial court's interpretation too broad, but not so broad as to have caused an incorrect outcome. It is this case that constitutes a harmless error. This comports well with what Posner (quoted in Section 2) indicates as the degree of deference on issues of law afforded a trial court by an appeals court (i.e., none). What might appear as deference with respect to reversing a verdict based on interpretation of the law is simply rational behavior by a Bayesian decision maker: the actual value of q_T is irrelevant; the issue concerns how $q_{\min}$ is related to $BR_A(q_D^*; q_A)$.

□ **Changes in appeals court resources.** We argued earlier that the appeals court has more resources to devote to research on the law than a trial court, making its signal of q_S more informative than the trial court's; this is why it should be able to overturn lower court decisions if it so chooses. By resources we mean that appeals courts have enough time and support personnel (such as law clerks) to thoroughly consider an issue. We examine the impact of a substantial reduction in these resources by examining an extreme case: what if it were common knowledge that q_A was uninformative and that A therefore must rely solely on the prior on q_S and the information inferred from D's decision to appeal? The resulting equilibrium involves pure reliance on the prior, as D's action becomes uninformative.

To see this, reconsider A's problem when q_A is uninformative. In this case $BR_A(t) = \mathrm{E}(q_S | q_D \in (-\infty, t])$, which means that $dBR_A(t)/dt \geq 0$ (by Milgrom and Roberts, 1982). Now A computes $L^*(t) \equiv L(\ell(\pi^P, \pi^D; BR_A(t)); \gamma)$; note that L^* is nondecreasing in t. D will choose to appeal if $\mathcal{A}L^*(t) + C \leq \mathcal{A}$. Thus, for all q_D, either this inequality is satisfied (because $L^* = 0$) or it is not (because $L^* = 1$). Hence, either all types appeal or none do: the decision to appeal is therefore uninformative and thus

[23] While we don't assume that any opinion written by the trial court necessarily conveys its estimated value of q, the appeals court can reconstruct the trial court's prior and compute q_T.

$BR_A(t) = E(q_S)$ for all t.[24] Therefore, if $E(q_S) < q_{\min}(\pi^P, \pi^D, \gamma)$, then all q_D-types appeal; otherwise no type appeals.[25]

Thus, resources devoted to the appeals court for research purposes have a double impact: if the appeals court has its own independent signal of q_S, it can also benefit from extracting further information from *D*'s decision to appeal. If it does not have its own informative signal, then it cannot learn anything from *D*'s decision to appeal, either. In this case, there is no reason for *D* to invest in obtaining a private signal. Furthermore, without an informative q_A, *A*'s caseload is totally dependent upon the trial court's estimate of q_S: overly ''tough'' trial courts will send many cases to appeal, including many poor cases. Contrast this with the previous analysis wherein q_A was informative: this induced defendants with poor cases to refrain from appeal.

6. Summary

■ This article is about appealing judgments. We have used the word ''appealing'' in two ways: as an adjective and as a verb. We used it as an adjective to address a trial court's assessment process by which it receives evidence and generates a judgment; we argued that rules of evidence and procedure act to enforce notions of responsiveness and procedural fairness in trial activities. We parsed the overall trial process into parts that would best be modelled as Bayesian (decisions about the credibility of evidence or about a yet higher court's likely breadth of interpretation of an issue of law) and those wherein a Bayesian model would generally violate the goals of responsiveness and procedural fairness. We focused on the second part in Section 3 by employing axiomatic methods to model the aggregation of credible evidence submitted by the litigants into the court's assessment, which could then be compared with an evidentiary standard so as to reach a (binary) decision as to whether a defendant should be held liable or not.

In Section 4 we used ''appealing'' in a second way, as a verb. Here we described a Bayesian model of an appeals court, whose primary job is to consider judgments made in a trial court that a litigant (we focused on the losing defendant) found sufficiently worthwhile to appeal. In the case of a trial court, the space of facts was largely unobservable to the court itself; the only observable portion was that brought by one or the other of the litigants. In the case of the appeals court, the space of past cases and existing statutory law is known to all parties, though how a yet superior court (e.g., a supreme court) would apply that law to the issue at hand is not known. Appeals courts have resources at their disposal to consider such issues, as do those who practice before them. More significantly, it is only because appeals courts have such resources that the action by a private party to bear the cost of appeal (and the risk of losing an appeal) makes such private actions potentially informative to the court. An informed appeals court (i.e., one that is imperfectly but privately signalled about the likely interpretation of the supreme court) can further improve its decision by accounting for the informational content of the (privately informed) litigant's action. An uninformed appeals court recognizes the strategic feedback that its lack of private information triggers: now the action of the privately informed litigant is content-free, reducing the incentive for private litigants to invest in information about the likely interpretation of the supreme court.

---

[24] Recall that $E(q_S)$ need not be the same as q_T, due to the appeals court's greater general knowledge of, and experience in, deciding issues of law.

[25] If $E(q_S) \geq q_{\min}(\pi^P, \pi^D, \gamma)$, an appeal by *D* is an out-of-equilibrium action that doesn't change *A*'s decision, for if it did, all q_D-types would defect and appeal.

With a privately informed appeals court, the equilibrium involved a cutoff value of the defendant's signal: only those with signals at or below the cutoff appealed. Higher awards at trial increase the cutoff, while higher costs of appeal decrease it. An increased evidentiary standard also increases the cutoff, resulting in an increased probability of appeal. These changes affect not only the marginal type, but the inframarginal types, too. In the case of the award and the cost of appeal, the impact of a change on the likelihood of the inframarginal type's losing its appeal is the same as that for the marginal type. For the evidentiary standard, the impact on the inframarginal type's likelihood of losing its appeal is ambiguous.

Finally, the combination of the axiomatic model of evidence aggregation at the trial level with the Bayesian model of legal interpretation generally led to local stability of judicial decisions, and in particular allowed for a notion of "harmless error" (error made by a trial court and noted by the appeals court that, nonetheless, should not lead to a reversal). Even if the appeals court's estimate of q were lower than that used by the trial court, it might not fall below $q_{\min}$. Thus, small differences in interpretation of the law are unlikely to lead to reversals, providing a type of stability of decision sometimes thought to reflect collegial deference, but here seen simply as rational decision making.

Appendix

■ This Appendix provides proofs of claims made in the main text as well as supplementary material, such as the discussion of relaxing the symmetry axiom in Theorem 1.

Claim A1. For the class of Bayesian inference models wherein the joint distribution over the true liability (denoted ρ) and the evidence (π^P, π^D) is strictly positive on $[0, 1] \times \Pi$, reflexivity cannot hold.

Proof. To see this formally in the current context, let $\sigma(\rho, \pi^P, \pi^D)$ be the subjective joint density of the true liability and the submitted evidence pair, in a Bayesian model a decision maker wants to infer ρ from (π^P, π^D). Assume that $\sigma(\cdot, \cdot, \cdot) > 0$ on its domain $([0, 1] \times \Pi)$; this is a fairly typical assumption in such models. Using π^P and π^D, a Bayesian decision maker would estimate ρ by computing $\int u\sigma(u, \pi^P, \pi^D)\, du / \int \sigma(v, \pi^P, \pi^D)\, dv$, that is, by forming the posterior density of ρ given (π^P, π^D) (namely, $\sigma(\rho, \pi^P, \pi^D)/\int \sigma(v, \pi^P, \pi^D)\, dv$) and then finding the expected value (all integrals are taken over [0, 1]). Reflexivity in this case is the requirement that $z = \int u\sigma(u, z, z)\, du / \int \sigma(v, z, z)\, dv$ for all z in [0, 1]. This is equivalent to solving $\int (z - u)\sigma(u, z, z)\, du = 0$. Note that, for $z = 0$ (or $z = 1$), this would imply that $\sigma(\rho, z, z)$ is zero for almost all values of ρ, which contradicts the assumption that $\sigma(\cdot, \cdot, \cdot)$ is strictly positive on its domain. *Q.E.D.*

In other words, to have a Bayesian inference model of the assessment process that satisfies reflexivity, a very special prior distribution would be needed. In this view of evidence as the outcome of strategic search, evidence is both statistically and strategically related to true liability. In models in which evidence is not statistically, but only strategically, related to type (as in signalling or mechanism design models of hidden information), it may be possible to infer type precisely from evidence submissions. These also fail to be reflexive: if signalling involves distortion, then submissions are not truthful, while in a mechanism designed to induce truth-telling, the distortions are built into the mechanism itself.

Proof of Theorem 1. For convenience in reading the proof of Theorem 1, we use the following mnemonic devices: *smono.* = strict monotonicity, *int.* = interiority, *sym.* = symmetry, *hom.* = homogeneity, *bp.* = burden of production, and *ind.* = independence of presentation. The proof applies results from the analysis of functional equations (see Aczél, 1966).

First, *sym.* and *ind.* imply a condition called "bisymmetry," wherein $\ell(\ell(u, w), \ell(v, z)) = \ell(\ell(u, v), \ell(w, z))$; note that here v and w were switched. This holds because, requiring *ind.*, $\ell(\ell(u, w), \ell(v, z)) = \ell(\ell(u, z), \ell(v, w))$, is (via *sym.*) equivalent to requiring $\ell(\ell(u, w), \ell(v, z)) = \ell(\ell(u, z), \ell(w, v))$ (note the switch of w and v). Applying *ind.* again to this yields $\ell(\ell(u, w), \ell(v, z)) = \ell(\ell(u, v), \ell(w, z))$, which is the bisymmetry property. Aczél shows that bisymmetry, along with reflexivity ($\ell(x, x) = x$), *int.*, *smono.* and *sym.* implies that there exists a continuous, strictly monotonic function $k(x)$, k: $[0, 1] \rightarrow [0, 1]$, such that (see Aczél, 1966): $\ell(x, y) = k^{-1}((k(x) + k(y))/2)$. This means that ℓ is a *quasi-arithmetic mean.* Quasi-arithmetic means that satisfy *hom.* have k-functions with a specific form (see Aczél, 1966):

$$k(x) = \begin{cases} ax^q + b & a \neq 0,\ b,\ q \neq 0, \quad \text{all real} \\ a \ln(x) + b & a \neq 0,\ b, \quad \text{both real.} \end{cases}$$

Substituting this into the previous expression for $\ell(x, y)$ and simplifying yields

$$\ell(x, y; q) = \begin{cases} \{(x^q + y^q)/2\}^{1/q} & q \neq 0 \\ (xy)^{.5} & q = 0. \end{cases}$$

Finally, *bp.* implies that $q \leq 1$. To see this, observe that $L(\ell(0, y); \gamma) = 0$, $\forall y \in [0, 1]$, $\forall \gamma > .5$, if and only if $\ell(0, y; q) \leq .5$ $\forall y \in [0, 1]$. Using the result immediately above, this clearly holds strictly for $q \in [0, 1)$ and just holds for $q = 1$ when $y = 1$. Since ℓ is monotonic in q (see below), we therefore know that $q > 1$ is ruled out. Finally, for $q < 0$, consider $\lim_{x\to 0} \ell(x, y; -r)$, with $r = -q > 0$. This can be written as $\lim_{x\to 0} (2/(1/x^r + 1/y^r))^{1/r}$, which is zero for all y, so *bp.* implies $q \leq 1$. *Q.E.D.*

□ **Discussion of relaxing symmetry.** If the symmetry condition is relaxed, then the independence-of-presentation property must be modified slightly to provide a property called *bisymmetry*; this was the property used in the proof above and was implied by symmetry and independence of presentation. Thus, here we drop symmetry and modify the presentation property to the following:

$$\forall u, v, w, z,\ \ell(\ell(u, w), \ell(v, z)) = \ell(\ell(u, v), \ell(w, z)). \qquad \text{(Bisymmetry)}$$

This axiom can be interpreted to mean that aggregation of evidence *across* sublists (as shown on the left) yields the same assessment as aggregation of evidence *within* lists, followed by aggregation over the two aggregates. Thus, for example, on the right $\ell(u, v)$ is the application of the assessment function to the two sublist values u and v associated with P's evidence (that is, $x = \ell(u, v)$); similarly, $\ell(w, z)$ applies the assessment function to D's sublists (that is, $y = \ell(w, z)$).

With these changes, it can be shown that there is a unique *two-parameter* family of continuous functions that satisfy strict monotonicity, interiority, homogeneity, burden of production, and bisymmetry (the proof is parallel to that shown above and employs Corollary 6 in Aczél and Dhombres (1989), which allows for nonsymmetric functions). Note that the domain of q in the theorem is $(-\infty, 0]$, not $(-\infty, 1]$ as in Theorem 1 earlier. This reflects the deletion of the symmetry property. If we substitute complementarity ($\ell_{xy} > 0$, which Shepherd (1999) finds empirically) for burden of production, then the original domain is restored.

Theorem A1. The family of functions, indexed by the parameters λ and q, given by

$$\ell(x, y; q, \lambda) = \{\lambda x^q + (1 - \lambda)y^q\}^{1/q}, \qquad q \in (-\infty, 0),\ q \neq 0,\ \lambda \in (0, 1), \quad \text{and} \quad \ell(x, y; 0, \lambda) = x^{\lambda}y^{1-\lambda},$$

is the unique family of continuous functions satisfying strict monotonicity, interiority, homogeneity, burden of production, and bisymmetry.

The proof below concerning the effect of q on $\ell(x, y; q)$ is readily extended to this family (that is, $\partial\ell(x, y; q, \lambda)/\partial q > 0$ for $x \neq y$ $(x, y > 0)$, $q \in (-\infty, 0]$, $\lambda \in (0, 1)$). Moreover, it can be shown that $\partial\ell(x, y; q, \lambda)/\partial\lambda$ $(>, <)$ 0 as x $(>, <)$ y. Thus, if x is greater than y, increasing the weight associated with x increases the court's assessment of D's liability, while if x is less than y, increasing this weight lessens the assessment.

A natural interpretation of λ is that it represents anticipated bias (ideological, psychological, etc.) of the judge or jury toward one or the other of the litigants. It can be readily shown that $\ell(\pi^P, \pi^D; q, \lambda)$ is convex in λ. If P and D are uncertain about the precise value of this bias, then a prior density on the possible values that λ might take on would be employed to form the expected assessment, $\mathrm{E}(\ell(\pi^P, \pi^D; q, \lambda))$; convexity implies that this equals or exceeds $\ell(\pi^P, \pi^D; q, \mathrm{E}(\lambda))$. The expected value of $\ell(\pi^P, \pi^D; q, \lambda)$ with respect to λ is relevant since the assessment of the likelihood of liability typically enters in a linear fashion into payoff functions for pretrial decisions (for example, to settle a case or to acquire evidence). Now observe that $\ell(\pi^P, \pi^D; q) = \ell(\pi^P, \pi^D; q, .5)$. Thus, for the same evidence pair, anticipated value of q, and evidentiary standard, uncertainty about asymmetric treatment of the litigants' evidence (holding $\mathrm{E}(\lambda) = .5$ and assuming a common prior distribution for λ), raises the parties' *ex ante* expectation of the court's assessment. In the discussion of appeals either $\ell(\pi^P, \pi^D; q,)$ or $\ell(\pi^P, \pi^D; q, \lambda)$ could be employed; for simplicity we will maintain symmetry and thus employ $\ell(\pi^P, \pi^D; q,)$ in our analysis.

Claim A2. $\ell(x, y; q)$ is increasing in q for $x, y > 0$, $x \neq y$, $q \in (-\infty, 1]$.

Proof. (i) First, assume $q \neq 0$. Using the functional form for $\ell(x, y; q)$ from Theorem 1 in Section 3,

$$\partial\ell(x, y; q)/\partial q = (1/q)^2((x^q + y^q)/2)^{1/q}[q(x^q \ln x + y^q \ln y)/(x^q + y^q) - \ln((x^q + y^q)/2)].$$

Let $w = x^q$ and $z = y^q$. Then the above derivative is positive if and only if

$$w \ln w + z \ln z > (w + z)\ln((w + z)/2) \quad \text{for } w \text{ and } z > 0, \qquad w \neq z,$$

which holds (see Hardy, Littlewood, and Pólya, 1952). (ii) Next, $\lim_{q\to 0} \partial\ell(x, y; q)/\partial q = (xy)^{.5}(\ln x - \ln y)^2/8 > 0$ for $x \neq y$, $x, y > 0$. Note that $\partial\ell(x, y; q)/\partial q = 0$ when $x = y$ $(x, y > 0)$. *Q.E.D.*

Proof of Theorem 2. We show below that there is a unique equilibrium for the model, by considering three mutually exclusive and exhaustive possibilities.

Case (i) $V(p; p) < 0$. In this case, since V is nonincreasing in its arguments, $V(t; q_D) < 0$ for all t and q_D, so it is optimal for all D types not to appeal, regardless of A's beliefs.

Case (ii) $V(1; 1) > 0$. In this case, since V is nonincreasing in its arguments, $V(t; q_D) > 0$ for all t and q_D, so it is optimal for all D types to appeal, regardless of A's beliefs.

Case (iii) $V(p; p) \geq 0$ and $V(1; 1) \leq 0$. Since V is nonincreasing in its arguments, there exists $q_D^* \in [p, 1]$ such that $V(q_D^*; q_D^*) = 0$. Moreover, this value of q_D^* is unique. To see this, first notice that $V(q_D^*; q_D^*) = 0$ implies that $q^*(q_D^*) \in (p, 1)$, since $q^*(q_D^*) = p$ implies $V(q_D^*; q_D^*) = \mathcal{A} - \mathcal{A} - C < 0$ (a contradiction) and $q^*(q_D^*) = 1$ implies $V(q_D^*; q_D^*) = \mathcal{A} - C > 0$ (a contradiction). Recall that when $q^*(t)$ is interior, it is continuously differentiable and strictly decreasing. Thus, if $V(q_D^*; q_D^*) = 0$, then

$$dV(q_D^*; q_D^*)/dq_D^* = \partial V(q_D^*; q_D^*)/\partial t + \partial V(q_D^*; q_D^*)/\partial q_D = \mathcal{A}\mu_D(q^*(q_D^*)\,|\,q_D^*)q^{*\prime}(q_D^*) + \mathcal{A}\partial \mathrm{M}_D(q^*(q_D^*)\,|\,q_D^*)/\partial q_D.$$

Both these terms are strictly negative; thus, $V(z; z) > 0$ for $z < q_D^*$ and $V(z; z) < 0$ for $z > q_D^*$. It remains to verify that $V(q_D^*; q_D) \geq 0$ for all $q_D \leq q_D^*$, and $V(q_D^*; q_D) < 0$ for all $q_D > q_D^*$. This follows immediately from the facts that $V(q_D^*; q_D^*) = 0$ and $\partial V(q_D^*; q_D^*)/\partial q_D = \mathcal{A}\partial \mathrm{M}_D(q^*(q_D^*)\,|\,q_D)/\partial q_D < 0$ by Assumption 3. *Q.E.D.*

References

ACZÉL, J. *Lectures on Functional Equations and Their Applications.* New York: Academic Press, 1966.

——— AND DHOMBRES, J. *Functional Equations in Several Variables.* Cambridge: Cambridge University Press, 1989.

BRAEUTIGAM, R., OWEN, B.M., AND PANZAR, J. ‘‘An Economic Analysis of Alternative Fee-Shifting Systems.’’ *Law and Contemporary Problems,* Vol. 47 (1984), pp. 173–185.

COFFIN, F.M. *On Appeal: Courts, Lawyering, and Judging.* New York: W.W. Norton, 1994.

DANZON, P.M. ‘‘Contingent Fees for Personal Injury Litigation.’’ *Bell Journal of Economics,* Vol. 14 (1983), pp. 213–224.

DAUGHETY, A.F. AND REINGANUM, J.F. ‘‘Keeping Society in the Dark: On the Admissibility of Pretrial Negotiations as Evidence in Court.’’ *RAND Journal of Economics,* Vol. 26 (1995), pp. 203–221.

——— AND ———. ‘‘Stampede to Judgment: Persuasive Influence and Herding Behavior by Courts.’’ *American Law and Economics Review,* Vol. 1 (1999), pp. 158–189.

——— AND ———. ‘‘On the Economics of Trials: Adversarial Process, Evidence and Equilibrium Bias.’’ *Journal of Law, Economics and Organization,* forthcoming.

FARMER, A. AND PECORINO, P. ‘‘Does Jury Bias Matter?’’ *International Review of Law and Economics,* forthcoming.

FROEB, L.M. AND KOBAYASHI, B.H. ‘‘Competition in the Production of Costly Information: An Economic Analysis of Adversarial Versus Court-Appointed Presentation of Expert Testimony.’’ Working Paper, Owen Graduate School of Management, Vanderbilt University, August 1993.

——— AND ———. ‘‘Naive, Biased, yet Bayesian: Can Juries Interpret Selectively Produced Evidence?’’ *Journal of Law, Economics and Organization,* Vol. 12 (1996), pp. 257–276.

FOSTER, J.E. AND SHNERYEROV, A. ‘‘Path Independent Inequality Measures.’’ *Journal of Economic Theory,* Vol. 91 (2000), pp. 199–222.

Goldsboro Christian Schools, Inc. v. *U.S.* and *Bob Jones University* v. *U.S.,* 456 U.S. 922 (U.S. Sup. Ct., 1982).

GRAHAM, M.H. *Federal Rules of Evidence in a Nutshell,* 3d ed. St. Paul, Minn.: West Publishing Co., 1992.

HARDY, G., LITTLEWOOD, J.E., AND PÓLYA, G. *Inequalities,* 2d ed. Cambridge: Cambridge University Press, 1952.

HAUSE, J.C. ‘‘Indemnity, Settlement, and Litigation, or, I'll Be Suing You.’’ *Journal of Legal Studies,* Vol. 18 (1989), pp. 157–179.

James, F., Jr. and Hazard, G.C., Jr. *Civil Procedure*, 3d ed. Boston: Little, Brown and Company, 1985.

Katz, A. ''Measuring the Demand for Litigation: Is the English Rule Really Cheaper?'' *Journal of Law, Economics and Organization,* Vol. 3 (1987), pp. 143–176.

———. ''Judicial Decisionmaking and Litigation Expenditure.'' *International Review of Law and Economics,* Vol. 8 (1988), pp. 127–143.

Landes, W.M. ''Sequential Versus Unitary Trials: An Economic Analysis.'' *Journal of Legal Studies,* Vol. 22 (1993), pp. 99–134.

Lewis, T.R. and Poitevin, M. ''Disclosure of Information in Regulatory Proceedings.'' *Journal of Law, Economics and Organization,* Vol. 13 (1997), pp. 50–73.

Lipman, B.L. and Seppi, D.J. ''Robust Inference in Communication Games with Partial Provability.'' *Journal of Economic Theory,* Vol. 66 (1995), pp. 370–405.

Milgrom, P.R. and Weber, R.J. ''A Theory of Auctions and Competitive Bidding.'' *Econometrica,* Vol. 50 (1982), pp. 1089–1122.

——— and Roberts, J. ''Relying on the Information of Interested Parties.'' *RAND Journal of Economics,* Vol. 17 (1986), pp. 18–32.

Plott, C.R. ''Legal Fees: A Comparison of the American and English Rules.'' *Journal of Law, Economics and Organization,* Vol. 3 (1987), pp. 185–192.

Posner, R.A. *The Federal Courts: Challenge and Reform.* Cambridge, Mass.: Harvard University Press, 1996.

———. *Economic Analysis of Law,* 5th ed. Boston: Little, Brown and Company, 1998.

———. ''An Economic Approach to the Law of Evidence.'' *Stanford Law Review,* Vol. 51 (1999), pp. 1477–1546.

In re Rhone-Poulenc Rorer Inc., 51 F.3d 1293, (7th Circuit, 1995).

Rubinfeld, D.L. and Sappington, D.E.M. ''Efficient Awards and Standards of Proof in Judicial Proceedings.'' *RAND Journal of Economics,* Vol. 18 (1987), pp. 308–315.

Sanchirico, C.W. ''Enforcement by Hearing: An Integrated Model of Evidence Production.'' Working Paper no. 9798-05, Department of Economics, Columbia University, 1997a.

———. ''Evidence Production, Adversarial Process and the Private Instigation of Suits.'' Working Paper no. 9798-06, Department of Economics, Columbia University, 1997b.

———. ''The Burden of Proof in Civil Litigation: A Simple Model of Mechanism Design.'' *International Review of Law and Economics,* Vol. 17 (1997c), pp. 431–447.

Seidmann, D.J. and Winter, E. ''Strategic Information Transmission with Verifiable Messages.'' *Econometrica,* Vol. 65 (1997), pp. 163–169.

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

Shepherd, G.B. ''An Empirical Study of the Economics of Pretrial Discovery.'' *International Review of Law and Economics,* Vol. 19 (1999), pp. 245–263.

Shin, H.S. ''The Burden of Proof in a Game of Persuasion.'' *Journal of Economic Theory,* Vol. 64 (1994), pp. 253–264.

———. ''Adversarial and Inquisitorial Procedures in Arbitration.'' *RAND Journal of Economics,* Vol. 29 (1998), pp. 378–405.

Skaperdas, S. ''Contest Success Functions.'' *Economic Theory,* Vol. 7 (1996), pp. 283–290.

Sobel, J. ''Disclosure of Evidence and Resolution of Disputes: Who Should Bear the Burden of Proof?'' In A.E. Roth, ed., *Game-Theoretic Models of Bargaining.* Cambridge: Cambridge University Press, 1985.

Spitzer, M. and Talley, E. ''Judicial Auditing.'' Working Paper no. 98-22, University of Southern California Law School, October 1998.

Yeazell, S.C. *Federal Rules of Civil Procedure, with Selected Statutes.* Boston: Little, Brown and Co., 1996.