Disclosure and Legal Advice*

Yeon-Koo Che† Sergei Severinov‡

This Version: December 2, 2014

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, either if (s)he can credibly advise his client to suppress some favorable evidence or if there is a cost associated with legal advice. The effect is socially undesirable in the former case, but it is desirable in the latter case although the benefit rests on its purely dissipative role rather than on his expertise. These results provide a general perspective for understanding the role of private information and expert advice in disclosure.

KEYWORDS: Lawyer advice, disclosure of evidence, regulating adjudicators' inferences.

*The authors thank Andrew Daughety, Bob Hall, Ken Judd, Navin Kartik, Bart Lipman, Mitchell Polinsky, and Kathy Spier, 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. Severinov acknowledges support from Social Sciences and Humanities Research Council of Canada.

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

‡University of British Columbia, email: sseverinov@gmail.com

1 Introduction

Lawyers play a prominent role in the modern day adjudication process. One notable aspect of their role involves advising clients on disclosing information to the court. Lawyers can advise their clients which evidence is unfavorable and should be withheld and which evidence is favorable and thus should be disclosed. Although lawyers often have a disclosure duty before the tribunal (particularly, in civil cases), 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] The goal of this paper is to understand whether and to what extent the lawyers can affect the outcome of a trial by influencing the amount and the nature of information reaching the court.

To this end, we study adjudication of a dispute between two parties, say defendant and a plaintiff, who may obtain legal advice. Formally, the dispute is modeled as an evidence disclosure game. The adjudicator, or the judge, decides whether to "convict" or "acquit" the defendant based on all information available to her. Part of the judgment-relevant information is the evidence the parties themselves may (or may not) possess. The main strategic decision for a party is whether to disclose evidence truthfully or to withhold it. The judge's ruling depends also on another piece of information, which reflects the legal rules and standards on interpreting that evidence and other public evidence surrounding that case. The lawyers can assess the latter piece of information better than the parties. That is, a lawyer can assess whether a party's evidence is favorable or unfavorable and how strong his case would be without its disclosure. A party advised by a lawyer can thus make a more informed decision about disclosure. We study this particular role of lawyers, and focus on understanding how the lawyer advice influences parties' disclosure behavior and the judge's inference and her ruling.

The resulting model introduces rich strategic interactions in a disclosure game. First, the lack of common knowledge about the existence of evidence makes the judge's inference nontrivial, since nondisclosure need not imply a party's concealment of unfavorable information. Hence, an equilibrium typically would not involve full disclosure, much in contrast to the unraveling that is typical in verifiable disclosure games (see Milgrom (1981) and Grossman (1981)). Second, a judge's inference is influenced by what a lawyer advises his client to disclose. In this sense, lawyer advising adds a new dimension both to the parties' strategic disclosure and the quality of the judge's inference.

Finally, the act of seeking lawyer advice itself, especially when it is costly, may signal whether a party possesses relevant evidence, and if so, what that evidence may be. This signal gener-

[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.

ally influences the judge's inference and her ruling. Our model accommodates these strategic interactions.

Our model produces several surprising results about the role of advising on disclosure. First, we find lawyer advising to be irrelevant — both privately and socially — under a baseline scenario where the advice is costless and the lawyers employ disclosure strategies that satisfy a certain credibility requirement, that is, to disclose information if and only if it is favorable to her client. This irrelevance finding is surprising since a lawyer observes additional judgment-relevant information that can improve his client's disclosure decision. Indeed, lawyers' advice in general affects the parties' equilibrium disclosure behavior. Yet, the change in the disclosure behavior does *not* affect the outcome of adjudication. This irrelevance holds regardless of whether one or both parties obtain legal advice, and whether the adjudicator makes a Bayesian inference based on the parties' disclosure strategies or follows an ad-hoc rule satisfying certain reasonable properties.

We then extend our baseline scenario to identify two circumstances in which legal advice on disclosure *does* affect the outcome of a trial. First, a lawyer's 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. However, the total welfare of the parties falls if both parties hire lawyers. 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 a rule by which she assigns the burden of proof and thus the way in which she draws an inference about the defendant's guilt. This last result provides a rationale for placing restrictions on the adjudicators' interpretations of evidence or lack thereof.

Lawyer advising can also affect the adjudication outcome when hiring a lawyer is costly. This cost provides a means by which the parties without evidence can credibly signal the lack of evidence and avoid a prejudicial inference by the adjudicator. In essence, hiring a lawyer buys one "the right to be silent without prejudice." The parties with unfavorable evidence also hire lawyers and often end up withholding evidence. However, the parties with moderately unfavorable evidence — those who would seek legal advice had it been free — do not hire a lawyer and disclose their evidence. Overall, the cost of legal advice increases the disclosure of private information, which in turn improves the quality of adjudication — in fact, more so as the cost increases.

Our analysis 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 several aspects not captured in our simple model. Yet, the advisory role of lawyers in disclosure is 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 as a useful benchmark — a 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 when the adjudicator is Bayesian and the lawyers use the strategy of disclosing all favorable evidence. In this case, the equilibrium outcome is socially optimal. However, this conclusion no longer holds in an equilibrium where the lawyers use the strategies of withholding some favorable evidence. In this case, the social harm associated with "strategic withholding" can be mitigated by a rule which allocates all burden of proof to one party. These results contribute to the understanding of evidentiary rules and procedures adopted by the courts.

Our modeling framework and the results are useful for understanding the role of advising more broadly, in settings other than dispute resolution. Indeed, the insight we develop on advising holds equally well in a setting where there is only one party. Often, decisions that have significant consequences for a party must be made based on the information provided by that party. Promotion and grant allocation, college admission and job application, product introduction and promotion are some relevant examples. A party facing a decision in such a context often seeks advice from mentors, counselors or consultants regarding strategies of information revelation. Our results offer basic necessary conditions for such advice to be relevant.

The current paper contributes to the literature of verifiable disclosure games. The literature originated from the seminal contributions by Grossman (1981), Milgrom (1981), Milgrom and Roberts (1986), and was further developed by Lipman and Seppi (1995) and Seidmann and Winter (1997). The key result of this literature is the so-called "unraveling," namely that conflicting interests can lead to full revelation of the parties' private, but non-falsifiable, information. The common knowledge of an agent's possession of information is crucial for this result. We relax this common knowledge assumption, as in Verrecchia (1983) and Shin (1994, 1998). As mentioned above, relaxation of common knowledge makes the judge's inference nontrivial. More recently, Kartik, Suen and Xu (2013) analyze disclosure of verifiable information in the context of a persuasion game.

Our paper is also related to the literature on the adversarial system of disclosure. Dewatripont and Tirole (1999) study the desirability of adversarial system in a broad organization design context. Shin (1998) compares adversarial and inquisitorial litigation systems. Froeb and Kobayashi (1996, 2001) assess the implications of endogenous evidence production. Sobel (1985), Hay and Spier (1997), Sanchirico (1998, 2000, 2001) and Bernardo et al. (2000) explore the allocation of the burden of proof and evidence production from the standpoint of litigation costs and deterrence. Sanchirico and Triantis (2008) study costly evidence fabrication and efficiency of adjudication in the context of contractual disputes.[2] Seidmann (2005), Mialon (2005) and

[2]Levy (2005) studies the effect of career concerns on judges' decision making. Also related is the literature on cheap talk, which includes Crawford and Sobel (1982) and Krishna and Morgan (2001) among others.

Leshem (2010) investigate the effect of the defendant's right to silence, with and without adverse inference by the adjudicator, on the adjudication outcomes and welfare. Shchepetova (2014) compares costly evidence production in inquisitorial and adversarial systems. In the follow-up paper, Turkay (2013) studies how the severity of legal punishment affects evidence disclosure behavior. Hadfield and Leshem (2012) provide a comprehensive review of the law and economics literature on attorney-client relationship and, in particular, the role of the confidentiality rules. None of these papers deal with the role of lawyers in disclosure — the focus of this paper.

The issue of legal advice has received relatively little formal treatment in the literature. Legal scholars have recognized the factors favoring and disfavoring the lawyer-aided adversarial system but disagree on the relative importance of those factors. Proponents argue that 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 that while the lawyers' ability to suppress evidence based on legal expertise undoubtedly benefits their clients, its social implications are ambiguous. 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, focusing instead 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 players' strategies. Among other benefits, this latter approach enables us to study how the rules and restrictions on inferences may affect the adjudication outcomes. In another related piece, Iossa and Jullien (2012) study the market for lawyers and focus on the match between the nature of the legal dispute and the quality of the lawyers hired by the litigants, as well as on the effect of the lawyer's reputation on the adjudicators.

2 Model

Two parties, 1 and 2, are in a dispute, which is adjudicated by an adjudicator in a tribunal. It is convenient to interpret parties 1 and 2 as a defendant and a plaintiff in a litigation. However, our model applies 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[3]. 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 be observed only by the parties to the dispute. The evidence

[3]For jury interpretation to apply, the jury must be given instructions regarding the content and application of the law by the judge.

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).[4] Obviously, $\sum_{i,j=0,1} p_{ij} = 1$, and we assume that $p_{ij} > 0$ for all $i, j = 0, 1$. We allow for possible correlation in the parties' abilities to observe 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 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 common knowledge.

There is another piece of judgment-relevant information, $\theta \in [0, 1] =: \Theta$, which is observed only 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.[5] The disputing parties have limited knowledge of the law and incomplete understanding of the legal process, so they can learn θ only by hiring lawyers. Lawyers understand the body of the law, 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. Ultimately, the lawyers' ability — and the litigants' *in*ability — to observe θ introduces a potentially productive role for the lawyers.

We assume that (s, θ) is drawn from $S \times \Theta$ according to an absolutely continuous cdf, $F(s, \theta)$ with a positive density $f(s, \theta)$ in the interior of $S \times \Theta$. From the ex-ante perspective, θ is random because the law and legal standards as well as the circumstance as perceived by the court may vary across cases. 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

[4] "Observing" s means either possessing that evidence or having a proof of its existence.

[5] 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 description of situations where s is not disclosed.

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. 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).[6] 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 advising. For instance, one could view the two regimes as both involving lawyer advising but differing in the quality of advising.

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 lawyer advising 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.

• **Evidence disclosure behavior:**

If a party does not hire a lawyer, then his decision to disclose s is based solely on s. In contrast, if a party hires a lawyer, he can make the disclosure decision based on the lawyer's advice, i.e., her knowledge of θ.

A lawyer prefers his client to prevail in court, and there are no agency issues in the attorney-client relationship. So, a client will communicate s to his lawyer truthfully, and a lawyer will explain the legal issues, i.e. communicate θ to the client truthfully. Therefore, a lawyer-advised party can simply be viewed as informed of both s (if he observes s) and θ.

Formally, party i's disclosure strategy is a function ρ_i mapping $S \times \Theta$ to $[0, 1]$, with $\rho_i(s, \theta)$ representing the probability that party $i \in \{1, 2\}$ discloses s for given θ. If a party is unrepresented, he does not observe θ, so $\rho_i(\cdot, \theta)$ must satisfy the condition $\rho_i(\cdot, \theta) = \rho_i(\cdot, \theta')$ for any θ, θ'.

• **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. A binary decision is common, and is more general than may appear at first glance. For instance, there may be no ambiguity about the size of damages, leaving the liability as the only object of dispute.[7]

The judge's ruling depends on (s, θ) if s has been disclosed, and on θ alone 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's (defendant's) culpability. If $g(s, \theta) > 0$, then the judge finds party 1

[6] 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.

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

culpable and 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 has no real consequence.

We assume that the function $g(\cdot, \cdot)$ is common knowledge between all players, including the lawyers and parties 1 and 2, and that $g(s, \theta)$ is increasing and continuous in both arguments. So, lower s and θ are more favorable for party 1, and vice versa. 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 less favorable for the defendant. To make the judge's decision problem nontrivial, we assume that $\int g(s, 1) f(s|1) ds > 0$ and $\int g(s, 0) f(s|0) ds < 0$. This implies 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 for party 1 ($\theta = 0$) to the least favorable ($\theta = 1$) for her.[8] 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 (s, θ) 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.[9]

The adjudication criterion $g(s, \theta)$ can be rationalized by a society's objective that the judge follows. Suppose the society would like to minimize the cost associated with a wrong decision, i.e. "convicting the innocent or exonerating the guilty." Let c_1 and c_2 be the cost of ruling mistakenly for party 1, the defendant, ("exonerating the guilty") and for party 2 ("convicting the innocent"), respectively, and let $\pi(s, \theta)$ be the probability that party 1 is guilty for given (s, θ). Then, if the judge convicts party 1 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 $g(s, \theta) := \pi(s, \theta) - \frac{c_2}{c_1 + c_2}$.[10] We assume throughout that the judge follows the criterion g whenever the evidence s is disclosed by either party.

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

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

[10] 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 levels between 0.75 and 0.9 with the standard of "proof beyond a reasonable doubt."

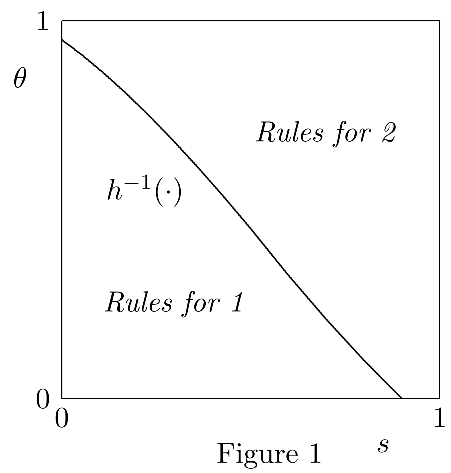

Figure 1

If no party discloses s, then the judge still makes a decision based on some "belief" about g, formed by θ she observes and by some inference about the parties' disclosure decisions. The adjudicator's decision rule in the event of nondisclosure, 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 allow for a large class of posteriors involving 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 $i \in \{1,2\}$ has observed s but has not disclosed. 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.[11] 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 there is no ambiguity.)

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. Throughout, we will only assume that

[11]The value of (1) corresponds 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 judge applies the same criterion, i.e. the coefficients (a, b_1, b_2, c) remain constant, regardless of whether a party is lawyer-advised 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).[12] Depending on the values of these variables, the adjudication criterion in (1) accommodates a variety of different decision procedures and burden-of-proof allocations. 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*; 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. In other words, the burden of proof is put on party 1.[13]

Likewise, if $b_2 > 0$ and $b_1 = c = 0$, then the burden of proof is put on 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. 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. We will say that the burden of proof shifts from party 2 to party 1 as b_1 increases and b_2 decreases, and vice versa.

This general model includes as a special case a *fully Bayesian* judge, 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. 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.[14] Therefore, it is important to allow for non-Bayesian — as well as

[12] 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 no role.

[13] In our terminology, the judge puts the "burden of proof" on a party if non-disclosure of information causes the judge to believe that this party has withheld evidence and hence to form the most negative inference against that party. As a result, the judge rules against this party with a high probability, i.e. for a large range of values of θ, in the case of non-disclosure. Still, the judge may end up ruling for the party bearing the burden of proof even when no evidence is produced. This is not inconsistent with the standard legal definition of the burden of proof. Indeed, a party bearing the burden of proof typically loses the court case when she does not produce evidence; but such a party could, and does occasionally, win the case when there is no *prima facie* evidence supporting the other side. Our model should be interpreted as considering this latter case.

[14] One of the most well-known examples 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 and members of the jury are very unlikely to make such a conviction when only this type of evidence is presented. Several legal

Bayesian — burden-of-proof allocations.

The judge's inference rule may also reflect 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 is allowed to make from certain evidence or lack thereof, because of concerns about their prejudicial effect. Our model allows us to study the implications of such restrictions. From this perspective, the coefficient b_i represents the extent to which the rule allows the judge to be predisposed 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, that is, when party 1 discloses 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 strategies (with a slight abuse of notation) is given by

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

We shall often suppress the dependence of the posterior on the the inference rule (b_1, b_2, c), unless confusion is likely.

• **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). We assume that, whenever the evidence is disclosed by either party, the judge follows the criterion g. This could be because in case s is disclosed, the ruling must be based on an immutable legal rule described by the criterion $g(s,\theta)$, and any deviation from it would constitute an "error" of law. By contrast, there is more ambiguity, so there is more scope for judge's discretion, when crucial evidence is not disclosed. This approach also allows us to focus on a nontrivial inference problem facing the judge in the event of nondisclosure.

Our ultimate interest is in the equilibrium outcome of the trial given the information available to the parties. Formally, 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 party i does not observe

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 $b_i > 0$.

s. In particular, an equilibrium (ρ_1, ρ_2, δ) induces the following outcome function:

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

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

3 Irrelevance of Lawyer Advising

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

3.1 No Advising

In this regime, neither party 1 nor party 2 has a lawyer. Thus, the parties 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 uses a cutoff strategy because her posterior $\mathbb{E}[g|\hat{s}, \hat{s}, \theta]$ is monotonic in θ, which follows from two effects. First, a higher θ is by itself stronger evidence of 1's culpability, holding s fixed. Second, there is also an inference effect. Monotone likelihood ratio property (MLRP) implies that a higher θ makes a high value of s more likely. So nondisclosure is more likely to be a result of party 1's concealment of unfavorable s (rather than his not observing s), given the parties' cutoff strategies. Obviously, this inference effect adds to the judges's suspicion of 1's culpability. Figure 2 graphs the two thresholds $\hat{s}$ and $\hat{\theta}$.

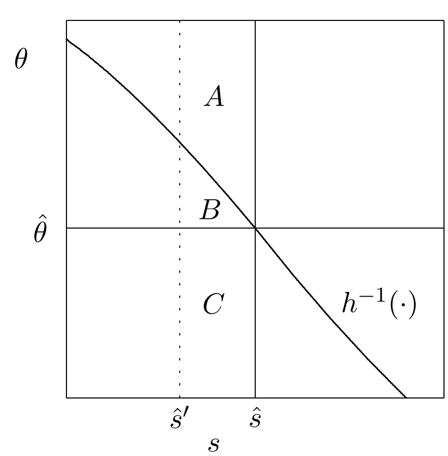

Figure 2

Crucially, the two thresholds $\hat{s}$ and $\hat{\theta}$ cross each other on the curve g; i.e., $g(\hat{s}, \hat{\theta}) = 0$ or $\hat{s} = h(\hat{\theta})$. The intuition for this is as follows. Suppose that, facing the judge's threshold $\hat{\theta}$, party 1 deviates by withholding any $s \in [\hat{s}', \hat{s})$. Let us show that such a deviation is not profitable. To this end, assume that party 2 does not disclose, otherwise 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 this deviation makes no difference, for the judge's ruling will be the same whether party 1 discloses s or not. But if θ happens to be 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 his favor. 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 party 2 will adopt cutoff strategies with threshold $\hat{s} = h(\hat{\theta})$: party 1 discloses her/his evidence if and only if $s < \hat{s}$, and party 2 discloses if and only if $s > \hat{s}$. 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 use the same threshold.

It is instructive to study 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 this result. Its proof is provided in the Appendix and also establishes the cutoff nature of the equilibrium.

Proposition 1 *If no party is advised by a lawyer, there exists a unique Perfect Bayesian equilibrium in which party 1 discloses s if and only if $s < h(\hat{\theta}^*(b_1, b_2))$, and party 2 discloses s if and only if $s > h(\hat{\theta}^*(b_1, b_2))$. Following nondisclosure, the judge rules for party 1 if $\theta < \hat{\theta}^*(b_1, b_2)$ and for party 2 if $\theta > \hat{\theta}^*(b_1, b_2)$. The threshold $\hat{\theta}^*(b_1, b_2)$ decreases in b_1 and increases in b_2 (i.e. party 1 discloses more and party 2 discloses less as the burden of proof shifts to party 1).*

3.2 Two-sided Advising

In this regime, both parties receive lawyer advice and learn θ. Hence, unlike in the no-advising case, the parties make their disclosure decisions based on both s and θ. Recall that the judge's ruling in case of disclosure follows the criterion $g(s, \theta)$. So, party 1 has a weakly dominant strategy of disclosing s if and only if $s < h(\theta)$, or $g(s, \theta) < 0$. Disclosing $s < h(\theta)$ leads to a sure win for party 1, whereas withholding it 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 lawyer advising. In most legal systems, 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 a strategy may also be problematic. Furthermore, the judge could simply refuse to believe that a lawyer is not following a dominant strategy, in which case deviation from the dominant strategy has no effect. Finally, if there is even a 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. For these reasons, we focus on the dominant disclosure strategies here. Later, we will consider what happens when these arguments do not apply and examine equilibria supported by "dominated" disclosure strategies.

Note that the disclosure behavior of a lawyer-advised party using a dominant strategy is different from the equilibrium behavior of an unadvised party, since the disclosure threshold of a represented party varies with θ. In the no-advising case, the parties employ a common disclosure threshold $\hat{s} = h(\hat{\theta}^*(b_1, b_2))$ that does not vary with θ. Consequently, the judge's posterior is $\mathbb{E}[g|h(\theta), h(\theta), \theta]$ in the two-sided advising case, whereas her posterior in the no-advising case is $\mathbb{E}[g|\hat{s}, \hat{s}, \theta]$ with $\hat{s} \equiv h(\hat{\theta}^*(b_1, b_2))$. These posteriors differ in their magnitudes for almost all θ, but, remarkably, they have the same sign for all θ. In particular,

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

This leads the judge to adopt the same default ruling under both regimes. Intuitively, this is so because whenever the realized value of θ is equal to the judge's threshold $\hat{\theta}$, the disclosure behavior of the parties is identical in both regimes. Party 1 discloses $s < h(\hat{\theta})$, whether she is represented or not. Party 2 discloses $s > h(\hat{\theta})$. Hence, we arrive at the following result.

Proposition 2 *If both parties are lawyer-advised, 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 $g(s, \theta) > 0$.*

The fact that the threshold $\hat{\theta}^*(b_1, b_2)$ is the same in the no-advising and the two-sided advising cases will be crucial for the irrelevance result shown later.

3.3 One-sided Advising

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 sign from negative to positive at $\hat{\theta}^*(b_1, b_2)$, the following result is immediate.

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 (3)) in her default ruling. If party 1 obtains legal advice, she discloses s if and only if $g(s, \theta) < 0$. If party 1 does not obtain legal advice, she 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 behavior of the party without a lawyer, in a way described in Proposition 1.*

The key result of subsequent interest is that the judge follows the same default ruling as in the other regimes. As explained before, this is due to the fact that the parties' different disclosure strategies affect the magnitude of the judge's equilibrium posterior assessment, but do not affect its sign.[15]

3.4 Irrelevance of Lawyer Advising

A striking feature emerging from our analysis is that the judge's equilibrium default ruling strategy is the same across all three regimes. At the same time, Propositions 1, 2, and 3 show that the parties follow different disclosure strategies across the regimes. Nevertheless, we will show below that the differences in the parties' disclosure behavior do not lead to any real differences in the outcome of the trial.

This irrelevance is, in fact, a result of a more general property of our disclosure/adjudication game, which is described in the following Lemma. Let $\phi^{\hat{\theta}}(x_1, x_2, s, \theta)$ denote the probability with

[15] In fact, this result holds even more generally. Suppose each party randomizes in hiring a lawyer, the judge has some arbitrary beliefs about the parties' decisions to hire lawyer, and this belief may even be inaccurate. The main logic determining the equilibrium behavior remains unchanged, leading to the same characterization of the judge's default ruling.

which the judge rules for party 2 in the state (x_1, x_2, s, θ) when she follows a default ruling strategy with threshold $\hat{\theta}$.

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., whether either party obtains legal advice or not, mutual best responses by the two parties in disclosure lead to the same outcome characterized by the following outcome function $\phi^{\hat{\theta}}(x_1, x_2, s, \theta)$:*

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

The Decision Equivalence Lemma establishes that the judge's cutoff strategy uniquely determines the equilibrium adjudication outcome, regardless of the parties' use of legal advice. An insight into this result can be gained by comparing the regimes of no advising and two-sided advising. Figure 3 illustrates the case in which the judge follows a threshold $\hat{\theta}$.

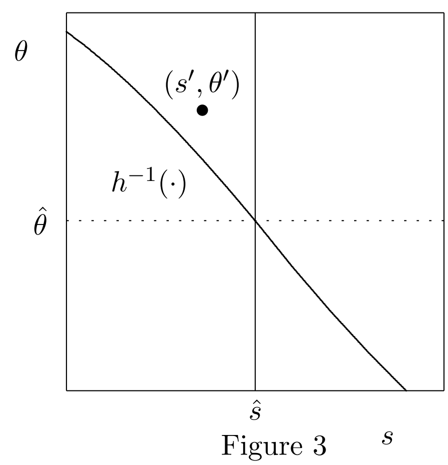

Figure 3

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 the difference in disclosure behavior, there is no difference in the adjudication outcome: the judge's ruling is unfavorable to party 1 in either case. A similar illustration can be provided for any other pair (s, θ). 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), in each advising regime, and that the parties employ undominated strategies in* *disclosure. Then, regardless of the advising regime, there is a unique equilibrium characterized by the outcome function $\phi^{\hat{\theta}^*(b_1,b_2)}(\cdot)$ described by (4).*

Importantly, the irrelevance result does not depend on the judge's inference rule (b_1, b_2, c). It holds if the judge follows a Bayesian rule (i.e., $(b_1, b_2) = (\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$) or 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 lawyer advising has no effect on the adjudication outcome.

Our analysis focuses on one particular aspect of legal representation — the role of lawyers as gate-keepers of information reaching the court. In practice, lawyers perform a number of valuable tasks that are not captured by our model, so our result should not be interpreted as suggesting that legal representation is never useful. Nevertheless, the robustness of our irrelevance result is surprising. To the extent that information transmission matters in adjudication, however, our irrelevance result clarifies and qualifies the sense in which lawyers influences this process. In particular, the irrelevance result identifies the circumstances under which *legal advice matters*. The next section considers one such case.

4 Relevance of Lawyer Advising: Withholding Favorable Evidence

Thus far, we have focused on equilibria in undominated strategies. A lawyer adopting such a strategy would never advise a client to suppress ex post favorable evidence. Although such a strategy can be seen as focal, deviating from it and sometimes suppressing favorable evidence may influence the judge's posterior and thus ultimately her ruling.

To illustrate this point, suppose both parties have hired lawyers. Suppose further that party 1 follows a strategy of never disclosing any evidence while party 2 discloses favorable evidence. This will induce the judge to treat party 1 more favorably in case of non-disclosure, since non-disclosure by party 1 is neutral about the content of the evidence, while party 2's non-disclosure will likely indicate his concealment of evidence favorable to party 1. Specifically, the judge's posterior in case of non-disclosure 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. 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 $\mathbb{E}[g|h(\theta), 1, \theta]$, which is more favorable for party 2 than if both adopt their dominant strategies. Of course, this strategy of suppressing potentially favorable evidence makes sense only if the judge rules favorably despite the suppression. Since the judge will not rule favorably if θ is sufficiently unfavorable, the strategy will work only for a range of θ's. Let

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

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

Observe that $\hat{\theta}_-(b_1, b_2, c) \leq \hat{\theta}^*(b_1, b_2) \leq \hat{\theta}_+(b_1, b_2, c)$, and that if $(b_1, b_2, c) > (0, 0, 0)$, then $\hat{\theta}_-(b_1, b_2, c) < \hat{\theta}_+(b_1, b_2, c)$.[16] If $\theta > \hat{\theta}_+(b_1, b_2, c)$, then $\mathbb{E}[g|0, h(\theta), \theta] > 0$, so the judge will rule for party 2 unless party 1 presents evidence $s < h(\theta)$. Hence, it will never be optimal for party 1 to suppress evidence $s < h(\theta)$ for such θ. Nevertheless, party 1 may be able to sway the judge's posterior by suppressing all evidence, both favorable and unfavorable, when $\theta < \hat{\theta}_+(b_1, b_2, c)$ since the judge will rule for party 1 even without disclosure. That is, employing the strategy of withholding all evidence makes disclosing inherently favorable information "unnecessary." By the same logic, party 2 can credibly sway the judge's posterior by suppressing all evidence (given that party 1 follows his dominant strategy), if $\theta > \hat{\theta}_-(b_1, b_2, c)$.

These arguments suggest the parties can manipulate the judge's posterior within the interval $[\hat{\theta}_-(b_1, b_2, c), \hat{\theta}_+(b_1, b_2, c)]$. Indeed, any $\hat{\theta} \in [\hat{\theta}_-(b_1, b_2, c), \hat{\theta}_+(b_1, b_2, c)]$ can be supported as an equilibrium threshold for the judge's default ruling when both parties are advised by lawyers:

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 $\theta > \hat{\theta}$. In this equilibrium, party 1 discloses evidence s if and only if $s < h(\theta)$ and $\theta > \hat{\theta}$, and party 2 discloses s if and only if $s > h(\theta)$ and $\theta < \hat{\theta}$.*
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, absent disclosure, the judge rules for party 1 if $\theta < \hat{\theta}$ and for party 2 if $\theta > \hat{\theta}$. In this equilibrium, party 1 discloses evidence s if and only if $s < h(\theta)$ and $\theta > \hat{\theta}$, and party 2 discloses s if and only if $s > h(\hat{\theta})$.
Conversely, any cutoff default ruling strategy by the judge has a threshold in $[\hat{\theta}^*(b_1, b_2), \hat{\theta}_+(b_1, b_2, c)]$.
An analogous 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 seemingly favorable evidence along with unfavorable evidence. Since such a strategy requires conditioning disclosure 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 seemingly favorable evidence.

[16]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 are satisfied by assumption.

The benefit of withholding seemingly favorable evidence 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 withhold favorable evidence, 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 is $\hat{\theta}^*(b_1, b_2)$, the threshold used when neither party hires a lawyer. In this sense, legal advice becomes relevant once weakly dominated strategies are available.

Of course, it is not possible for both parties to be strictly better off from hiring lawyers. This 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 playing the 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 enhance and maintain them. For example, some criminal defense lawyers are known to call very few witnesses. Sometimes, a lawyer may rest the case without calling any witnesses at all if he or she believes 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 her 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 advising 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 scope for manipulation 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 in the next section.

5 Welfare implications of Legal Advising 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 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, 2000; 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 anyway. If not, some restrictions may be warranted, and it is important to investigate the reasons.

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 the 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 that it solves $[WP]$.*

In light of Propositions 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 parties obtaining legal advice play a weakly dominant strategy, since a 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 seemingly 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 lawyer advising is privately valuable but moves the outcome away from the social optimum. So lawyer advising 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*

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

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 (i.e. either $b_1 = 0$ or $b_2 = 0$), provided that the parties follows undominated strategies. This finding is consistent with the evidence law that assigns the entire burden of proof to one of the parties, e.g., the plaintiff. It is interesting that a 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 $(b_1, b_2) = (\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$, it is possible to lower both coefficients without altering the threshold, until the lower of the two becomes zero.

The second part of the proposition indicates that such allocation of the burden-of-proof may be socially desirable. Lower values of (b_1, b_2) reduce the scope of the judge's discretion in adjusting her inference based on the parties' disclosure strategies, since the range of equilibrium outcomes around the optimal outcome shrinks as (b_1, b_2) falls. 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 $(\frac{p_{10}}{p_{00}}, \frac{p_{01}}{p_{00}})$. These two findings imply that committing adjudicators to an inference rule different from the Bayesian could be socially desirable.[17]

6 Can Money Buy Justice?: A Signaling Role of Costly Legal Advice

So far, we have assumed that legal advice is available for free. This assumption helps to isolate the effect of lawyers' expertise and provides a foundation for normative analysis. But in reality, lawyer advice is often costly, more so if a lawyer has a good reputation or a high profile. It is thus natural to extend our model to introduce positive lawyer cost. The most important finding of this section is that lawyer advice, when available at a cost, changes the parties' disclosure behavior in a way that affects both private and social welfare.

This extension also enables us to study when individuals choose to hire lawyers and whether this can affect adjudication outcomes. There is a concern that prominent lawyers can skew the justice in favor of their clients. Our analysis provides insight into this issue.

[17] 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)).

To extend our model in this direction, we need to add more structure to it. In particular, the cost of lawyer advising must be measured in units comparable to the value of winning. To this end, we assume that the cost of legal advice is $w > 0$, and that a party derives a value $v > 0$ when (s)he wins the dispute and zero when (s)he loses, all measured in monetary units. We then consider a game in which each party with a signal s chooses whether to hire a lawyer at the cost of w or not, followed by the disclosure game studied earlier.

Further, we simplify our model in several ways. First, we consider a symmetric environment in which each party observes evidence with probability $p \in (0, 1)$, independent of the other party (i.e., $p_{10} = p_{01} = p(1-p), p_{11} = p^2, p_{00} = (1-p)^2$), and that $b_1 = b_2 = b$. Next, we assume s and θ to be independently distributed according to cdf $F(s, \theta) = K(s)L(\theta)$ with $K(.)$ and $L(.)$ having densities k and l, respectively. Next, we assume that the nature of the dispute is symmetric between the two parties:

Assumption 2 *(symmetry) For all $(s, \theta) \in [0, 1]^2$, $k(s) = k(1-s)$, $l(\theta) = l(1-\theta)$, and $g(s, \theta) = -g(1-s, 1-\theta)$.*

Among other things, this assumption means that the strength of the case for party 1 in state (s, θ) is the same as the strength of the case for party 2 in state $(1-s, 1-\theta)$. The assumption implies that $g(\frac{1}{2}, \frac{1}{2}) = 0$, or $h^{-1}(\frac{1}{2}) = \frac{1}{2}$.

With costly legal advice, there are two types of equilibria: an "uninformative" equilibrium in which legal advice has no signaling role, and an "informative" one in which lawyer advising becomes a signal. We begin with the uninformative equilibrium.

Proposition 8 **(Uninformative equilibrium)** *There exists an equilibrium in which no party retains a lawyer, and the parties follow the same strategies as in the equilibrium described in Proposition 1.*

The proof of this proposition is straightforward. Suppose that the judge follows a default ruling strategy with threshold $\hat{\theta}^*$ characterized by (3), regardless of whether any party has retained a lawyer. Then, it is a best response for parties not to hire lawyers and to follow the disclosure strategies with threshold $h(\hat{\theta}^*)$. Such strategies in turn rationalize the judge's beliefs and her default ruling behavior. If a party, say party 1, deviates and hires a lawyer, then disclosing s if and only if $g(s, \theta) < 0$ is his dominant strategy. Thus, the judge's ruling is sequentially rational after this deviation also. Hence, the uninformative equilibrium can be sustained.

There exists another, more interesting, equilibrium in which legal representation plays a non-trivial signaling role. In this equilibrium, a lawyer is hired for two reasons. First, a party without evidence hires a lawyer to add "credibility" to his nondisclosure. Second, a party with sufficiently weak (likely unfavorable) evidence does so to make sure that they disclose only favorable evidence and to imitate those without any evidence. Specifically, for any cost $w > 0$, for an associated threshold $\hat{s}(w) \in [\frac{1}{2}, 1]$ (defined below), consider the following *lawyer signaling strategies*:

- *Party 1 [resp. party 2] retains a lawyer either if he observes no evidence or if he observes* $s > \hat{s}(w)$ *[resp.* $s < 1 - \hat{s}(w)$*]. In the latter case, he discloses* s *if and only if* $g(s, \theta) < 0$ *[resp.* $g(s, \theta) > 0$*].*

- *If party 1 [resp. party 2] observes* $s \in [0, \hat{s}(w)]$ *[resp.* $s \in [1 - \hat{s}(w), 1]$*], he does not hire a lawyer and discloses* s*.*

- *If* s *is disclosed, then the judge rules for party 1 if and only if* $g(s, \theta) < 0$*. If* s *is not disclosed, and either both sides have retained lawyers or no side has retained a lawyer, then the judge rules for 1 if and only if* $\theta < \frac{1}{2}$*. If* s *is not disclosed and only party 1 [resp. party 2] has retained a lawyer, then the judge rules for party 1 if and only if* $g(0, \theta) < 0$ *[resp.* $g(1, \theta) < 0$*].*

Note that the judge's beliefs off the equilibrium path are prejudiced against a party without a lawyer. Particularly, *if there is no disclosure and only party 1 (party 2), has retained a lawyer, the judge believes that* $s = 0$ *(*$s = 1$*), i.e. the evidence is the worst for party 2 (party 1).*

To complete the description, we need to characterize the threshold $\hat{s}(w)$. It must satisfy

$$v(1 - p)[L(\tfrac{1}{2}) - L(h^{-1}(\hat{s}(w)))] = w, \tag{6}$$

To understand (6), note that party 1 with signal $s = \hat{s}(w)$ must be indifferent between hiring a lawyer *and* withholding s and not hiring a lawyer *and* disclosing s. The first strategy has cost w but raises party 1's chance of winning if party 2 has not observed $\hat{s}(w)$. (If party 2 observes $\hat{s}(w)$, then he will disclose it because $\hat{s}(w) \geq \frac{1}{2}$ and hence $\hat{s}(w) \in [1 - \hat{s}(w), 1]$, so party 1's decision would not make any difference.) Suppose party 2 does not observe $\hat{s}(w)$. Then, if party 1 does not hire a lawyer and discloses $\hat{s}(w)$, he wins if $g(\hat{s}(w), \theta) < 0 \Leftrightarrow \theta < h^{-1}(\hat{s}(w))$, whereas hiring a lawyer allows him to win if $\theta < \frac{1}{2}$. Hence, hiring a lawyer increases party 1's probability of winning by $(1 - p)[L(\frac{1}{2}) - L(h^{-1}(\hat{s}(w)))]$. Therefore, for party 1 with evidence $\hat{s}(w)$ to be indifferent, the expected gain from winning more often (the LHS of (6)) must equal the cost of hiring a lawyer, w (the RHS of (6)).

Since $\hat{s}(w) \leq 1$, the cost of lawyer advising must not exceed $\overline{w} := v(1 - p)[L(\frac{1}{2}) - L(h^{-1}(1))]$ for a lawyer to ever be hired. For any $w \in [0, \overline{w}]$, there is a unique solution $\hat{s}(w)$ to (6): its LHS is nondecreasing in $\hat{s}(w)$, equals to zero at $s = \frac{1}{2}$, and equals $\overline{w}$ at $s = 1$. Hence, $\hat{s}(w)$ lies in $[\frac{1}{2}, 1]$ for any $w \in [0, \overline{w}]$. Further, $\hat{s}(w)$ is increasing in w and equals 1 when $w = \overline{w}$.

We are now in a position to state the main result of this section.

Proposition 9 **(Lawyer signaling equilibrium)** *If $w \in [0, \overline{w}]$, then the lawyer signaling strategies constitute a perfect Bayesian equilibrium.*

In the signalling equilibrium, almost all types who hire a lawyer strictly prefer to do so. So, legal advice *does* have private value. Specifically, by hiring a lawyer, a party without evidence buys "the right to remain silent without prejudice," since a non-disclosing party without a lawyer suffers from a very negative inference.

The concern that those who can afford prominent lawyers can buy justice is real in this equilibrium. To see it, consider party 1 with evidence $\hat{s}(w) - \epsilon$ and the same party with evidence $\hat{s}(w) + \epsilon$, for small ϵ. The latter type incurs the cost of hiring a lawyer (whereas the former does not) and consequently wins with a higher probability than the former, despite having a weaker case. In this sense, expensive lawyers can buy justice. But the marginal type himself does not benefit from representation because he has to pay legal fees for an increased chance of winning.

The signaling equilibrium also affects social welfare. To the extent that the lawyers' fees are neutral transfers from the litigants to the lawyers, net social welfare can be measured by the adjudicator's objective. By this criterion, the effect of costly lawyer signaling is quite intuitive. A higher cost of non-disclosure (in terms of negative inference) means that the parties are compelled to disclose more. More disclosure leads to more accurate adjudication. So lawyer signaling — hence costly legal advice — improves welfare. More precisely, since in the lawyer signaling equilibrium party 1 (resp. party 2) does not hire a lawyer and discloses s if $s \leq \hat{s}(w)$ (resp. $s \geq 1 - \hat{s}(w)$) where $\hat{s}(w) > \frac{1}{2}$, this equilibrium entails extra disclosure by the amount represented in Figure 3 with a rightward shift of party 1's threshold and a leftward shift of party 2's threshold.

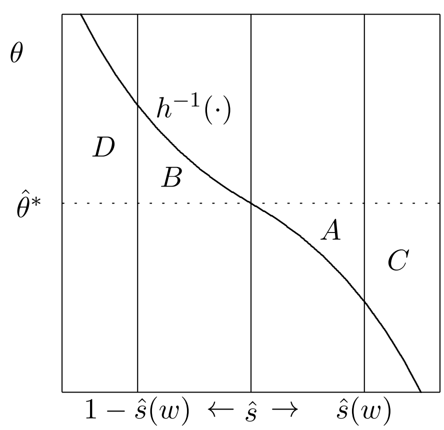

Figure 3

Compared with the uninformative equilibrium (which is equivalent to the equilibrium without representation), the lawyer signaling equilibrium entails different outcomes in two cases: (i) party 1 alone observes evidence and the state is in area A (so party 1 does not disclose); (ii) party 2 alone observes evidence and the state is in area B (so party 2 does not disclose). In both cases, wrong adjudication decisions are made in the uninformative equilibrium but correct adjudication is made in the lawyer signaling equilibrium. Still, the lawyer signaling equilibrium involves incorrect decisions if party 1 (resp. 2) alone observes evidence and the state is in area C (resp. D). But as w increases toward $\overline{w}$, the regions of erroneous ruling shrink, and the social welfare increases. When $w \geq \overline{w}$, there is full disclosure, so adjudication decisions are always correct. The following result shows that costly legal advice can be socially valuable.

Corollary 1 *The lawyer signaling equilibrium yields higher social welfare than the uninformative equilibrium. The social welfare is increasing in the lawyer fee w on $(0, \overline{w}]$ and attains the first-best at $w = \overline{w}$, in which case all evidence is disclosed.*

The social value of lawyers in our model comes from greater disclosure. Legal fees induce more disclosure, as the partes with 'good' evidence decide to forego costly legal representation and to disclose, since an unrepresented party faces a negative inference after non-disclosure.

Our analysis also has implications for regulating adjudicators' inferences. Indeed, it shows that negative inferences from nondisclosure can improve parties' disclosure incentives and lead to a better outcome. At minimum, this suggests that a careful examination of the rules governing adjudicators' inferences is warranted. This is notable since adjudicators are often prohibited from drawing negative inferences against parties refusing to disclose their information, lest such inference distort the judgment. However, little is known about how such inferences affect the parties' disclosure incentives.

Thus far, we have characterized two equilibria. A natural question is whether there are any other equilibria. We show that, given plausible restrictions, no other equilibrium exists.

Proposition 10 *Suppose that parties do not randomize in seeking legal advice,* [18] *the represented parties always follow the dominant strategy of disclosing all favorable evidence and withholding all unfavorable evidence, and the judge follows a threshold strategy in her default ruling. Then there are no other equilibria besides the lawyer signaling equilibrium of Proposition 9 and the uninformative equilibrium of Proposition 8.*

[18] If we allow such randomization, then there is only one other possible equilibrium scenario: an uninformed party randomizes between hiring and not hiring a lawyer. Although we have not been able to completely rule out this type of equilibria, it is easy to show that the qualitative nature of the outcome in such an equilibrium would be the same as in the lawyer signalling equilibrium of Proposition 9.

7 Conclusions

In this paper, we have studied the effect of lawyer advising on disclosure and adjudication. Our analysis was concerned with the role of lawyers as gate-keepers of information reaching the court. We have shown that lawyer advising does not affect the adjudication outcome if legal advice is costless and the lawyers cannot credibly suppress favorable evidence. At the same time, lawyer advising can affect the adjudication outcome if lawyers can leverage their legal expertise to engage in more sophisticated strategic behavior. First, they can do so by credibly withholding some favorable evidence and thereby affecting the judge's inference in their clients favor when evidence is not disclosed. Second, lawyer advising can play a signaling role. In our signalling equilibrium, hiring a lawyer allows a litigating party to avoid a negative inference from non-disclosure, so legal representation "buys a right to be silent without being prejudice" which ultimately affects the litigation outcome.

Besides illuminating 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. There is 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 may draw from nondisclosure of evidence.

More broadly, our results shed light on the role of advising in settings other than legal disputes. Agents and divisions often compete for resources within organizations. Resource allocation decisions — which could take such forms as merit assignment, promotion of the employees, budget allocation between divisions — in turn 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 basic insights regarding the role of advising in such circumstances.

8 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: Suppose that $\mathbb{E}[g|\hat{s},\hat{s},\theta] \geq 0$. Recall that

$$\mathbb{E}[g|\hat{s},\hat{s},\theta] := \int_0^1 g(s,\theta)f(s|\theta)ds + b_1\int_{\hat{s}}^1 g(s,\theta)f(s|\theta)ds + b_2\int_0^{\hat{s}} g(s,\theta)f(s|\theta)ds. \tag{7}$$

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

$$\int_0^1 g(s,\theta')f(s|\theta')ds > \int_0^1 g(s,\theta)f(s|\theta)ds \tag{8}$$

and

$$\frac{\int_{\hat{s}}^1 g(s,\theta')f(s|\theta')ds}{1-F(\hat{s}|\theta')} > \frac{\int_{\hat{s}}^1 g(s,\theta)f(s|\theta)ds}{1-F(\hat{s}|\theta)}. \tag{9}$$

Note that $\int_{\hat{s}}^1 g(s,\theta)f(s|\theta)ds \geq 0$, for otherwise (7) would imply that $\mathbb{E}[g|\hat{s},\hat{s},\theta] < 0$, in contradiction to our original assumption. Further, by MLRP, $1-F(\hat{s}|\theta') \geq 1-F(\hat{s}|\theta)$. So, $\int_{\hat{s}}^1 g(s,\theta')f(s|\theta')ds \geq \int_{\hat{s}}^1 g(s,\theta)f(s|\theta)ds$.

Thus, $\int_0^1 g(s,\theta)f(s|\theta)ds$, the first term in (7), is strictly increasing in θ, and $b_1\int_{\hat{s}_1}^1 g(s,\theta)f(s|\theta)ds$, the second term in (7), 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 (8) and (9), 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 (8) and (9), 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

$$\mathbb{E}[g|h(\theta),h(\theta),\theta] = \int_0^1 g(s,\theta)f(s|\theta)ds + b_1\int_{h(\theta)}^1 g(s,\theta)f(s|\theta)ds + b_2\int_0^{h(\theta)} g(s,\theta)f(s|\theta)ds. \tag{10}$$

The first term of (10) is increasing in θ by (8). Now consider the second term. We have:

$$\int_{h(\theta')}^1 g(s,\theta')f(s|\theta')ds = \int_{h(\theta)}^1 g(s,\theta')f(s|\theta')ds + \int_{h(\theta')}^{h(\theta)} g(s,\theta')f(s|\theta')ds > \int_{h(\theta)}^1 g(s,\theta)f(s|\theta)ds, \tag{11}$$

where the inequality follows from (9) 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

$$\int_0^{h(\theta)} g(s, \theta') f(s|\theta') ds > \int_0^{h(\theta)} g(s, \theta) f(s|\theta) ds. \tag{12}$$

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 (12).

Combining the observations, we conclude that (10) 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)

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

Similarly, party 2 discloses s if (only if)

$$P_0(s) > (\geq) \Pr\{\theta < h^{-1}(s) \mid s\}. \tag{14}$$

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

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

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 $s > \hat{s}$ ($s < \hat{s}$).[19] ▌

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

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

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

Hence, the optimality of the cutoff strategies with threshold $\hat{s}$, together with (13) and (14), 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 $h(\hat{\theta}^*)$.*[20]

Proof: Recall from (3) that

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

It then follows from Lemma A2 that

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

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

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

proving that the judge's cutoff default ruling strategy with threshold $\hat{\theta}^*$ is optimal. The uniqueness of the 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 $\hat{s} \in S$.

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

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 cutoff strategy with a threshold $\hat{\theta} \in \Theta$. We show that any combination of the parties' best response disclosure strategies leads to the same outcome, regardless of whether either party has obtained legal advice. 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 opponent 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. Therefore, 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 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 θ (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 to the contrary that there exists an equilibrium strategy combination $(\hat{\theta},\rho_1,\rho_2)$ s.t. $\hat{\theta}\notin[\hat{\theta}_-(b_1,b_2,c),\hat{\theta}_+(b_1,b_2,c)]$. Consider first $\hat{\theta}>\hat{\theta}_+(b_1,b_2,c)$. 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:

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

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

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

$$\begin{aligned}&-p_{00}\int_0^1 g(s,\hat{\theta})f(s|\theta)dsl(\hat{\theta}) - p_{10}\int_{h(\hat{\theta})}^1 g(s,\hat{\theta})f(s|\hat{\theta})dsl(\hat{\theta}) - p_{01}\int_0^{h(\hat{\theta})} g(s,\hat{\theta})f(s|\hat{\theta})dsl(\hat{\theta})\\ &\substack{>\\<}\ 0 \quad \text{if} \quad \hat{\theta}\substack{<\\>}\ \hat{\theta}^*(\tfrac{p_{10}}{p_{00}},\tfrac{p_{01}}{p_{00}}).\end{aligned}$$

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, implies 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 $\mathbb{E}[g|0,h(\theta),\theta;\frac{p_{10}}{p_{00}},\frac{p_{01}}{p_{00}},\frac{p_{11}}{p_{00}}]\ge 0$. 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 $\hat{\theta}^*(\frac{p_{10}}{p_{00}},\frac{p_{01}}{p_{00}}) = \hat{\theta}^*(\hat{b}_1,0)\ge\hat{\theta}_-(\frac{p_{10}}{p_{00}},\frac{p_{01}}{p_{00}},\frac{p_{11}}{p_{00}})$ 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 $(0,0,0)$. ▮

Proof of Proposition 9: Given the strategies adopted by parties 1 and 2, the judge's ruling strategy is rational under her Bayes-consistent beliefs. In particular, the symmetry between the two parties implies that the judge's threshold of $\frac{1}{2}$ is optimal when both parties are represented. We next show that the disputing parties' strategies are sequentially rational. Given the symmetry, it suffices to check only for party 1's incentives for deviation. The proof consists of several steps.

Step 1: *If party 1 observes $s \in [0, 1]$ and does not retain a lawyer and the judge and party 2 follow the candidate equilibrium strategies, then it is optimal for party 1 to disclose s.*

Proof: If party 1 discloses s, he wins if and only if $\theta < h^{-1}(s)$. So, his expected payoff is equal to

$$vL(h^{-1}(s)). \tag{15}$$

If party 1 does not disclose, the expected outcome depends on the value of s. Suppose, first, that $s < 1 - \hat{s}(w)$. With probability $1 - p$, party 2 does not observe s. He then hires a lawyer and makes no disclosure. With probability p, party 2 observes s. He then hires a lawyer and discloses s if $\theta > h^{-1}(s)$. As a result, the judge rules for party 1 if and only if $\theta < h^{-1}(1)$. So, party 1's payoff from nondisclosure is

$$vL(h^{-1}(1)). \tag{16}$$

Next, suppose that $s > 1 - \hat{s}(w)$. Then, if party 2 observes s, he does not retain a lawyer and discloses s. If party 2 does not observe s, then he retains a lawyer and makes no disclosure. Hence, given the judge's strategy, party 1's payoff from nondisclosure is

$$vpL(h^{-1}(s)) + v(1 - p)L(h^{-1}(1)). \tag{17}$$

Since $L(h^{-1}(s))$ is nonincreasing in s, each of (16) and (17) is less than (15). Therefore, if party l observes $s \in [0, 1]$ and does not retain a lawyer, it is optimal for him to disclose. ▌

Step 2: *If party 1 observes $s \in [0, 1]$ and the judge and party 2 follow the candidate equilibrium strategies, then it is optimal for party 1 to retain a lawyer if and only if $s \geq \hat{s}(w)$.*

Proof: If party 1 retains a lawyer, then it is optimal for him to disclose s if and only if $\theta \leq h^{-1}(s)$. Let us compute party 1's expected payoff associated with this strategy. If party 2 also observes s (which happens with probability p), then he discloses either if $s > 1 - \hat{s}(w)$ (without hiring a lawyer) or if $s \leq 1 - \hat{s}(w)$ and $\theta > h^{-1}(s)$ (after hiring a lawyer). Either way, party 1 wins if and only if $\theta \leq h^{-1}(s)$.

If party 2 does not observe s (which happens with probability $1 - p$), he retains a lawyer and does not disclose. Given the judge's default ruling strategy when both sides are represented, party 1 wins if and only if $\theta < \max\{h^{-1}(s), \frac{1}{2}\}$. Thus, party 1's expected payoff when he hires a lawyer after observing s is

$$vpL(h^{-1}(s)) + v(1 - p)L(\max\{h^{-1}(s), \tfrac{1}{2}\}) - w. \tag{18}$$

Suppose next that party 1 does not retain a lawyer after observing s. By Step 1, he will then always disclose s and receive the payoff given by (15). By (6), (18) is greater than (15) if and only if $s \geq \hat{s}(w)$. So the strategy of hiring a lawyer if and only if $s \geq \hat{s}(w)$ is, indeed, optimal. ▌

STEP 3: *If party 1 does not observe s and the judge and party 2 follow the candidate equilibrium strategies, then it is optimal for party 1 to retain a lawyer.*

PROOF: Suppose, indeed, that party 1 retains a lawyer. Given the other players' strategies, party 1 wins if $\theta \leq \frac{1}{2}$ and party 2 does not disclose, and if $\theta \leq h^{-1}(s)$ and party 2 discloses s. Hence, party 1's expected payoff is equal to

$$vp\left[K(1-\hat{s}(w))L(\tfrac{1}{2}) + \int_{1-\hat{s}(w)}^{1} L(h^{-1}(s))k(s)ds\right] + v(1-p)L(\tfrac{1}{2}) - w. \tag{19}$$

Meanwhile, if party 1 does not retain a lawyer, then his payoff becomes

$$vp\left[K(1-\hat{s}(w))L(h^{-1}(1)) + \int_{1-\hat{s}(w)}^{1} L(h^{-1}(s))k(s)ds\right] + v(1-p)L(h^{-1}(1)). \tag{20}$$

Subtracting (20) from (19) gives

$$\begin{aligned}&v\left[pK(1-\hat{s}(w)) + 1 - p\right](L(\tfrac{1}{2}) - L(h^{-1}(1)) - w > v(1-p)(L(\tfrac{1}{2}) - L(h^{-1}(1)) - w \\ &\geq v(1-p)(L(\tfrac{1}{2}) - L(h^{-1}(\hat{s}(w))) - w = 0.\end{aligned}$$

Hence, it is optimal for party 1 to retain a lawyer in this case. ▌

Steps 1-3 establish that party 1's candidate equilibrium strategy constitutes his best response to party 2's and the judge's strategies. By symmetry, the same is true for party 2. The optimality of the judge's strategy was established earlier. ▌

PROOF OF PROPOSITION 10: We first show that there is no equilibrium in which each party hires a lawyer with probability 1. The proof is by contradiction. Suppose to the contrary that such an equilibrium exists. Let $\hat{\theta}$ be the judge's threshold characterizing his default strategy in case of non-disclosure. If $h^{-1}(0) > \hat{\theta}$, then party 1 strictly prefers to remain unrepresented and disclose s when s is sufficiently close to zero. On the other hand, if $h^{-1}(0) \leq \hat{\theta}$ then $h^{-1}(1) \leq \hat{\theta}$ because $h^{-1}(.)$ is decreasing. In this case, party 2 strictly prefers to remain unrepresented and disclose s when s is sufficiently close to 1.

We next rule out the existence of equilibria in which some party, say party 1, remains unrepresented with probability 1 when he does not learn s, while some informed types of this party hire a lawyer with positive probability. The proof is also by contradiction. So, suppose such an equilibrium exists. Since party 1's actions do not affect the outcome when party 2 discloses, let us focus on the case in which party 2 does not disclose s. On the equilibrium path, party 2 does not disclose with positive probability, because he does not learn s with a positive probability.

Nondisclosure by both parties when party 1 is represented and party 2 has not deviated from the candidate equilibrium leads the judge to conclude that party 1 possesses evidence s s.t. $g(s, \theta) > 0$. This is because party 1 hires a lawyer only when he is informed and, by assumption, represented parties disclose all favorable evidence. So, the judge will always rule against represented party 1 if there is no disclosure and party 2's nondisclosure occurs on the equilibrium path. Therefore, in the candidate equilibrium, represented party 1 wins the dispute if and only if $g(s, \theta) < 0$. Yet, he could have done strictly better by not hiring a lawyer and disclosing with probability 1, generating a contradiction. ▮

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.

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.

Hadfield, G. and S. Leshem (2012), "Attorney Client Confidentiality," in *Procedural Law and Economics*, ed. Sanchirico, C.W. Encyclodedia of Law and Economics, Vol.8, 52-67.

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

Iossa,E. and B.Jullien (2012), "The Market for Lawyers and the Quality of Legal Services," *Rand Journal of Economics*, 43 (4), 677-704.

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.

Kartik, N., Suen, W., and Xu,F. (2013), "Multi-Sender Disclosure of Verifiable Information," mimeo, Honk Kong University.

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

Leshem, S. (2010), "The Benefits of a Right to Silence for The Innocent," *Rand Journal of Economics*, 41(2), 398416.

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.

Sanchirico, C.W. (1998), "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. (2000), "Games, information, and evidence production: with application to English legal history," *American Law and Economcis Reveiew* 2(2), 342-380.

Sanchirico, C.W. (2001), "Relying on the Information of Interestedand Potentially DishonestParties," *American Law and Economcis Reveiew* 3 (2), 320-357.

Sanchirico, C.W. and G. Triantis (2008), "Evidentiary Arbitrage: The Fabrication of Evidence and the Verifiability of Contract Performance," *Journal of Law, Economics, and Organization*, 24(1), 72-94.

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

Seidmann, D.J. and Winter, E. (1997), "Strategic Information Transmission with Verifiable Messages," *Econometrica*, 65, 163-170.

Shchepetova, A. (2014), "Inquisitorial vs. Adversarial system and the Right to be Silent," *University of Toulouse,* mimeo.

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.

Turkay, E. (2013), "Legal Punishments and Evidence Disclosure: A Game Theoretical Approach," mimeo, Vassar College.

Verrecchia, R.E. (1983), "Discretionary Disclosures," Journal of Accounting and Economics, 5, 179-194.

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