Lawyer Advising in Evidence Disclosure

Che, Yeon-Koo and Severinov, Sergei

Columbia University, University of Essex

6 May 2007

Lawyer Advising in Evidence Disclosure*

Yeon-Koo Che† Sergei Severinov‡

April 11, 2007

ABSTRACT: This paper examines how the advice that lawyers provide to their clients affects the disclosure of evidence and the outcome of adjudication, and how the adjudicator should allocate the burden of proof in light of these effects. Despite lawyers' expertise in assessing the evidence, their advice is found to have no effect on adjudication, if the lawyers follow disclosure strategies that are undominated in a certain sense. A lawyer's advice *can* influence the outcome to his client's favor, if he can credibly advise his client to suppress some favorable evidence, but this effect is socially undesirable.

KEYWORDS: Legal advice, disclosure of evidence, burden-of-proof allocation, regulating adjudicators' inferences.

---

*The authors thank Andrew Daughety, Bob Hall, Ken Judd, Navin Kartik, Bart Lipman, Mitch Polinsky, and Kathy Spier, the participants at the 2005 Decentralization Conference, Duke-Northwestern-Texas IO Conference, seminar participants at the University of Arizona, Stanford Law and Economics Seminar, and Hoover Brown Bag Lunch, for helpful comments. This paper was completed while the second author was a National Fellow at the Hoover Institution, Stanford University, and he thanks the Hoover Institution for hospitality. The first author acknowledges financial support from National Science Fellowship (SES 0319061) and Shoemaker Fellowship.

†Department of Economics, Columbia University, email: yc2271@columbia.edu

‡Fuqua School of Business, Duke University, email: sseverin@duke.edu

1 Introduction

Lawyers play a highly prominent role in the modern day adjudication process. An important aspect of their role involves advising clients regarding disclosing information to the court. Lawyers can advise their clients to withhold unfavorable information and disclose favorable one. Although lawyers often (particularly, in civil cases) have a disclosure duty before the tribunal, the rules of confidentiality and attorney-client privilege enable them to suppress evidence during discovery and trial, particularly when the opposing party and the tribunal are unaware of the existence of the evidence.[1] As Kaplow and Shavell (1989) point out, "the combination of carefully crafted responses, limited testimony and the adversary's inability to conceive of (or expend the resources to ask) every possible question may well result in a significant gap between the information learned by the adversary's lawyer and that possessed by the client." To the extent that lawyers can affect the amount and nature of information reaching the court, they could affect the outcome of a trial. Our aim is to understand this effect.

To this end, we consider an evidence disclosure game played by disputing parties who may obtain lawyer advice. Specifically, we study a dispute between two parties, say defendant and a plaintiff, which is tried by an adjudicator called "judge." In the trial, each party presents his privately-held evidence, and the judge rules either to "convict" or "acquit" the defendant based on the disclosed evidence. The party's evidence is "hard" in that he cannot manipulate it, so the main decision facing each party is whether to disclose his evidence or not. This decision is not trivial, however, since the judge's ruling depends not just on the evidence itself but also on another piece of information reflecting the legal rules and standards that are applied to interpret that evidence and the other public evidence surrounding that case. We assume that the lawyers can assess this latter information better than the parties, so a lawyer can assess whether a party's evidence is favorable or unfavorable and how strong his case would be without its disclosure. A lawyer-represented party can thus make a more informed decision about disclosure. We study this particular role of lawyers.

At first glance, lawyers' expertise appears to leave no doubt about the value of their advice, at least to their clients. Armed with the knowledge of the law, a lawyer should be able to improve her client's disclosure decision. Surprisingly, however, this benefit does not always materialize. We find legal advice to be irrelevant — both privately and socially — when the parties employ disclosure strategies strategies that satisfy a certain credibility requirement: i.e., a party discloses information if and only if it is favorable. In that case, legal advice indeed affects one's disclosure behavior, but it does *not* affect the outcome of adjudication. This irrelevance holds regardless of

[1] The attorney-client privilege protects privileged information in testimony at trial. Federal Rules of Civil Procedure (Rules 26(b)(1) and 26(b)(3)) limit discovery of privileged information and trial preparation materials.

whether one or both parties obtain legal advice, and whether the adjudicator makes a Bayesian inference based on the parties' disclosure strategies or she simply follows an ad-hoc rule satisfying some reasonable properties.

We then extend our model to identify a circumstance in which legal advice on disclosure is relevant. We show that a lawyer advice matters if she can credibly follow a strategy of suppressing some favorable evidence. Such a strategy can skew the inference by the court, and thus the adjudication outcome, in favor of his client. This role of lawyers generates a private incentive for hiring lawyers, but the aggregate welfare falls if both parties hire lawyers, which suggests a "prisoners' dilemma" type explanation for the prevalence of lawyer representation. Moreover, this role of lawyers distorts parties' disclosure in a socially undesirable way. We show that this harm can be remedied if the adjudicator commits ex ante to the way in which she assigns the burden of the proof and thus the way in which she draws an inference about the guilt. This last result provides some rationale for restrictions the law places on how judges and juries should interpret evidence or lack thereof.

Our analysis of legal representation has several broad implications. First, our model provides a useful framework for analyzing the advisory role of lawyers in dispute resolution. Admittedly, lawyer representation in the real world includes many important aspects, not all captured in our simple model. Yet, the advisory role of lawyers in disclosure remains an important one, and our model identifies ways in which this role may (or may not) affect the outcome of adjudication. In this sense, our model can serve ultimately as a useful benchmark — a useful building block for studying various aspects of lawyer representation.

Our paper also yields useful insights into various rules and restrictions on the inferences that adjudicators are allowed to draw from nondisclosure of evidence. First, we show that no such restrictions are warranted with or without legal representation, when the adjudicator is Bayesian and the lawyers disclose all favorable evidence. In this case, the equilibrium outcome is socially optimal. However, this conclusion no longer holds when the lawyers can credibly use strategies of withholding some favorable evidence. In this case, the extent of potential harm to social welfare can be reduced by committing the adjudicator to a rule which allocates all burden of proof to one party. These results contribute to the understanding of evidentiary rules and procedures adopted by courts.

Our modeling framework and the results may be useful for understanding the role of advising more broadly, namely in settings other than dispute resolution. Indeed, much of the insight we develop on advising does not depend on the "dispute" context, and holds equally well in a setting where there is only one agent. Often, decisions that have significant consequences on an agent must be made based on the information provided by that agent. Promotion and grant allocation, college admission and job application, product introduction and promotion are some relevant examples. An agent facing a decision in such a context (e.g. writing a grant proposal or preparing a case for promotion, choosing a strategy of college or job application, or a strategy of product introduction and promotion) often seeks or is encouraged to seek advice from mentors, counselors or consultants regarding strategies of information revelation. Our results offer some basic necessary condition for the advising to be relevant.

The issue of legal advice has received very little formal treatment in the literature. Legal scholars have recognized the factors favoring and disfavoring lawyer-aided adversarial system but disagree on the relative importance of those factors. Proponents argue that a vigorous adversarial competition among lawyers leads the court to focus on relevant evidence, thus making judicial fact-finding efficient (Luban, 1983; Bundy and Elhauge, 1992 and 1993). Critics point out that lawyers can mislead as much as inform the court (Frank, 1973). In particular, Kaplow and Shavell (1989) point out, via illustrative examples, that while the lawyers' ability to suppress evidence based on legal expertise undoubtedly benefits their clients, its social implications are ambiguous, thus casting doubt on its social benefit. Although the current paper is similar in spirit to the last study, there are important distinctions. First, these authors do not perform a full-fledged equilibrium analysis of the disclosure game, focusing rather on the effect of legal advice when possible outcomes are exogenously fixed. Second, they treat the adjudicators' inferences as exogenous, while we allow the inferences to depend on the strategies the players may employ, with or without the lawyers. Among other benefits, this latter approach enables us to study how the rules and restrictions on inferences may affect the adjudication outcomes.

This paper is also closely related to the economics literature on disclosure of non-falsifiable information. Grossman (1981), Milgrom (1981), Milgrom and Roberts (1986) and Lipman and Seppi (1995) find that conflicting interests can lead to full revelation of commonly shared information by the agents. Shin (1994, 1998) studies information revelation when the possession of information is not common knowledge, which makes the inference (or "burden of proof") nontrivial. In particular, our basic model is related to Shin (1998), who compares adversarial and inquisitorial litigation systems. Froeb and Kobayashi (1996), (2001) and Daughety and Reinganum (2000b) assess the implications of endogenous evidence production arising from the adversarial system. Sobel (1985), Hay and Spier (1997), Sanchirico (1997, 1998, 2001), and Bernardo et al. (2000) explore the allocation of the burden of proof and evidence production from the standpoint of litigation costs and/or deterrence.[2] Seidmann (2005) and Mialon (2005) investigate the effect of

[2]Dewatripont and Tirole (1999) study the desirability of adversarial system in a broad organization design context. Levy (2005) studies the effect of career concerns on judges' decision making. Also related, albeit with less relevance in the litigation setting, is the literature on cheap talk (or communication of falsifiable information), which includes Crawford and Sobel (1982) and Krishna and Morgan (2001).

the defendant's right to silence, with and without adverse inference by the adjudicator after such right is exercised, on the adjudication outcomes and welfare. None of these papers deal with the role of lawyers in disclosure — the focus of the current paper.

2 Model

Two parties, 1 and 2, are in dispute, which is adjudicated by a judge/jury in a tribunal. It is convenient to interpret parties 1 and 2 as a defendant and a plaintiff in a litigation, so we will invoke this interpretation throughout the paper. However, our model is fairly general and can apply equally well to a number of different settings. The adjudicator in our model can be either a judge or a jury or a combined entity, whom we shall call simply "the judge," throughout. Lawyers provide legal advice, if hired by the parties.

There are two pieces of judgment-relevant information that pertain to the case. First, there is evidence $s \in [0,1] =: S$ which may only be observed by the parties to the dispute. The evidence is observed with probability p_{00} by neither party, with probability p_{11} by both parties; and with probability p_{10} (resp. p_{01}) by party 1 only (resp. party 2 only).[3] Obviously, $\sum_{i,j=0,1} p_{ij} = 1$, and we assume that $p_{ij} > 0$ for all $i,j = 0,1$. Note that we allow for possible correlation in the parties' abilities to observe the evidence. The evidence is "hard" in the sense that, while it can be concealed, it cannot be fabricated or manipulated. For instance, the evidence can take the form of an unforgeable document or a non-perjuring witness. Equivalently, the evidence may be soft but perjury laws prevent the possessor of the evidence from falsifying it. It is well known that the non-falsifiability of information, coupled with conflict of interests of the players, leads to full revelation of information (Grossman, 1981; Milgrom and Roberts, 1986). Unraveling of this kind will not occur in our setting, however, since the possession of evidence is no longer a common knowledge.

The other piece of judgment-relevant information, $\theta \in [0,1] =: \Theta$ can only be observed by the lawyers and the judge. The variable θ represents the judge's interpretation of the laws and legal standards in application to the current case. Further, θ may also reflect the court's view regarding the evidence, as well as its interpretation of external circumstances surrounding the case, such as basic uncontested facts, police reports, the testimony by neighbors, etc. Thus, when s is disclosed, the judge's ruling depends on both s and θ, and when s is not disclosed, the ruling depends only on θ.[4] The disputing parties have limited knowledge of the law and incomplete understanding of the

[3] "Observing" s means either possessing that evidence or having proof of its existence (perhaps in the opponent's possession).

[4] Posner (1999) discusses a class of 'bare bones cases' in which very little evidence is presented by the parties, and the adjudicator has to rule on the basis of the law and a few uncontested facts. Such 'bare bones' cases fit the

legal process, so they can observe θ only by hiring lawyers. Lawyers understand the body of the law in the jurisdiction where they practice, as well as the judge's interpretation of the law and her possible biases. For instance, the lawyer and the judge may be able to assess more accurately how strong or weak the mitigating circumstances are for a litigant. Ultimately, the lawyers' ability — and the litigants' *in*ability — to observe θ serves a modeling purpose of introducing a productive role for the lawyers.

We assume that (s, θ) is drawn from $S \times \Theta$ according to an absolutely continuous cdf, $F(s, \theta)$ which has a positive density $f(s, \theta)$ in the interior of $S \times \Theta$. From the ex-ante perspective, θ is random because it describes the realized state of the law and legal standards — one out of many possibilities, as well as a particular realization of commonly known evidence. Since s and θ reflect the nature of underlying case, they may be correlated. We assume that s and θ satisfy the (weak) Monotone Likelihood Ratio Property (MLRP):

Assumption 1 **(MLRP)** *For all $s' \geq s$ and $\theta' \geq \theta$, $f(s', \theta')/f(s, \theta') \geq f(s', \theta)/f(s, \theta)$.*

To understand the value of legal advice, we will compare two regimes. In the first regime, the parties are not represented by lawyers and do not receive any legal advice. In the second regime, both parties are represented by lawyers, at no cost to them. Self representation serves as a benchmark necessary for our analysis, but it is not without practical relevance. Although few parties represent themselves in civil or criminal trials in state or federal courts in the U.S., many litigants do so in municipal courts and administrative trial procedures. Also, in small claims courts — which comprise a significant share of trials in the U.S. — legal representation is expressly forbidden in most states (California, New York, Arizona, and others).[5] Further, our comparison should not be narrowly interpreted as pertaining only to the two regimes; rather, it applies to any increase in the quality of lawyer representation. For instance, one could view the two regimes as involving lawyer representation but differing only in the quality of representation.

The time line of the events in both regimes is as follows. At date 0, (s, θ) is realized. At date 1, parties 1 and 2 observe the evidence s with probabilities $p_{10} + p_{11}$ and $p_{01} + p_{11}$, respectively, while the judge and the lawyers learn θ. At date 2 (trial), party 1 and party 2 simultaneously and independently decide whether to disclose the evidence s to a judge, provided that the respective party has observed it. In the representation regime, this decision is taken with the help of a lawyer providing legal advice. At date 3, the judge rules either for party 1 or for party 2.

description of situations where s is not disclosed.

[5]See Spurrier (1980) for detail. The problem of withholding evidence is particularly relevant in this case, since the discovery process is very limited and the trials focus on a few key elements of evidence.

• Evidence disclosure behavior:

If a party is not represented by a lawyer, then his decision whether to disclose s is based solely on s. In contrast, if a party is represented by a lawyer, the lawyer can provide advice based on his knowledge of θ. In particular, a lawyer can advise the client whether disclosing s is beneficial to him for given θ.

We assume that a lawyer would prefer his client to prevail in court, and that there are no agency issues in the attorney-client relationship. Given the congruence of interests between the lawyer and the client, the client will have full incentives to communicate s to his lawyer, and likewise the lawyers will have the incentives to explain the legal issues, i.e. to communicate θ to the client truthfully. Therefore, the represented party can simply be viewed as informed of both s (if he observes s) and θ.

Thus, the difference between representation and no representation in our model boils down to the information available to the party when he makes the disclosure decision. In the regime of no representation, the disclosure decision is made solely on the basis of s itself, while in the regime of representation it is based on both s and θ. Formally, party i's disclosure strategy is a function ρ_i that maps $S \times \Theta$ to $[0, 1]$, with $\rho_i(s, \theta)$ representing the probability that party $i = 1, 2$'s discloses s, in state θ. If the party is not represented, he does not observe θ, so $\rho_i(\cdot, \theta)$ must satisfy the requirement that $\rho_i(\cdot, \theta) = \rho_i(\cdot, \theta')$ for any $\theta \neq \theta'$.

• Judge's Adjudication behavior:

In the last stage of the game, the judge makes a binary decision, ruling either for party 1 or party 2. For instance, in a criminal trial, the judge convicts or acquits the defendant. Such a binary decision is quite common, and it is more general than may appear at first glance. For instance, there may be no ambiguity about the size of damages in case the plaintiff prevails, leaving the liability as the only source of dispute.[6]

The judge's ruling depends on (s, θ) if s has been disclosed, and on θ only if s has not been disclosed. The judge's decision given (s, θ) is described by a function $g(s, \theta)$, interpreted as the her assessment of party 1 (defendant)'s culpability. Specifically, if $g(s, \theta) > 0$, then the judge finds party 1 culpable and thus rules for party 2. If $g(s, \theta) < 0$, the judge finds party 1 innocent and rules for him. The judge is indifferent if $g(s, \theta) = 0$, but since the distribution $F(s, \theta)$ is absolutely continuous, how a tie is broken in this case has no real consequence.

We assume that the function $g(s, \theta)$ is increasing and continuous in both arguments. Thus, lower s and θ are more favorable for party 1, and higher values of them are more favorable for

[6]The binary feature can also be justified in an idealistic Beckerian world in which any defendant found liable is subject to a sanction equaling his maximum wealth limit.

party 2. In a tort setting, a higher value of s would mean that the defendant (party 1) is more likely to have caused a harm, while a higher value of θ indicates that the law and legal standards are more unfavorable toward the defendant. To make the judge's decision problem nontrivial, we assume that $\int g(s,1)f(s|1)ds > 0$ and $\int g(s,0)f(s|0)ds < 0$, which means that publicly available information and legal standards have enough inherent variability that the judge's unconditional belief about the culpability swings from one side to the other as θ changes from the most favorable to party 1 (i.e., $\theta = 0$) to the most unfavorable (i.e., $\theta = 1$).[7] Since $g(s,\theta)$ is monotonically increasing in both arguments, there exists a strictly decreasing continuous function $s = h(\theta)$ such that $g(h(\theta),\theta) = 0$ for all $\theta \in [\underline{\theta},\overline{\theta}]$, where $\overline{\theta} := \max\{\theta|\exists s' \in S \text{ s.t. } g(s',\theta) = 0\}$ and $\underline{\theta} := \min\{\theta|\exists s'' \in S \text{ s.t. } g(s'',\theta) = 0\}$. This function partitions the evidence/legal environment space into two regions where the judge rules for party 1 and party 2 respectively when she observes both s and θ, as depicted in Figure 1.[8]

The adjudication criterion $g(s,\theta)$ can be justified by a society's objective that the judge follows. Suppose the society minimizes the cost associated with a wrong decision, i.e. "convicting the innocent or exonerating the guilty," with appropriate costs assigned to each type of mistake. Let c_1 and c_2 be the cost of ruling mistakenly for party 1 ("exonerating the guilty") and for party 2 ("convicting the innocent"), respectively, and let $\pi(s,\theta)$ be the probability that for given (s,θ) the party 1, the defendant, is guilty. Then, if the judge convicts party 1 (i.e., the defendant) with probability z, the expected cost of a mistake is

$$(1 - \pi(s,\theta))c_2 z + \pi(s,\theta)c_1(1 - z).$$

To minimize this cost, the judge should choose $z = 1$ if $\pi(s,\theta) - \frac{c_2}{c_1+c_2} > 0$ and should choose $z = 0$ otherwise. Our model accommodates this behavior if we let $g(s,\theta) := \pi(s,\theta) - \frac{c_2}{c_1+c_2}$.[9]

We assume that the judge's adjudication criterion $g(\cdot,\cdot)$ is a common knowledge to all players, including the lawyers and parties 1 and 2. This assumption does not rule out the possibility that the adjudication criterion may be biased, as the legislature or higher courts may impose standards of proofs that differ from the one corresponding to the costs c_1 and c_2 the society assigns.

[7] This assumption is mainly to simplify exposition. Its only use is to allow for nontrivial analysis in Section 5.

[8] The two regions have nonempty interiors given the above assumption.

[9] Different standards of proof and evidence adopted by the courts are consistent with this model. Indeed, let $\alpha := \frac{c_2}{c_1+c_2}$. If $\alpha = 0.51$, then the judge can be said to follow the rule of preponderance of evidence. The interval of $(0.6, 0.7)$ corresponds to the standard of "clear and convincing evidence." According to Posner (1999) judges associate probability level between 0.75 and 0.9 with the standard of "proof beyond a reasonable doubt."

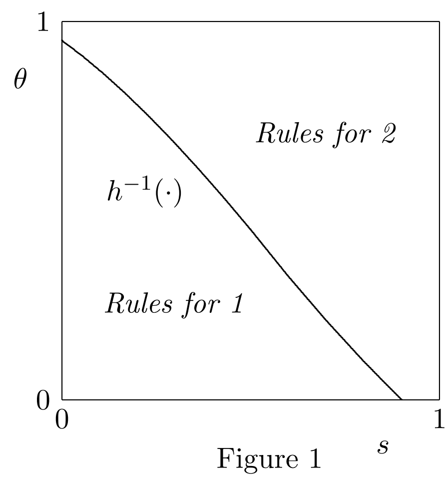

Figure 1

If no party discloses s, then the adjudicator decides based on θ and, possibly on her inference about the parties' disclosure decisions. The adjudicator's decision rule in this case, henceforth referred to as *default ruling strategy*, is described by the function $\delta : \Theta \mapsto [0,1]$, where $\delta(\theta)$ denotes the probability with which she rules for party 2 if she observes signal θ and no evidence is disclosed. The judge's default ruling strategy will depend on her *posterior assessment of party 1's culpability*, or simply her *posterior*. We model the posterior quite generally, allowing for both Bayesian and non-Bayesian updating as special cases. Specifically, the judge's posterior is given by

$$\mathbb{E}[g|\rho_1(\cdot),\rho_2(\cdot),\theta;a,b_1,b_2,c] := a\mathbb{E}_\emptyset[g|\theta] + b_1\mathbb{E}_1[g|\theta] + b_2\mathbb{E}_2[g|\theta] + c\mathbb{E}_{12}[g|\theta], \tag{1}$$

which is a weighted average of expected culpability criterion based on alternative evidence scenarios, with nonnegative constants a, b_1, b_2 and c used as weights. The first term, $E_\emptyset[g|\theta] := \int_0^1 g(s,\theta)f(s|\theta)ds$, is party 1's expected culpability given the presumption that no party has observed the evidence s; $\mathbb{E}_i[g|\theta] := \int_0^1 g(s,\theta)(1-\rho_i(s,\theta))f(s|\theta)ds$ is the (normalized) expectation of g given the presumption that only party $i = 1,2$ has observed s but has not disclosed; and the last expectation term, $\mathbb{E}_{12}[g|\theta] := \int_0^1 g(s,\theta)(1-\rho_1(s,\theta))(1-\rho_2(s,\theta))f(s|\theta)ds$, is based on the presumption that both have observed s but neither has disclosed it.[10] Absent disclosure of s, the judge applies this posterior, ruling in favor of 2 if and only if $\mathbb{E}[g|\rho_1(\cdot),\rho_2(\cdot),\theta] > 0$. (The dependence of the posterior on (a,b_1,b_2,c) will be suppressed when this does not generate any ambiguity.)

[10]The value of (1) would correspond to a weighted expectation of the adjudication criterion with arbitrarily fixed weights a, b_1, b_2, c if we normalize it dividing by $a + b_1\int_0^1(1-\rho_1(s,\theta))f(s|\theta)ds + b_2\int_0^1(1-\rho_2(s,\theta))f(s|\theta)ds + c\int_0^1(1-\rho_1(s,\theta))(1-\rho_2(s,\theta))f(s|\theta)ds$. Since the judge's ruling depends only on the sign of (1), all our results are invariant to this normalization. So, for brevity we work with (1) without normalizing it.

The coefficients, (a, b_1, b_2, c), henceforth referred to as the judge's *inference rule*, reflect how the judge weighs alternative evidence scenarios in her inference formation/burden of proof assignment. Throughout, we will only assume that the judge applies the same criterion, i.e. the coefficients (a, b_1, b_2, c) remain constant, regardless of whether a party is represented by a lawyer or not.

Since only the sign of the posterior matters for the judge's decision, we normalize by setting $a = 1$, and focus on the values of (b_1, b_2, c).[11] Depending on the values of these variables, our adjudication criterion in (1) accommodates a variety of different decision procedures and burden-of-proof allocation rules. For example, if $b_1 = b_2 = c = 0$, then the judge bases her decision only on the prior expectation of g. In this case, the judge is *completely non-Bayesian* in the sense that she does not account for the possibility that one of the parties may be withholding evidence. If $b_1 > 0$ and $b_2 = c = 0$, then the judge never attributes nondisclosure of evidence to party 2's withholding it. In other words, the burden of proof is put on party 1. Likewise, if $b_2 > 0$ and $b_1 = c = 0$, then the burden of proof is assigned to party 2. If both b_1 and b_2 are strictly positive, then the judge assigns some weight to either party withholding the evidence, so the burden of proof is split between the two parties. One important case arises when the judge is *fully Bayesian*, i.e. $(b_1, b_2, c) = (\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}})$. In this case, the judge's posterior assessment assigns accurate probability weights to alternative scenarios of evidence withholding.

In the legal literature, the burden of proof is defined as an obligation to provide evidence which proves or disproves allegations. A party that bears the burden of proof loses a case if he or she is not able to meet it by submitting such evidence. Our modeling of the burden of proof is consistent with this notion. Indeed, as we will show below, if a single party bears the entire burden of proof in our model (e.g. if b_1 is sufficiently large and $b_2 = c = 0$), then this party can only win the case by providing favorable evidence. Moreover, by varying the coefficients b_1, b_2, c we are able to quantify the effect of burden-of-proof allocation, and show how the extent of disclosure by a party depends on its share of the burden-of-proof.

There is an active debate in the legal literature regarding the appropriate allocation of the burden of proof, as well as the applicability of Bayesian approach. It is widely acknowledged that adjudicators are prone to biases and errors in computing the true statistical odds of events (see Tribe (1971)) and are often reluctant to convict on the basis of simple statistical likelihood.[12]

[11] The only loss is when $a = 0$. This case is arbitrarily closely approximated by $a \approx 0, a > 0$. Further, c will be seen to play a limited role.

[12] One of the most well-known examples extensively discussed in the literature to highlight the problems with this kind of cases and the use of statistical evidence is the so-called Blue Bus/Grey Bus case. In this case, a plaintiff has been negligently hit by a bus in the location where Blue Bus Company operates a greater number of buses than Grey Bus Company. A direct application of 'more likely than not' criterion should lead the court to convict Blue Bus company on the basis of the 'bare bones' statistical evidence that blue buses are more numerous and, therefore, are more likely to have hit the plaintiff. Yet, experimental results (see Wells (1992)) show that judges

Therefore, it is important to allow for non-Bayesian — as well as Bayesian — burden-of-proof allocations.

The judge's inference rule may also reflect the legal rules and procedures intended to regulate the adjudicator's behavior. Evidence laws often restrict the admissibility of certain types of evidence and limit the inferences which a judge or a jury are allowed to make from certain evidence or lack thereof, because of concerns about their prejudicial effect. Our model will allow us to study the implications of such restrictions. From this perspective, the coefficient b_i can be interpreted as the extent to which the rule allows the judge to be "rationally" prejudiced against party i in interpreting his nondisclosure.

For a later purpose, it is useful to consider a *posterior assessment* arising when the parties follow *cutoff* strategies. Suppose party 1 employs a strategy of disclosing her/his evidence if and only if $s < \hat{s}_1$ and party 2 discloses if and only if $s > \hat{s}_2$. The judge's posterior under such cutoff strategies (with a slight abuse of notation) is given by

$$\begin{aligned}\mathbb{E}[g|\hat{s}_1, \hat{s}_2, \theta; b_1, b_2, c] := \int_0^1 g(s,\theta)f(s|\theta)ds + b_1 \int_{\hat{s}_1}^1 g(s,\theta)f(s|\theta)ds + b_2 \int_0^{\hat{s}_2} g(s,\theta)f(s|\theta)ds \\ + c\int_{\{\hat{s}_1 \le s \le \hat{s}_2\}} g(s,\theta)f(s|\theta)ds. \qquad (2)\end{aligned}$$

• **Equilibrium concept and outcome:**

In each regime, we focus on Perfect Bayesian equilibria in the parties' disclosure strategies and the judge's default ruling strategy, summarized by a triple, (ρ_1, ρ_2, δ). This means that the judge's ruling in the event of disclosure is essentially "hard wired" to follow the criterion g. We do so to focus on a more nontrivial inference problem facing the judge in the event of nondisclosure. But this assumption can be also justified by the view that the criterion $g(s, \theta)$ (in case s is disclosed) is given by an immutable legal rule, so any deviation from the criterion would constitute an "error" of law. A broader justification is that the judge has a more scope for discretion (say in burden of proof allocation) or that there is more ambiguity in decision when crucial evidence is not disclosed than when it is disclosed.

The object of our interest is not just parties' disclosure behavior per se, but ultimately the decision that gets made in equilibrium given the information available to the parties. Formally,

and members of the jury are very unlikely to make such conviction when only such evidence is presented. Several legal scholars (e.g. Posner (1999) and Thompson (1989)) explain the reluctance to convict by the fact that it is quite implausible that the statistic is the only evidence available to the plaintiff. That is, absence of other evidence should lead the adjudicator to infer that the plaintiff is concealing some evidence indicating that the bus actually belonged to the other bus company. The latter point of view is consistent with the judges and juries following an adjudication criterion such as (2) with $b_i > 0$.

an *adjudication outcome* is a function, $\phi : X_1 \times X_2 \times S \times \Theta \mapsto [0,1]$, that maps the state of the world (x_1, x_2, s, θ) into the probability that the judge rules for party 2, where $x_i \in \{0,1\}, i = 1,2$, with $x_i = 1$ if party i observes s and $x_i = 0$ if not. In particular, an equilibrium (ρ_1, ρ_2, δ) induces an outcome function via

$$\phi(x_1, x_2, s, \theta) = \delta(\theta)(1-x_1\rho_1(s,\theta))(1-x_2\rho_2(s,\theta))+\mathbb{I}_{\{g(s,\theta)\geq 0\}}\left[1-(1-x_1\rho_1(s,\theta))(1-x_2\rho_2(s,\theta))\right],$$

where $\mathbb{I}_{\{\cdot\}}$ takes 1 in the event of $\{\cdot\}$ and zero otherwise. We are interested in comparing the adjudication outcomes induced by equilibria under different legal regimes.

3 Irrelevance of Legal Advice

In this section, we characterize equilibrium outcomes across legal regimes that differ in the availability of (costless) legal advice. We then compare them.

3.1 No Representation

In this regime, neither party 1 nor party 2 has a lawyer. Thus, they must decide whether to disclose the evidence s without being certain about the value of θ, and thus without knowing whether this disclosure will lead to a favorable or an unfavorable ruling by the judge.

We shall establish that there exists a unique perfect Bayesian equilibrium. In this equilibrium, both parties and the judge adopt cutoff strategies. In particular, there exists a common threshold $\hat{s}$ such that party 1 discloses s if and only if $s < \hat{s}$, and party 2 discloses s if and only if $s > \hat{s}$. Absent disclosure, the judge rules for party 1 if $\theta < \hat{\theta}$ and for party 2 if $\theta > \hat{\theta}$, for some threshold $\hat{\theta}$ that makes the judge indifferent. The judge's cutoff strategy is explainable by her posterior $\mathbb{E}[g|\hat{s}, \hat{s}, \theta]$ being monotonic in θ, which follows from two effects. First, a higher θ is by itself a stronger evidence of 1's culpability holding s fixed. Second, there is also an inference effect: Given MLRP, a higher θ makes a high value of s more likely, which makes nondisclosure more attributable to party 1's concealing unfavorable s, given the parties' cutoff strategies. Obviously, this inference effect adds to the judge's suspicion of 1's culpability. Figure 2 graphs the two thresholds $\hat{s}$ and $\hat{\theta}$.

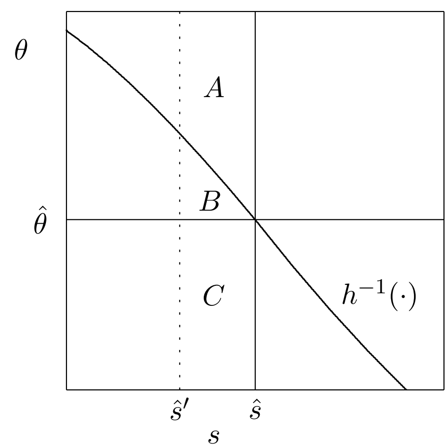

Figure 2

An important fact is that the two thresholds $\hat{s}$ and $\hat{\theta}$ cross each other on the curve g; i.e., $g(\hat{s}, \hat{\theta}) = 0$ or $\hat{s} = h(\hat{\theta})$. This can be explained as follows. Suppose that, facing the judge's threshold $\hat{\theta}$, party 1 deviates by withholding some s with $\hat{s}' < s < \hat{s} = h(\hat{\theta})$. We can show that such a deviation is not profitable. To this end, assume that party 2 does not disclose s, or else party 1's disclosure wouldn't matter. If θ is either below $\hat{\theta}$ or above $h^{-1}(s)$ (regions C or A in Figure 2), then the deviation makes no difference, for the judge will rule exactly the same either way. But if θ lies between $h^{-1}(s)$ and $\hat{\theta}$ (region B), then withholding s will result in a ruling against party 1 whereas disclosing it would result in a ruling in favor of him. So withholding any $s < h(\hat{\theta})$ is never profitable. A similar argument shows that disclosing $s > h(\hat{\theta})$ can never pay for party 1.

This argument shows why party 1, and by symmetry party 2, will adopt a cutoff strategy with threshold $\hat{s} = h(\hat{\theta})$. Substituting this into (2), the judge's equilibrium posterior becomes $\mathbb{E}[g|h(\hat{\theta}), h(\hat{\theta}), \hat{\theta}]$. Hence, her cutoff threshold is given by:

$$\hat{\theta}^*(b_1, b_2) := \inf\{\hat{\theta} \in \Theta \,|\, \mathbb{E}[g|h(\hat{\theta}), h(\hat{\theta}), \hat{\theta}; b_1, b_2, c] > 0\}, \tag{3}$$

where $\hat{\theta}^*(b_1, b_2) := 1$ if the set in the RHS is empty. Note that the threshold does not depend on c since the last expectation term, $\mathbb{E}_{12}[\cdot]$, in (2) vanishes when the parties employ the same threshold in the cutoff strategies.

It is interesting to see how the judge's threshold varies with her inference rule. As the burden of proof shifts from party 2 to party 1 (i.e., as b_1 rises and/or b_2 falls), the judge's threshold falls. This, in turn, causes an increase in the parties' common disclosure threshold, $h(\hat{\theta}^*(b_1, b_2))$. Hence, the party with an increased burden of proof discloses more evidence and the party with a decreased burden discloses less evidence. The next proposition summarizes the result. Much of its proof (which appears in the Appendix, along with all subsequent proofs) is concerned about demonstrating the cutoff structure of the equilibrium itself.

PROPOSITION 1 *If no party is represented by a lawyer, there exists a unique Perfect Bayesian equilibrium in which party 1 discloses s if and only if $s < h(\hat{\theta}^*(b_1, b_2))$, party 2 discloses s if and only if $s > h(\hat{\theta}^*(b_1, b_2))$, and the judge rules for party 1 if $\theta < \hat{\theta}^*(b_1, b_2)$ and for party 2 if $\theta > \hat{\theta}^*(b_1, b_2)$ following nondisclosure. As the burden of proof shifts from party 2 to party 1, the latter discloses more and the former discloses less; i.e. $\hat{\theta}^*(b_1, b_2)$ decreases, and hence $h(\hat{\theta}^*(b_1, b_2))$ increases, in $(b_1, -b_2)$.*

3.2 Full Representation

In this regime, both parties are represented by lawyers and learn θ through legal advice. Hence, unlike the no representation case, the parties make their disclosure decisions based on both s and θ. Recall that the judge's ruling in case of disclosure is "hard-wired" to follow the criterion $g(s, \theta)$. Given this convention, party 1 has a weakly dominant strategy of disclosing s if and only if $s < h(\theta)$, or $g(s, \theta) < 0$. Specifically, disclosing $s < h(\theta)$ leads to a sure win for party 1, whereas withholding may entail an unfavorable ruling. Likewise, withholding $s > h(\theta)$ is a dominant strategy for party 1 because the judge may rule for party 1 without disclosure, but will rule against him for sure if s is disclosed. By the same logic, party 2's weakly dominant strategy is to disclose s if and only if $s > h(\theta)$, or $g(s, \theta) > 0$.

Dominant strategies have an intuitive appeal in our model, particularly with representation. In the US legal system, a lawyer has a positive duty to explore all avenues of defense, and withholding exculpatory evidence may contravene this obligation. Getting a client's consent for such strategy may also be problematic. Furthermore, the judge could simply refuse to believe that a lawyer is not following a dominant strategy. Finally, if there is even small uncertainty about the judge's default ruing, then disclosing all favorable evidence and withholding all unfavorable evidence is the unique optimal strategy for either party. Any of these possibilities rule out any other equilibria under representation. For this reason, we focus on the dominant disclosure strategies here. Later, we shall consider what happens when these arguments to not apply and examine equilibria supported by "dominated" disclosure strategies.

The disclosure behavior of a party represented by a lawyer is clearly different from that of a non-represented party. Remarkably, this change in disclosure behavior does *not* affect the judge's default ruling. Indeed, given the dominant disclosure strategies by the parties, the judge's posterior is given by $\mathbb{E}[g|h(\theta), h(\theta), \theta]$. This posterior is different in magnitude from that held by the judge in the case of no representation, but the two posteriors always have the same sign, since

$$\mathbb{E}[g|h(\theta), h(\theta), \theta] \gtrless 0 \text{ if } \theta \lessgtr \hat{\theta}^*(b_1, b_2).$$

Hence, we arrive at the following result.

Proposition 2 *If both parties are represented, there exists a unique equilibrium in undominated strategies in which, absent disclosure, the judge rules for party 1 if $\theta < \hat{\theta}^*(b_1, b_2)$ and for party 2 if $\theta > \hat{\theta}^*(b_1, b_2)$. Party 1 discloses s if and only if $g(s, \theta) < 0$, and party 2 discloses s if and only if $g(s, \theta) > 0$.*

That the threshold $\hat{\theta}^*(b_1, b_2)$ is the same in the no representation and full representation cases is surprising, and it is worth exploring the logic behind it. In the no representation case, the parties employ a (common) threshold $\hat{s} = h(\hat{\theta}^*(b_1, b_2))$ that does not vary with θ, whereas in the full representation case, their threshold $h(\theta)$ varies with θ and thus in general differs from $h(\hat{\theta}^*(b_1, b_2))$. This difference in disclosure strategies leads to different inferences and different posterior assessments by the judge. Yet, the judge's posterior changes in sign from negative to positive with θ, around the same threshold $\hat{\theta}^*(b_1, b_2)$ in both regimes. So, the posterior has the same sign under both regimes, even though their exact magnitudes will typically be different. Hence, the judge's default ruling is the same under both regimes.

3.3 Partial Representation

The results of the previous subsections generalize to the regime in which only one side hires a lawyer. Suppose without loss of generality that party 1 hires a lawyer and party 2 does not. Let $\hat{\theta}$ denote the threshold which the judge uses in her default ruling strategy when s is not disclosed. Focusing as before on undominated strategies, party 1 will disclose s if and only if $s < h(\theta)$, just as in Subsection 3.2. As established in Subsection 3.1, party 2's unique optimal strategy is to disclose s if and only if $s > h(\hat{\theta})$. So, when the judge observes θ but not s, her posterior becomes

$$\mathbb{E}[g|h(\theta), h(\hat{\theta}), \theta].$$

Since this posterior is monotonic in θ and changes the sign from negative to positive at $\hat{\theta}^*(b_1, b_2)$, the following result is immediate.[13]

Proposition 3 *If only one party hires a lawyer, there exists a unique equilibrium in undominated strategies. In this equilibrium the judge uses a cutoff strategy with threshold $\hat{\theta}^*(b_1, b_2)$ (defined in*

[13]The result holds more generally in the following sense. Suppose each party randomizes on hiring a lawyer, and the judge has some arbitrary beliefs about the parties' decisions to hire lawyers. The behavior described in Proposition 3 continues to be an equilibrium in this environment.

(3)) in her default ruling. If party 1 obtains legal advice, he discloses s if and only if $g(s,\theta) < 0$. If party 1 does not obtain legal advice, reveals s if and only if $s < h(\hat{\theta}^*(b_1, b_2))$. A symmetric characterization applies to party 2. A shift in the burden of proof affects only the party without a lawyer, in a way described in Proposition 1.

3.4 Irrelevance of Representation

A striking feature of all three regimes is that the judge's equilibrium default ruling strategy is the same across all three regimes. The judge adopts a cutoff strategy with the same threshold $\hat{\theta}^*(b_1, b_2)$, regardless of whether the parties obtain legal advice. This does not mean that the parties disclose the same evidence. Propositions 1, 2 and 3 clearly show that the set of s's revealed to the judge differs across the regimes. Nevertheless, we will show that the difference in the parties' disclosure strategies does not amount to any real difference in the outcome of the trial.

This irrelevance is in fact a result of a more general property of equilibrium behavior in our disclosure/adjudication game. This property is described in the following lemma.

Lemma 1 **(Decision Equivalence)** *Suppose that the judge adopts a cutoff strategy with threshold $\hat{\theta} \in \Theta$ in her default ruling. Regardless of the legal regime, i.e. of whether either party obtains legal advice or not, any combination of the best response disclosure strategies by the two parties lead to the same outcome characterized by the following outcome function:*

$$\phi^{\hat{\theta}}(x_1, x_2, s, \theta) = \begin{cases} \mathbb{I}_{\{g(s,\theta)\geq 0\}} & \text{if } x_1 = x_2 = 1, \\ \mathbb{I}_{\{g(s,\theta)\geq 0 \text{ and } \theta \geq \hat{\theta}\}} & \text{if } x_1 = 1, x_2 = 0, \\ \mathbb{I}_{\{g(s,\theta)\geq 0 \text{ or } \theta \geq \hat{\theta}\}} & \text{if } x_1 = 0, x_2 = 1, \\ \mathbb{I}_{\{\theta \geq \hat{\theta}\}} & \text{if } x_1 = x_2 = 0, \end{cases} \tag{4}$$

where $x_i = 1$ if party i observes the evidence and $x_i = 0$) if he doesn't, and $\phi^{\hat{\theta}}(x_1, x_2, s, \theta)$ is the probability of ruling for party 2 in the state of the world (x_1, x_2, s, θ) given the adjudicator's default ruling strategy.

The Decision Equivalence Lemma shows that the judge's cutoff strategy uniquely determines the equilibrium adjudication outcome, regardless of the parties' use of legal advice. Some insight into this result can be gained from comparing the cases of no representation and full representation. Figure 3 illustrates the case in which the judge follows a threshold $\hat{\theta}$.

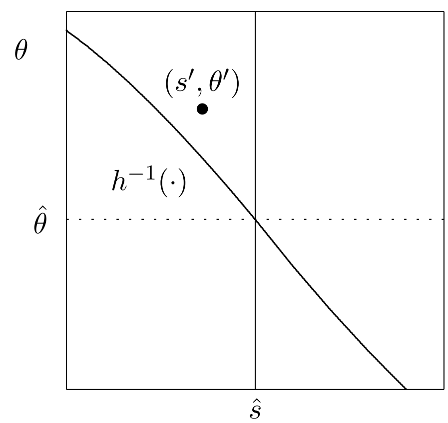

Figure 3

Then, by Propositions 1 and 2, the parties without legal advice follows a fixed threshold $\hat{s} = h(\hat{\theta})$, while the party with legal advices follows a contingent threshold $h(\hat{\theta})$. Suppose that (s', θ') occurs and party 1 observes s'. Without legal advice, party 1 will disclose s' being unaware of θ', and the judge will rule for party 2. With legal advice, party 1 would not disclose s', but the judge will nevertheless rule for party 2. Hence, despite different disclosure behavior, there is no difference in the adjudication outcome: the judge's ruling is unfavorable to party 1 in either case. Combining Propositions 1-3 with Lemma 1, we obtain our key result:

Proposition 4 **(Irrelevance of legal advice)** *Suppose that the judge applies the same inference rule, (b_1, b_2, c), regardless of the representation regime (i.e. of whether both or any party are represented or not), and the parties employ undominated strategies in disclosure. Then there is a unique equilibrium outcome which does not depend on the representation regime and is characterized by the outcome function $\phi^{\hat{\theta}^*(b_1,b_2)}(\cdot)$.*

It is worth emphasizing that the application of the same inference rule does not mean that the judge makes the same inferences from nondisclosure across different regimes. As seen above, parties employ different disclosure strategies in different regimes, which causes the judge to make different inferences from nondisclosure. Yet, the same outcome arises in all regimes. Importantly, the irrelevance result does not depend on the judge's inference rule: it holds if the judge is fully Bayesian (i.e., $(b_1, b_2) = (\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$), or if she follows any other inference rule (b_1, b_2, c). A change in (b_1, b_2) would typically cause the threshold $\hat{\theta}^*(b_1, b_2)$ to shift, but would not alter the fact that the representation regime has no effect on the adjudication outcome.

The robustness of the irrelevance result is surprising and may appear contradictory to the lawyers' prominent roles in high-profile trials. One should not take the irrelevance result as suggesting that legal representation is never useful, for lawyers perform a number of valuable tasks that are not captured by our model. Our analysis focuses on one particular aspect of legal representation — the role of lawyers as gate-keepers of information reaching the court. To the extent that this role is crucial from the information elicitation perspective, however, our irrelevance result clarifies and qualifies the sense in which lawyers can make a difference. The irrelevance result also helps us to identify the circumstances that *may make lawyer advice matter*. The next section considers one such circumstance.

4 Relevance: Withholding Favorable Evidence

Thus far, we have focused on equilibria in undominated strategies. Lawyers adopting such strategies never advise their clients to suppress ex post favorable evidence. Although such a strategy is compelling for reasons discussed above, deviating from this strategy — i.e., suppressing favorable evidence — may influence the judge's posterior and thus the ultimate ruling in favor of the lawyer's client. To gain some intuition behind this, fix the disclosure strategy of party 2, and suppose party 1 follows the strategy of never disclosing any evidence. This will influence the judge's inference from nondisclosure more favorably toward party 1, now attributing nondisclosure more to party 2's concealing evidence favorable for party 1. So, the judge's default ruling will become more favorable toward the latter.

Below, we investigate whether such behavior can occur in an equilibrium and what effect it has on the adjudication outcome. We continue to focus on equilibria in which the judge follows a cutoff strategy, which is reasonable and intuitive.[14]

To evaluate precisely how the adoption of (weakly) dominated strategies influences the judge's inference, suppose that both parties have retained lawyers, and that party 2 (or his lawyer) follows the dominant strategy of disclosing all favorable evidence, but the lawyer for party 1 advises him to withhold s regardless of its value. Given these strategies, the judge's posterior becomes $\mathbb{E}[g|0, h(\theta), \theta]$ which is less than $\mathbb{E}[g|h(\theta), h(\theta), \theta]$, and is therefore more favorable to party 1 than if both adopt their dominant strategies. Intuitively, this change in posterior reflects a shift in the judge's inference of nondisclosure, mentioned above. Likewise, if party 1 adopts the dominant strategy but party 2 adopts the strategy of never disclosing his private evidence, then the judge forms a posterior $\mathbb{E}[g|h(\theta), 1, \theta] > \mathbb{E}[g|h(\theta), h(\theta), \theta]$, which is more favorable for party 2 than if both adopt their dominant strategies. Define

$$\hat{\theta}_+(b_1, b_2, c) := \inf\{\theta \mid \mathbb{E}[g|0, h(\theta), \theta; b_1, b_2, c] > 0\};$$

$$\hat{\theta}_-(b_1, b_2, c) := \inf\{\theta \mid \mathbb{E}[g|h(\theta), 1, \theta; b_1, b_2, c] > 0\}.$$

[14] An earlier version of this paper presents other equilibria. Since they do not add any new insight, we omit them here.

Using the same argument as in Lemma A2 in the Appendix, we can show that $\mathbb{E}[g|0, h(\theta), \theta] > 0$ if and only if $\theta > \hat{\theta}_+(b_1, b_2, c)$, and $\mathbb{E}[g|h(\theta), 1, \theta] > 0$ if and only if $\theta > \hat{\theta}_-(b_1, b_2, c)$, and also that $\hat{\theta}_-(b_1, b_2, c) \leq \hat{\theta}^*(b_1, b_2) \leq \hat{\theta}_+(b_1, b_2, c)$. If $(b_1, b_2, c) > (0, 0, 0)$, then at least one inequality is strict, so the region $[\hat{\theta}_-(b_1, b_2, c), \hat{\theta}_+(b_1, b_2, c)]$ is non-degenerate.[15]

If $\theta < \hat{\theta}_-(b_1, b_2, c)$, the judge will surely rule for party 1 because the judge's posterior is negative even if she holds the most favorable beliefs about party 2 — namely that party 2 does not disclose any evidence and party 1 discloses all favorable evidence. Similarly, if $\theta > \hat{\theta}_+(b_1, b_2, c)$, then the judge will rule for party 2 in the absence of disclosure because her posterior is positive even if she holds the most favorable belief for party 1. But if $\theta \in [\hat{\theta}_-(b_1, b_2, c), \hat{\theta}_+(b_1, b_2, c)]$, then the judge's default ruling is affected by the parties' disclosure strategies, because for such θ,

$$\mathbb{E}[g|0, h(\theta), \theta] < 0 < \mathbb{E}[g|h(\theta), 1, \theta],$$

So in this case, the judge's inference and hence her ruling can go either way depending on the equilibrium disclosure strategies. It turns out that any $\hat{\theta} \in [\hat{\theta}_-(b_1, b_2, c), \hat{\theta}_+(b_1, b_2, c)]$ can be supported as a an equilibrium threshold for the judge's default ruling under full representation, as we show below.

Proposition 5 **(Relevance of legal advice with dominated strategies)** *(i) Suppose that both parties retain lawyers. Then, for any $\hat{\theta} \in [\hat{\theta}_-(b_1, b_2, c), \hat{\theta}_+(b_1, b_2, c)]$, there exists a perfect Bayesian equilibrium such that, absent disclosure, the judge rules for party 1 if $\theta < \hat{\theta}$ and for party 2 if $\theta > \hat{\theta}$. Conversely, any equilibrium cutoff default ruling strategy has a threshold in $[\hat{\theta}_-(b_1, b_2, c), \hat{\theta}_+(b_1, b_2, c)]$.*

(ii) Suppose only party 1 retains a lawyer. Then, for any $\hat{\theta} \in [\hat{\theta}^*(b_1, b_2), \hat{\theta}_+(b_1, b_2, c)]$, there exists a perfect Bayesian equilibrium such that in the absence of disclosure the judge rules for party 1 [resp. party 2] if $\theta < \hat{\theta}$. Conversely, any cutoff default ruling strategy by the judge has a threshold in $[\hat{\theta}^*(b_1, b_2), \hat{\theta}_+(b_1, b_2, c)]$. A symmetric, but opposite, characterization holds if only party 2 retains a lawyer.

Proposition 5 shows that a range of different posteriors by the judge — and thus a range of different default rulings — is sustainable when the parties adopt legal strategies of withholding favorable evidence. The reason is simple. Suppose party 1 (or his lawyer) adopts a strategy of suppressing favorable evidence $s < h(\theta)$, while party 2 follows the weakly dominant strategy. This

[15] Using equation (2), it is easy to see that $\hat{\theta}_-(b_1, b_2, c) \leq \hat{\theta}^*(b_1, b_2) \leq \hat{\theta}_+(b_1, b_2, c)$ because $\mathbb{E}[g|0, h(\theta), \theta] < \mathbb{E}[g|h(\theta), h(\theta), \theta] < \mathbb{E}[g|h(\theta), 1, \theta]$. Further, $\hat{\theta}^+(b_1, b_2, c) > \hat{\theta}^-(b_1, b_2, c)$ if $\hat{\theta}^+(b_1, b_2, c) > 0$ and $\hat{\theta}^-(b_1, b_2, c) < 1$. Given $(b_1, b_2, c) > (0, 0, 0)$, the former holds if $\mathbb{E}[g(s, 0)] < 0$ and the latter holds if $\mathbb{E}[g(s, 1)] > 0$, both of which hold by assumption.

will lead the judge to make a more favorable inference, and thus a favorable ruling, for party 1 in case of nondisclosure, which in turn justifies party 1's withholding of the favorable information. The resulting outcome with a threshold $\hat{\theta} > \hat{\theta}^*(b_1, b_2)$ is clearly more beneficial for party 1. Since such a strategy requires conditioning disclosure of evidence on the realized level of θ, it cannot be played without the legal expertise of a lawyer. In this sense, we have identified a source of private value of legal advice — namely, the ability to advise a party to withhold favorable evidence.

This benefit is most evident when only one party has access to legal advice, the case illustrated in (ii) of Proposition 5. Suppose initially neither party retains a lawyer. Then, by Proposition 1, the judge employs a threshold of $\hat{\theta}^*(b_1, b_2)$ in her default ruling. Suppose now only party 1 hires a lawyer. Provided that the lawyer can credibly advise party 1 to adopt a dominated strategy, the judge can be induced to employ a more favorable threshold in $[\hat{\theta}^*(b_1, b_2), \hat{\theta}_+(b_1, b_2, c)]$. More importantly, any equilibrium threshold in this situation is (weakly) more favorable for party 1 than $\hat{\theta}^*(b_1, b_2)$, the threshold used when neither party hires a lawyer. In this sense, legal advice becomes relevant once (weakly) dominated strategies are available. Nevertheless, if both parties hire lawyers, the parties cannot be both better off than if neither party hired a lawyer. The last fact implies that the game of hiring lawyers has the structure of a prisoner's dilemma, which may explain why both parties would hire lawyers in equilibrium.

Clearly, the relevance of legal advice rests on the credibility of weakly dominated strategy of withholding favorable evidence. One way in which a lawyer can achieve such credibility is by building a reputation through repeated trials. Casual observation suggests that lawyers are concerned about their reputations and undertake specific steps to develop/enhance them. For example, some criminal defense lawyers are known to call very few witnesses. Sometimes, a lawyer would rest the case without calling any witnesses at all if (s)he considers that the case has not been proven by the prosecutors. Our results indicate that such behavior helps to build a lawyer's reputation, which could skew the court's ruling in favor of that lawyer and his clients. From this perspective, the above result can be interpreted in terms of the relative reputation of the lawyers representing the two sides. A lawyer with a good reputation can make one better off, while a lawyer with a bad reputation can make one worse off, relative to the no representation case. Interestingly, good reputation in our context means being known for presenting limited evidence, while bad reputation is being known for presenting too much evidence, sometimes unnecessarily.

The interval, $[\hat{\theta}_-(b_1, b_2, c), \hat{\theta}_+(b_1, b_2, c)]$, can be interpreted as a range of manipulation of the outcomes by the lawyers. It is worth noting that this range depends on the inference rule (b_1, b_2, c). In particular, this range vanishes as $(b_1, b_2, c) \to (0, 0, 0)$; i.e., as the judge's inference rule becomes completely non-Bayesian. This feature points to a possible rationale for regulating the judge's inference rule — a topic addressed more fully in the next section.

5 Welfare implications of legal advice and Evidence Laws

In this section, we consider the welfare implications of our results. Recall that the judge's inference rule, (b_1, b_2, c), influences the parties' disclosure behavior, and ultimately determines the equilibrium outcome. It is thus important to understand how the inference rule affects the welfare and how it should be optimally chosen.

Of particular interest is whether the Bayesian inference rule $(b_1, b_2, c) = (\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}})$ is optimal. This issue has special policy relevance for the law of evidence which imposes various restrictions on the inferences that adjudicators are allowed to make from evidence or from lack thereof. For instance, the burden of proof is sometimes assigned to one of the parties (e.g., the plaintiff); in criminal cases a negative inference cannot be made when the defendant refuses to testify; a certain inference is to be made even with insufficient evidence (i.e. prima facie rules); and self-interest alone does not constitute a reason for discounting evidence (Daughety and Reinganum, 2000a, 2000b; Posner, 1999). If the Bayesian inference rule is optimal, then such restrictions would not be necessary, for rational adjudicators would strive to adjudicate in a Bayesian fashion any way. If not, some restrictions may be warranted and it is important to investigate the reason.

Our model allows us to address this issue in a simple fashion. It is natural to take the judge's objective as our welfare criterion. When maximizing this criterion, we can treat the judge's threshold as the only choice variable since, by Decision Equivalence (Lemma 1), the judge's threshold completely pins down the outcome. So, we can formulate our welfare inquiry as the following mechanism design problem:

$$[WP] \qquad \max_{\hat{\theta}\in\Theta} \sum_{i,j=0,1} p_{ij}\mathbb{E}\left[\phi^{\hat{\theta}}(i,j,s,\theta)g(s,\theta)\right],$$

where $\phi^{\hat{\theta}}$ is the outcome induced by the cutoff rule with threshold $\hat{\theta}$ (see Lemma 1). This objective function puts a positive value on the ruling for party 2 if and only if party 1 is truly culpable, i.e. $g(s,\theta) > 0$ (see the discussion in Section 2). The analysis of $[WP]$ yields the following result.

Proposition 6 *A cutoff ruling strategy with a Bayesian threshold, $\hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$, is socially optimal in the sense it solves $[WP]$.*

In light of Lemmas 1-3, the optimal threshold $\hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$ can be implemented by a laissez-faire policy in any legal regime, as long as the partie(s) with a legal advice plays a weakly dominant strategy, for any rational judge would use the Bayesian inference rule. In this sense, Proposition 6 argues against regulating adjudicators' inferences.

The same conclusion does not hold, however, if lawyers can credibly use the strategies of not disclosing all favorable evidence. Proposition 5 shows that, in this case, any threshold $\hat{\theta} \in$

$[\hat{\theta}_-(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}}), \hat{\theta}_+(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}})]$ can arise in equilibrium even under Bayesian inference rule. Thus, if a lawyer, due to his reputation, credibly uses a disclosure strategy which shifts the threshold away from $\hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$, then legal representation is privately valuable but moves the outcome away from social optimum. So legal advice is socially harmful in this case.

In light of this result, the following proposition provides a possible justification for regulating adjudicators' inferences.

PROPOSITION 7 *(i) If $(p_{10}, p_{01}) \gg (0,0)$, then there exists $(b_1, b_2) \ll (\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$ with either $b_1 = 0$ or $b_2 = 0$ such that $\hat{\theta}^*(b_1, b_2) = \hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$. (ii) There exists (b_1, b_2) with either $b_1 = 0$ or $b_2 = 0$, such that*

$$\hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}) \in [\hat{\theta}_-(b_1, b_2, 0), \hat{\theta}_+(b_1, b_2, 0)] \subset [\theta_-(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}}), \theta_+(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}})]. \tag{5}$$

The first part of Proposition 7 shows that the outcome attained under the Bayesian inference rule can be replicated by an inference rule that assigns no burden of proof to one of the parties, provided that the parties follows undominated strategies. This finding is consistent with the evidence law that often assigns the entire burden of proof to one of the parties, e.g. the plaintiff. It is interesting that an one-sided burden-of-proof allocation can replicate the adjudication outcome attainable under the Bayesian rule. The reason is that, as observed earlier, the judge's threshold rises with b_1 and falls with b_2. Hence, starting from $(b_1, b_2) = (\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$, there is a way to lower both coefficients without altering the threshold, until the lower of the two becomes zero.

The second part of the proposition develops a sense in which such a burden-of-proof allocation may be socially desirable. Lower values of (b_1, b_2) do not alter the judge's threshold but reduce the scope of the judge's discretion in adjusting her inferences based on the parties' disclosure strategies. In this sense, such a burden-of-proof allocation makes the judge's ruling less susceptible to manipulation by the lawyers, compared with the Bayesian rule $(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$. In fact, as noted in the previous section, the range of equilibrium outcomes around the optimal outcome shrinks as (b_1, b_2) falls. These two findings imply that committing adjudicators to an inference rule different from the Bayesian could be socially desirable.[16]

6 Conclusion

In this paper, we have studied the effect of legal representation on the adjudication outcomes. Our analysis was concerned with the role of lawyers as gate-keepers of information reaching the court. We have compared outcomes under legal representation and self-representation, and have shown

[16]These arguments for regulating judges' inferences differ from the existing ones based on the cost of evidence production (see Hay and Spier (1997) for example) or on the ex ante deterrence (see Bernardo et al. (2000)).

that legal representation does not affect the adjudication outcome if legal advice is costless and the lawyers cannot credibly suppress favorable information. At the same time, we have shown that lawyers can affect the adjudication outcome if they can leverage their legal expertise to engage in more sophisticated strategic behavior. They can do so by credibly withholding some favorable information and thereby affecting the judge's inference in their clients favor when evidence is not disclosed.

Besides the lawyers' role in disclosure, this paper yields useful insights regarding several related issues. First, our findings help to understand the implications of quality differences in legal advice. Recent legal cases have generated a concern that high-profile lawyers may influence the outcome to the point of jeopardizing fair adjudication. Our model can be used to understand the effect of quality differences in lawyering, once we interpret self-representation as representation by an inexperienced lawyer. Second, our study also helps to understand various rules regulating the adjudicators interpretation of evidence and restrictions on inferences they can or cannot draw from nondisclosure of evidence.

More broadly, our results could shed light on the role of advising in settings outside legal disputes. Agents and divisions often compete for resources within organizations, and resource allocating decisions — which could take such forms as merit assignment, promotion for workers, budget allocation between divisions — often depend on the information provided by those closely affected by the decisions. Advising agents regarding disclosure of information can affect both the quality of information transmission and the resource allocation decision itself. Our results offer some basic insight on the role of advising in such circumstances.

7 Appendix: Proofs

We first establish several lemmas that will be used in the proofs.

Lemma A1 *For any $\theta' > \theta$*

$$\int_0^{\hat{s}} g(s,\theta) f(s|\theta') ds \geq \min\left\{0, \int_0^{\hat{s}} g(s,\theta) f(s|\theta) ds\right\}.$$

Proof: Both $\frac{F(s|\theta)}{F(\hat{s}|\theta)}$ and $\frac{F(s|\theta')}{F(\hat{s}|\theta')}$ are cdf's on $[0,\hat{s}]$. By MLRP, $\frac{F(s|\theta')}{F(\hat{s}|\theta')}$ first-order stochastically dominates $\frac{F(s|\theta)}{F(\hat{s}|\theta)}$. Therefore, since $g(s,\theta)$ is increasing in s, we have: $\int_0^{\hat{s}} g(s,\theta) \frac{f(s|\theta')}{F(\hat{s}|\theta')} ds \geq \int_0^{\hat{s}} g(s,\theta) \frac{f(s|\theta)}{F(\hat{s}|\theta)} ds$, which can be rewritten as: $\int_0^{\hat{s}} g(s,\theta) f(s|\theta') ds \geq \frac{F(\hat{s}|\theta')}{F(\hat{s}|\theta)} \int_0^{\hat{s}} g(s,\theta) f(s|\theta) ds$. The latter inequality implies that, if $\int_0^{\hat{s}} g(s,\theta) f(s|\theta) ds \geq 0$, then $\int_0^{\hat{s}} g(s,\theta) f(s|\theta') ds \geq 0$. Also, note that $\frac{F(\hat{s}|\theta')}{F(\hat{s}|\theta)} \leq 1$. Hence, if $\int_0^{\hat{s}} g(s,\theta) f(s|\theta) ds \leq 0$, then $\int_0^{\hat{s}} g(s,\theta) f(s|\theta') ds \geq \int_0^{\hat{s}} g(s,\theta) f(s|\theta) ds$. ▮

Lemma A2 *Fix any $\hat{s} \in [0,1]$. If $\mathbb{E}[g|\hat{s},\hat{s},\theta] \geq 0$, then $\mathbb{E}[g|\hat{s},\hat{s},\theta'] > 0$ for any $\theta' > \theta$.*

Proof: Recall that

$$\mathbb{E}[g|\hat{s},\hat{s},\theta] := \int_0^1 g(s,\theta)f(s|\theta)ds + b_1 \int_{\hat{s}}^1 g(s,\theta)f(s|\theta)ds + b_2 \int_0^{\hat{s}} g(s,\theta)f(s|\theta)ds. \tag{6}$$

The result follows from several observations. By MLRP and monotonicity of $g(\cdot,\cdot)$, for $\theta' > \theta$,

$$\int_0^1 g(s,\theta')f(s|\theta')ds > \int_0^1 g(s,\theta)f(s|\theta)ds \tag{7}$$

and

$$\int_{\hat{s}}^1 g(s,\theta')f(s|\theta')ds > \int_{\hat{s}}^1 g(s,\theta)f(s|\theta)ds. \tag{8}$$

To get the second inequality, rewrite the set restriction as $\mathbb{I}_{\{s>\hat{s}\}}$, an increasing function of s. Therefore, the first term $\int_0^1 g(s,\theta)f(s|\theta)ds$ is strictly increasing in θ, and the second term $b_1 \int_{\hat{s}_1}^1 g(s,\theta)f(s|\theta)ds$ is nondecreasing in θ.

Now consider the third term. Suppose first that $\int_0^{\hat{s}} g(s,\theta)f(s|\theta)ds \geq 0$. Then by Lemma A1, $\int_0^{\hat{s}} g(s,\theta')f(s|\theta')ds \geq 0$. This, together with (7) and (8), implies that $\mathbb{E}[g|\hat{s},\hat{s},\theta'] > 0$.

Suppose next $\int_0^{\hat{s}} g(s,\theta)f(s|\theta)ds < 0$. Then, by Lemma A1 $\int_0^{\hat{s}} g(s,\theta')f(s|\theta')ds \geq \int_0^{\hat{s}} g(s,\theta)f(s|\theta)ds$. Combining this fact with (7) and (8), we again conclude that $\mathbb{E}[g|\hat{s},\hat{s},\hat{\theta}'] > 0$. ∎

Lemma A3 *If $\mathbb{E}[g|h(\theta),h(\theta),\theta] \geq 0$, then $\mathbb{E}[g|h(\theta'),h(\theta'),\theta'] > 0$ for any $\theta' > \theta$.*

Proof: Recall that

$$\mathbb{E}[g|h(\theta),h(\theta),\theta] = \int_0^1 g(s,\theta)f(s|\theta)ds + b_1 \int_{h(\theta)}^1 g(s,\theta)f(s|\theta)ds + b_2 \int_0^{h(\theta)} g(s,\theta)f(s|\theta)ds. \tag{9}$$

The first term of (9) is increasing in θ by (7). Now consider the second term. We have:

$$\int_{h(\theta')}^1 g(s,\theta')f(s|\theta')ds = \int_{h(\theta)}^1 g(s,\theta')f(s|\theta')ds + \int_{h(\theta')}^{h(\theta)} g(s,\theta')f(s|\theta')ds > \int_{h(\theta)}^1 g(s,\theta)f(s|\theta)ds, \tag{10}$$

where the inequality follows from (8) and the fact that $h(\theta') < h(\theta)$ and $g(s,\theta') \geq 0$ if $s > h(\theta')$.

Now consider the third term. Since $\int_0^{h(\theta)} g(s,\theta)f(s|\theta)ds < 0$, and $g(s,\theta)$ is strictly increasing in θ, Lemma A1 implies that

$$\int_0^{h(\theta)} g(s,\theta')f(s|\theta')ds > \int_0^{h(\theta)} g(s,\theta)f(s|\theta)ds. \tag{11}$$

Differentiating $\int_0^{h(\theta)} g(s,\theta)f(s|\theta)$ with respect to θ, we have

$$h'(\theta)g(h(\theta),\theta)f(h(\theta)|\theta) + \frac{d}{d\tilde{\theta}}\left[\int_0^{h(\theta)} g(s,\tilde{\theta})f(s|\tilde{\theta})ds\right]_{\tilde{\theta}=\theta} \geq 0,$$

since the first term vanishes and the second term is nonnegative by (11).

Combining the observations, we conclude that (9) is strictly increasing in θ. ▮

Proof of Proposition 1: The proof consists of several steps.

Step 1: *In any equilibrium, parties 1 and 2 use cutoff strategies with the same threshold, i.e. there exists $\hat{s}$ such that party 1 discloses evidence s if (only if) $s < (\leq)\hat{s}$, and party 2 discloses evidence s if (only if) $s > (\geq)\hat{s}$.*

Proof: Fix any equilibrium and suppose that party 1 has observed evidence s. Party 1's disclosure decision affects the outcome of the trial only if party 2 does not disclose the evidence. Let $P_0(s)$ denote the probability that the judge rules for party 1 in that equilibrium if s is not disclosed. This probability depends on s, because the judge's decision depends only on the value of θ, and s and θ are (weakly) affiliated. On the other hand, if party 1 discloses s, then the judge will rule for party 1 if $g(s,\theta) < 0$, or $\theta < h^{-1}(s)$. Thus, party 1 discloses s in that equilibrium if (only if)

$$P_0(s) < (\leq) \Pr\{\theta < h^{-1}(s) \mid s\}. \tag{12}$$

Similarly, party 2 discloses s if (only if)

$$P_0(s) > (\geq) \Pr\{\theta < h^{-1}(s) \mid s\}. \tag{13}$$

Thus, the disclosure incentives of the two parties are precisely the opposite.

To establish that parties 1 and 2 use cutoff strategies in any equilibrium, we show that, for any $s' > s$, $P_0(s') > \Pr\{\theta < h^{-1}(s') \mid s'\}$ if $P_0(s) = \Pr\{\theta < h^{-1}(s) \mid s\}$. Suppose the judge follows a default ruling strategy, $\delta(\theta)$, i.e. she rules for party 2 with probability $\delta(\theta)$ given θ and nondisclosure. Then, we have:

$$\begin{aligned}
P_0(s') &\equiv \int_0^1 (1-\delta(\theta))f(\theta|s')d\theta \\
&= \int_0^{h^{-1}(s)} (1-\delta(\theta))f(\theta|s')d\theta + \int_{h^{-1}(s)}^1 (1-\delta(\theta))\frac{f(\theta|s')}{f(\theta|s)}f(\theta|s)d\theta \\
&\geq \int_0^{h^{-1}(s)} (1-\delta(\theta))f(\theta|s')d\theta + \int_{h^{-1}(s)}^1 (1-\delta(\theta))\frac{f(h^{-1}(s)|s')}{f(h^{-1}(s)|s)}f(\theta|s)d\theta \\
&= \int_0^{h^{-1}(s)} (1-\delta(\theta))f(\theta|s')d\theta + \int_0^{h^{-1}(s)} \delta(\theta)\frac{f(h^{-1}(s)|s')}{f(h^{-1}(s)|s)}f(\theta|s)]d\theta \\
&\geq \int_0^{h^{-1}(s)} (1-\delta(\theta))f(\theta|s')d\theta + \int_0^{h^{-1}(s)} \delta(\theta)\frac{f(\theta|s')}{f(\theta|s)}f(\theta|s)]d\theta \\
&= \int_0^{h^{-1}(s)} f(\theta|s')d\theta = \Pr\{\theta < h^{-1}(s) \mid s'\} > \Pr\{\theta < h^{-1}(s') \mid s'\}.
\end{aligned}$$

The first and the last two equalities in this sequence hold by definition. The two non-strict inequalities hold by MLRP. The equality between them holds because

$$P_0(s) = \Pr\{\theta < h^{-1}(s) \mid s\} \quad \Leftrightarrow \quad \int_0^1 (1-\delta(\theta))f(\theta|s)d\theta = \int_0^{h^{-1}(s)} f(\theta|s)d\theta$$
$$\Leftrightarrow \quad \int_{h^{-1}(s)}^1 (1-\delta(\theta))f(\theta|s)d\theta = \int_0^{h^{-1}(s)} \delta(\theta)f(\theta|s)d\theta.$$

The lone strict inequality holds because $h^{-1}(\cdot)$ is strictly decreasing, and s and θ are affiliated.

A symmetric argument establishes that, for all $s'' < s$, $P_0(s'') < \Pr\{\theta < h^{-1}(s'') \mid s''\}$ if $P_0(s) = \Pr\{\theta < h^{-1}(s) \mid s\}$.

In combination, these results imply the existence of a common threshold $\hat{s} \in [0,1]$ s.t. party 1 discloses (withholds) s if $s < \hat{s}$ ($s > \hat{s}$) and party 2 discloses (withholds) s if $s > \hat{s}$ ($s < \hat{s}$).[17] ▌

Step 2: *In any equilibrium, the judge follows a cutoff strategy in her default ruling; i.e., there exists $\hat{\theta}$ such that $\delta(\theta) = 0$ if $\theta < \hat{\theta}$ and $\delta(\theta) = 1$ if $\theta > \hat{\theta}$.*

Proof: By Step 1, the parties follow cutoff disclosure strategies with some common threshold $\hat{s}$. Hence, the judge's posterior on party 1's culpability when she observes θ is given by $\mathbb{E}[g|\hat{s},\hat{s},\theta]$. Then, by Lemma A2, there exists $\hat{\theta} \in \Theta$ such that $\delta(\theta) = 0$ if $\theta < \hat{\theta}$ and $\delta(\theta) = 1$ if $\theta > \hat{\theta}$. ▌

Step 3: *If $\hat{s}$ is the parties' common threshold and $\hat{\theta}$ is the judge's threshold, then $\hat{s} = h(\hat{\theta})$.*

Proof: Since the parties' strategies must constitute best responses to the judge's default ruling strategy with threshold $\hat{\theta}$, we must have

$$P_0(s) = \Pr\{\theta < \hat{\theta} \mid s\}.$$

Hence, the optimality of the cutoff strategies with threshold $\hat{s}$, together with (12) and (13), implies that $\Pr\{\theta < \hat{\theta} \mid s\} < (\leq) \Pr\{\theta < h^{-1}(s) \mid s\}$ if (only if) $s < (\leq) \hat{s}$. Similarly, $\Pr\{\theta < \hat{\theta} \mid s\} > (\geq) \Pr\{\theta < h^{-1}(s) \mid s\}$ if (only if) $s > (\geq) \hat{s}$. Therefore, $\hat{s} = h(\hat{\theta})$. ▌

Step 4: *It is an equilibrium for the judge to follow a cutoff strategy with threshold $\hat{\theta}^*$ and for the parties to follow cutoff strategies with a common threshold $h(\hat{\theta}^*)$.*[18]

Proof: Recall from (3) that

$$\hat{\theta}^* := \inf\{\theta \in \Theta \mid \mathbb{E}[g|h(\theta), h(\theta), \theta] > 0\}.$$

It then follows from Lemma A2 that

$$\mathbb{E}[g|h(\hat{\theta}^*), h(\hat{\theta}^*), \theta] \gtrless 0 \quad \text{if} \quad \theta \gtrless \hat{\theta}^*.$$

[17] If some party, say party 1, has a strict incentive for disclosing all s, then the statement remains valid with $\hat{s} = 1$.

[18] For brevity, we omit the dependence of $\hat{\theta}^*$ on b_1, b_2 and c.

So, the judge's cutoff strategy with threshold $\hat{\theta}^*$ is optimal when the parties adopt cutoff strategies with common threshold $h(\hat{\theta}^*)$. Likewise, Steps 1 and 3 show that the parties' cutoff strategies with common threshold $h(\hat{\theta}^*)$ are best responses to the judge's cutoff strategy with threshold $\hat{\theta}^*$. Hence, this strategy profile constitutes a perfect Bayesian equilibrium. ▌

STEP 5: *The equilibrium described in Step 4 is unique .*

PROOF: The uniqueness follows from the uniqueness of the judge's threshold, which in turn follows from Lemma A3. ▌

PROOF OF PROPOSITION 2: The weak dominance of the parties' disclosure strategies is already established in the text. Given the disclosure strategies, when the judge observes θ her posterior of party 1's culpability is given by $\mathbb{E}[g|h(\theta), h(\theta), \theta]$. Recall that

$$\hat{\theta}^* := \inf\{\theta \in \Theta \,|\, \mathbb{E}[g|h(\theta), h(\theta), \theta] > 0\}.$$

Lemma A3 then implies that

$$\mathbb{E}[g|h(\theta), h(\theta), \theta] \underset{<}{>} 0 \quad \text{if} \quad \theta \underset{<}{>} \hat{\theta}^*,$$

proving that the judge's cutoff default ruling strategy with threshold $\hat{\theta}^*$ is optimal. The uniqueness of equilibrium follows from the uniqueness of the equilibrium threshold, which in turn follows from Lemma A3 and the definition of $\hat{\theta}^*$. ▌

PROOF OF PROPOSITION 3: Suppose without loss of generality that party 1 has hired a lawyer but party 2 has not. (The opposite case is completely symmetric.) Then, party 1 has a dominant strategy of disclosing (withholding) s if $s > h(\theta)$ ($s < h(\theta)$). Just as in Proposition 1, party 2 will adopt a cutoff strategy with some threshold $\hat{s} \in S$.

Consider next the judge's default ruling strategy. Given θ and nondisclosure of s, the judge's posterior becomes $\mathbb{E}[g|h(\theta), \hat{s}, \theta]$. Lemmas A1 and A2 imply that this posterior is ordinally monotonic: i.e., $\mathbb{E}[g|h(\theta), \hat{s}, \theta] \geq 0$ implies $\mathbb{E}[g|h(\theta'), \hat{s}, \theta'] > 0$ for $\theta' > \theta$. Hence, the judge adopts a cutoff strategy with some threshold $\hat{\theta}$. Then, the same argument as in Proposition 1 can be used to establish that $\hat{s} = h(\hat{\theta})$. It then follows that $\hat{\theta} = \hat{\theta}^*$. Further, the equilibrium threshold $\hat{\theta}^*$ is unique by Lemma A2. ▌

PROOF OF LEMMA 1: Suppose that the judge follows a default ruling strategy of a cutoff form with a threshold $\hat{\theta} \in \Theta$. Let us show that any combination of the parties' best response disclosure strategies lead to the same outcome, regardless of whether either party has obtained legal advice or not. To begin, given the threshold $\hat{\theta}$, let $\mathcal{S}_1^{\hat{\theta}}$ be a set of party 1's disclosure strategies such that, if $\rho_1(s, \theta) \in \mathcal{S}_1^{\hat{\theta}}$, then $\rho_1(s, \theta) = 1$ for almost every (s, θ) with $\theta > \hat{\theta}$ and $g(s, \theta) < 0$, and

$\rho_1(s,\theta) = 0$ for almost every (s,θ) with $\theta < \hat{\theta}$ and $g(s,\theta) > 0$. Similarly, let $\mathcal{S}_2^{\hat{\theta}}$ be the set of disclosure strategies for party 2 such that, if $\rho_2(s,\theta) \in \mathcal{S}_2^{\hat{\theta}}$, then $\rho_2(s,\theta) = 0$ for almost every (s,θ) with $\theta > \hat{\theta}$ and $g(s,\theta) < 0$, and $\rho_2(s,\theta) = 1$ for almost every (s,θ) with $\theta < \hat{\theta}$ and $g(s,\theta) > 0$. In words, a party $i = 1,2$ following a strategy in $\mathcal{S}_i^{\hat{\theta}}$ will always present evidence that would overturn an unfavorable default ruling, and would never present evidence that will overturn a favorable ruling. Such a strategy is optimal for each party, regardless of the opponent's disclosure strategy. If the latter discloses then a party's strategy has no effect, whereas if the opponent does not disclose, then no other strategy can make the party strictly better off.

Note that, if the judge follows a default ruling strategy with a cutoff threshold $\hat{\theta}$, then any pair of parties' disclosure strategies $(\rho_1,\rho_2) \in \mathcal{S}_1^{\hat{\theta}} \times \mathcal{S}_2^{\hat{\theta}}$, induces the outcome in (4).

If a party $i = 1,2$ has obtained a legal advice, clearly all strategies in $\mathcal{S}_i^{\hat{\theta}}$ are feasible. Importantly, party i without legal advice also has access to a strategy in $\mathcal{S}_i^{\hat{\theta}}$. This can be seen by the fact that a simple cutoff strategy $\hat{\rho}_1(s,\theta) := \mathbb{I}_{\{s<h(\hat{\theta})\}}$ does not depend on $\hat{\theta}$ (so it is a feasible strategy for party 1 without lawyer advice), yet it belongs to $\mathcal{S}_1^{\hat{\theta}}$. Likewise, $\hat{\rho}_2(s,\theta) := \mathbb{I}_{\{s>h(\hat{\theta})\}}$ is feasible for party 2 when he has no legal advice but it belongs to $\mathcal{S}_2^{\hat{\theta}}$.

Finally, to complete the proof, fix any legal regime, and suppose $\rho_i(s,\theta)$ is party i's best response to some strategy of player j and the judge's threshold strategy $\hat{\theta}$. Then, we must have $\rho_i \in \mathcal{S}_i^{\hat{\theta}}$. Or else, one can show that the strategy is strictly worse for party i than the simple cutoff strategy $\hat{\rho}_i(s,\theta)$, which is available for that party in every legal regime. The argument for this result is essentially the same as the one provided prior to Proposition 1. ▌

Proof of Proposition 5: Before proceeding, it is useful to establish some preliminary results. The arguments employed in Lemmas A2 and A3 can be combined to show that $\forall \theta' > \theta$:

$$\mathbb{E}[g|0,h(\theta),\theta] \geq 0 \Longrightarrow \mathbb{E}[g|0,h(\theta'),\theta'] > 0$$
$$\mathbb{E}[g|h(\theta),1,\theta] \geq 0 \Longrightarrow \mathbb{E}[g|h(\theta'),1,\theta'] > 0.$$

From these, it follows that $\mathbb{E}[g|0,h(\theta),\theta] > 0$ if and only if $\theta > \hat{\theta}_+(b_1,b_2,c)$, and $\mathbb{E}[g|h(\theta),1,\theta] > 0$ if and only if $\theta > \hat{\theta}_-(b_1,b_2,c)$.

We first prove (i). Fix any $\hat{\theta} \in [\hat{\theta}_-(b_1,b_2,c), \hat{\theta}_+(b_1,b_2,c)]$. We shall prove that there exists an equilibrium in which the judge adopts a cutoff strategy with threshold $\hat{\theta}$. In this equilibrium, party 1 discloses s if and only if $s < h(\theta)$ and $\theta > \hat{\theta}$, whereas party 2 discloses s if and only if $s > h(\theta)$ and $\theta < \hat{\theta}$. Given these disclosure strategies, the judge's posterior becomes $\mathbb{E}[g|0,h(\theta),\theta] < 0$ if $\theta < \hat{\theta}$ and $\mathbb{E}[g|h(\theta),1,\theta] > 0$ if $\theta > \hat{\theta}$. Hence, it is optimal for the judge to rule for party 1 if and only if $\theta < \hat{\theta}$. Given the default ruling by the judge, party i's $(i = 1,2)$ disclosure strategy is in $\mathcal{S}_i^{\hat{\theta}}$ and hence constitutes a best response. The first statement is thus proven.

Next, consider the converse. To prove that any equilibrium threshold $\hat{\theta}$ must be in $[\hat{\theta}_-(b_1,b_2,c),\hat{\theta}_+(b_1,b_2,c)]$, suppose otherwise i.e. there exists an equilibrium strategy combination $(\hat{\theta},\rho_1,\rho_2)$ s.t. $\hat{\theta}\notin[\hat{\theta}_-(b_1,b_2,c),\hat{\theta}_+(b_1,b_2,c)]$. At first, consider $\hat{\theta}>\hat{\theta}_+(b_1,b_2,c)$. Then $\mathbb{E}[g|\rho_1,\rho_2,\tilde{\theta}]\leq 0$ for an arbitrary $\tilde{\theta}\in(\hat{\theta}_+(b_1,b_2,c),\hat{\theta})$, where

$$\mathbb{E}[g|\rho_1,\rho_2,\tilde{\theta}]=\int_0^1 g(s,\tilde{\theta})f(s|\tilde{\theta})ds+b_1\int_0^1(1-\rho_1(s,\tilde{\theta}))g(s,\tilde{\theta})f(s|\tilde{\theta})ds+b_2\int_0^1(1-\rho_2(s,\theta))g(s,\tilde{\theta})f(s|\tilde{\theta})ds.$$

Note that both parties never disclose the same s in equilibrium, because (ρ_1,ρ_2) are best response strategies and therefore $\rho_i\in S_i^{\hat{\theta}}$ for $i\in\{1,2\}$ (see the proof of Lemma 1). So the term associated with c disappears.

Let us compare $\mathbb{E}[g|\rho_1,\rho_2,\tilde{\theta}]$ and $E[g|0,h(\tilde{\theta}),\tilde{\theta}]$ term by term. First, $\int_0^1(1-\rho_2(s,\tilde{\theta}))g(s,\tilde{\theta})f(s|\tilde{\theta})ds\geq\int_0^{h(\theta)}g(s,\tilde{\theta})f(s|\tilde{\theta})ds$. Also, by Lemma 1, $\rho_1(s,\theta)=0$ if $g(s,\theta)>0$. Hence,

$$\int_0^1(1-\rho_1(s,\tilde{\theta}))g(s,\tilde{\theta})f(s|\tilde{\theta})ds\geq\int_0^1 g(s,\tilde{\theta})f(s|\tilde{\theta})ds.$$

Therefore, $\mathbb{E}[g|\rho_1,\rho_2,\tilde{\theta}]>0$ if $E[g|0,h(\tilde{\theta}),\tilde{\theta}]>0$. Since $\tilde{\theta}>\hat{\theta}_+(b_1,b_2,c)$, $E[g|0,h(\tilde{\theta}),\tilde{\theta}]>0$, so $\mathbb{E}[g|\rho_1,\rho_2,\tilde{\theta}]>0$ - a contradiction. Hence, $\hat{\theta}\leq\hat{\theta}_+(b_1,b_2,c)$. A symmetric argument proves that $\hat{\theta}\geq\hat{\theta}_-(b_1,b_2,c)$.

Turning now to part (ii), let us without loss of generality focus on the case in which only party 1 obtains legal advice. Choose any $\hat{\theta}\in[\hat{\theta}^*(b_1,b_2),\hat{\theta}_+(b_1,b_2,c)]$. Consider the disclosure strategies whereby party 1 discloses s if and only if $s<h(\theta)$ and $\theta>\hat{\theta}$ and party 2 (who does not have legal advice) discloses s if and only if $s>h(\hat{\theta})$. This pair of strategies constitute best responses given the judge's threshold $\hat{\theta}$. Under these disclosure strategies, the judge's posterior is $\mathbb{E}[g|0,h(\hat{\theta}),\theta]$ if $\theta<\hat{\theta}$ and $\mathbb{E}[g|h(\theta),h(\hat{\theta}),\theta]$ if $\theta>\hat{\theta}$. But $\mathbb{E}[g|0,h(\hat{\theta}),\theta]<0$ and $\mathbb{E}[g|h(\theta),h(\hat{\theta}),\theta]>0$ because $\hat{\theta}\in[\hat{\theta}^*(b_1,b_2),\hat{\theta}_+(b_1,b_2,c)]$. Hence, the judge's cutoff strategy is optimal. The proof of the converse is analogous to that for part (i) and is therefore omitted. ∎

Proof of Proposition 6: The objective function of $[WP]$ can be rewritten as follows:

$$p_{11}\int_{s>h(\theta)}g(s,\theta)f(s|\theta)l(\theta)dsd\theta+p_{10}\int_{\theta>\hat{\theta}}\int_{s>h(\theta)}g(s,\theta)f(s|\theta)l(\theta)dsd\theta+$$
$$p_{01}\left\{\int_{\theta>\hat{\theta}}\int_0^1 g(s,\theta)f(s|\theta)l(\theta)dsd\theta+\int_{\theta\leq\hat{\theta}}\int_{s>h(\theta)}g(s,\theta)f(s|\theta)l(\theta)dsd\theta\right\}+p_{00}\int_{\theta>\hat{\theta}}\int_0^1 g(s,\theta)f(s|\theta)l(\theta)dsd\theta,$$

where $l(\cdot)$ is the marginal density of θ.

Differentiating this with respect to $\hat{\theta}$ yields

$$-p_{00}\int_0^1 g(s,\hat{\theta})f(s|\theta)dsl(\hat{\theta})-p_{10}\int_{h(\hat{\theta})}^1 g(s,\hat{\theta})f(s|\hat{\theta})dsl(\hat{\theta})-p_{01}\int_0^{h(\hat{\theta})}g(s,\hat{\theta})f(s|\hat{\theta})dsl(\hat{\theta})$$
$$\gtrless 0 \quad \text{if} \quad \hat{\theta}\lessgtr\hat{\theta}^*\left(\tfrac{p_{10}}{p_{00}},\tfrac{p_{01}}{p_{00}}\right).$$

So the objective function of $[WP]$ attains its maximum at $\hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$. ▮

Proof of Proposition 7: To prove (i), let $\bar{\theta}^* := \hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$. Also, let

$$B := \left(\frac{p_{10}}{p_{00}}\right) \int_{h(\bar{\theta}^*)}^{1} g(s, \bar{\theta}^*) f(s|\bar{\theta}^*) ds + \left(\frac{p_{01}}{p_{00}}\right) \int_{0}^{h(\bar{\theta}^*)} g(s, \bar{\theta}^*) f(s|\bar{\theta}^*) ds.$$

If $B > 0$, set $b_1 = \hat{b}_1 \equiv \frac{B}{\int_{h(\bar{\theta}^*)}^{1} g(s,\bar{\theta}^*) f(s|\bar{\theta}^*) ds}$. Since $p_{01} > 0$, we have $0 < \hat{b}_1 < \frac{p_{10}}{p_{00}}$. Furthermore,

$$\begin{aligned}\mathbb{E}[g|h(\bar{\theta}^*), h(\bar{\theta}^*), \bar{\theta}^*; \hat{b}_1, 0, 0] &= \int_0^1 g(s, \bar{\theta}^*) f(s|\bar{\theta}^*) ds + \hat{b}_1 \int_{h(\bar{\theta}^*)}^{1} g(s, \bar{\theta}^*) f(s|\bar{\theta}^*) ds \\ &= \int_0^1 g(s, \bar{\theta}^*) f(s|\bar{\theta}^*) ds + B = \mathbb{E}[g|h(\bar{\theta}^*), h(\bar{\theta}^*), \bar{\theta}^*; \tfrac{p_{10}}{p_{00}}, \tfrac{p_{01}}{p_{00}}, \tfrac{p_{11}}{p_{00}}].\end{aligned}$$

This, together with Lemma A3, proves that

$$\hat{\theta}^*(\hat{b}_1, 0) = \bar{\theta}^* = \hat{\theta}^*(\tfrac{p_{10}}{p_{00}}, \tfrac{p_{01}}{p_{00}}),$$

as was to be shown. The case with $B < 0$ is treated symmetrically with $b_1 = 0$.

We now prove (ii). Without loss of generality, assume $B > 0$. Consider any inference rule $(b_1, 0, 0)$ with $b_1 > 0$. Clearly, $\mathbb{E}[g|0, h(\theta), \theta; b_1, 0, 0] > 0$ whenever $\mathbb{E}[g|0, h(\theta), \theta; \frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}}] \geq 0$. Hence, $\hat{\theta}_+(b_1, 0, 0) < \hat{\theta}_+(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}})$.

Recall that $\hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}) = \hat{\theta}^*(\hat{b}_1, 0) \geq \hat{\theta}_-(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}}, \frac{p_{11}}{p_{00}})$ and that $\hat{\theta}_+(b_1, 0, 0) - \hat{\theta}_-(b_1, 0, 0)$ monotonically converges to zero as b_1 gets small. Hence, there exists b_1 such that $(b_1, 0, 0)$ satisfies (5) in the statement of the Proposition. The case with $B < 0$ can be treated symmetrically. Finally, if $B = 0$, then $\hat{\theta}_-(0, 0, 0) = \hat{\theta}_+(0, 0, 0) = \hat{\theta}^*(0, 0) = \hat{\theta}^*(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$, so (5) is satisfied with the inference rule $(0, 0, 0)$. ▮

References

Bernardo, A.E., Talley, E., and Welch, I. (2000), "A Theory of Legal Presumption," *Journal of Law, Economics and Organization,* 16, 1-49.

Bundy, S.M., and Elhauge, E. (1991), "Do Lawyers Improve the Adversary System? A General Theory of Litigation Advice and Its Regulation," *California Law Review,* 79, 313-420.

Bundy, S.M., and Elhauge, E. (1993). "Knowledge about Legal Sanctions," *Michigan Law Review,* 92, 261-335.

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

Daughety, A.F., and Reinganum, J.F. (2000a). "Appealing Judgments," *Rand Journal of Economics*, 31, 502-525.

Daughety, A.F., and Reinganum, J.F. (2000b). "On the Economics of Trials: Adversarial Process, Evidence and Equilibrium Bias," *Journal of Law, Economics, and Organization*, 16, 365-394.

Dewatripont, M., and Tirole, J (1999). " Advocates," *Journal of Political Economy*, 107, 1-39.

Frank, J. (1973), *Courts on Trial: Myth and Reality in American Justice*, Princeton, Princeton University Press.

Froeb, L.M., and Kobayashi, B.H. (1996), "Naive, Biased, yet Bayesian: Can Juries Interpret Selectively Produced Evidence?" *Journal of Law, Economics, and Organization*, 12, 257-276.

Froeb, L.M., and Kobayashi, B.H. (2001), "Evidence Production in Adversarial vs. Inquisitorial Regimes," *Economics Letters*, 70, 267-272.

Grossman, S. (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality, " *Journal of Law and Economics*, 24, pp. 461-483.

Hay, B.L., and Spier, K.E. (1997), "Burdens of Proof in Civil Litigation: An Economic Perspective," *Journal of Legal Studies*, 26, 413-431.

Kaplow, L., and Shavell, S. (1989), "Legal Advice about Information to Present in Litigation: Its Effects and Social Desirability," *Harvard Law Review*, 102, 567-614.

Krishna, V., and J. Morgan (2001), "A Model Of Expertise," *The Quarterly Journal of Economics*, 116, 747-775.

Levy, G. (2005), "Careerist Judges," *Rand Journal Economics,* 36, 275-297.

Lewis, T.R., and Poitevin, M. (1997), "Disclosure of Information in Regulatory Proceedings" *Journal of Law, Economics, and Organization,* 13, 50-73.

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

Luban, D.J. (1983), "The Adversary System Excuse," in *The Good Lawyer: Lawyers' Roles and Lawyers' Ethics*, (D.J. Luban eds.), New Jersey: Rowman & Allanheld.

Mialon, H. (2005), "An Economic Theory of the Fifth Amendment," *Rand Journal of Economics*, 36 (4).

Milgrom, P. (1981), "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380-391.

Milgrom, P., and Roberts, J. (1986), "Relying on the Information of Interested Parties," *Rand Journal Economics*, 17, 18-32.

Posner, R.A. (1999), "An Economic Approach to the Law of Evidence," *Stanford Law Review*, 51, 1477-1546.

Rhode, D. (1985), "Ethical Perspectives on Legal Practice," *Stanford Law Review*, 37, 589-614.

Sanchirico, C.W. (1997), "A Burden of Proof in Civil Litigation: A Simple Model of Mechanism Design" *International Review of Law and Economics*, 17, 431-447.

Sanchirico, C.W. (1998), "Enforcement by Hearing: An Integrated Model of Evidence Production," *USC Law School Olin Working Paper* No. 98-19.

Sanchirico, C.W. (2001), "Relying on the Information of Interested — and Potentially — Dishonest Parties" *American Law and Economics Review*, 3, 320-357.

Seidmann, D. (2005), "The Effects of a Right to Silence," *Review of Economic Studies*, 72, 593-614.

Shin, H.S. (1994), "The Burden of Proof in a Game of Persuasion," *Journal of Economic Theory*, 64, 253-264.

Shin, H.S. (1998), "Adversarial and Inquisitorial Procedure in Arbitration," *Rand Journal of Economics*, 29, 378-405.

Sobel, J. (1985), "Disclosure of Evidence and Resolution of Disputes: Who Should Bear the Burden of Proof?" in *Game Theoretical Models of Bargaining*, A. Roth (ed.) Cambridge: Cambridge University Press.

Spurrier, R.L. (1980), *Inexpensive justice: Self-representation in the small claims court*, Kennikat Press.

Thompson, W. (1989), "Are Juries Competent to Evaluate Statistical Evidence?" *Law and Contemporary Problems*, 52, 9-41.

Tribe, L.H. (1971), "Trial by Mathematics: Precision and Ritual in the Legal Process," *Harvard Law Reeview*, 84, 1329-1393.

Wells, G. (1992), "Naked Statistical Evidence of Liability: Is Subjective Probability Enough?" *Journal of Personality and Social Psychology*, 62, 739-752.