Statistical Evidence and the Problem of Robust Litigation

Jesse Bull* and Joel Watson†

This Version: March 2019

Subsequently published in the RAND Journal of Economics, 50 (2019, Issue 4, Winter): 974-1003.

Abstract

We propose a new model of disclosure, interpretation, and management of hard evidence in the context of litigation and similar applications. A litigant has private information and may also possess hard evidence that can be disclosed to a fact-finder, who interprets the evidence and decides a finding in the case. We identify conditions under which hard evidence generates value that is robust to the scope of rational reasoning and behavior. These fail if the litigant's private information is sufficiently strong relative to the "face-value signal" of evidence, and then hard evidence may be misleading. Rules that exclude some relevant hard evidence can be justified.

*Florida International University; bullj@fiu.edu.

†UC San Diego; jwatson@ucsd.edu.

The authors are grateful to Kathy Spier (the editor in charge of this article) and two anonymous referees for their many thoughtful and detailed comments, which stimulated significant improvements of this article. The authors thank Alex Weiss for superb research assistance and also the following people for their input: Navin Kartik, Dan Klerman, Aaron Kolb, Alex Lee, Herb Newhouse, Joel Sobel, Eric Talley, Simon Wilkie, and seminar participants at Indiana University, USC, UC Riverside, UCF, the 2015 Law and Economic Theory Conference at Duke University, the 2017 American Law and Economics Association Conference at Yale University, the 2016 Canadian Law and Economics Association Conference at the University of Toronto, the Fall 2017 Midwest Economic Theory Conference at the Southern Methodist University, and the 2018 Canadian Economic Theory Conference at the University of Toronto.

1 Introduction

In common-law countries, judges have latitude to exclude relevant evidence from being presented at trial and from being considered by juries. The United States' Federal Rules of Evidence (FRE) Rule 403 describes various circumstances that warrant exclusion, and it offers a limited explanation:

> The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence.

The related Rule 404 provides for a particular circumstance in stating that "Evidence of a person's character or character trait is not admissible to prove that on a particular occasion the person acted in accordance with the character or trait." Both of these rules have origins in common law tradition. The existence of these rules, and the broad judicial practice of excluding certain types of evidence, raises two questions: How can excluding relevant evidence improve accuracy in the deliberations of a jury, as Rule 403 suggests, and what kinds of evidence should be excluded?

We address these questions with a theoretical model that explores the interaction between a single litigant and a fact-finder whom we take to be a jury. Both the litigant and the fact-finder are sophisticated and rational. The fact-finder must evaluate a claim being made by the litigant, which amounts to estimating the value of an underlying *state* of the world. The litigant has private unverifiable information, described as the litigant's *type*. In some contingencies the litigant also possesses hard evidence and must decide whether to disclose it, if allowed. The fact-finder reviews whatever hard evidence is presented, updates its beliefs about the state (obeying Bayes' rule), and then issues a finding in the case. Applications extend beyond litigation, as we describe below.

In our model, the meaning of hard evidence is endogenous because the fact-finder must assess the litigant's disclosure strategy in order to interpret evidence. We show that constructive hard evidence requires the litigant and fact-finder to have similar beliefs about this interpretation. In some circumstances, it is possible for beliefs to be out of alignment and, as a consequence, hard evidence can mislead the fact-finder and reduce social welfare by more than the value of evidence. For a court that is sensitive to such a worst-case scenario, which we refer to as seeking "robust litigation," it is optimal to exclude relevant hard evidence under some conditions.

To delve into the logic, note that hard evidence is, by definition, *statistical* in nature: An individual piece of evidence exists with different probabilities in various states of the world. Hard evidence rarely provides definitive proof (that is, certainty) that the state is in some set, but it gives a signal that allows a fact-finder to update from a prior to a posterior probability distribution of the state.

Consider, for example, a trial that focuses on the question of whether the defendant robbed a particular store at 10:00 p.m. on a given date. The defendant may enter into evidence a time-stamped surveillance video showing him at a stadium 20 miles away at 9:20 p.m. on the same date (the hard evidence in this example). This piece of hard evidence does not prove with certainty that the defendant is innocent. It is possible that traffic conditions on the day of the crime were such that the defendant could, by leaving the stadium at 9:25 p.m. and speeding through the city, reach the store before 10:00 p.m.[1] However, the defendant's image on the stadium's 9:20 p.m. surveillance video is perhaps more likely to exist in the state of the world in which the defendant did not rob the store than it would in the state of the world in which he did.

We observe that hard evidence produces information through two channels. The first channel is the exogenous signal provided by the simple existence or nonexistence of the hard evidence. We call this the *face-value signal*. The second channel is the signal that disclosure or nondisclosure of hard evidence provides about the litigant's private information, and so we call this the *litigant-type* signal. The litigant-type signal has an endogenous element because the fact-finder's posterior belief about the litigant's type depends on the fact-finder's assessment of the litigant's disclosure strategy (the probability that each type of litigant would disclose the hard evidence when it exists). It is informative if different types of the litigant would disclose hard evidence with different probabilities.

Let us use the robbery sketch described above to illustrate these two channels and to show how evidence can be misleading. If the jury is assured of seeing the time-stamped video of the defendant at the stadium whenever such a video exists, then the jury would extract from its disclosure (or nondisclosure) exactly the face-value signal. In this case, disclosure of the video would cause the jury to revise upward its probability assessment that

[1] There may also be errors in the estimate of the time of the robbery and/or the video time stamp.

the defendant is innocent. But suppose the jury believes that a type of defendant who likely knows he committed the crime (the "bad type") would disclose the video evidence with high probability when it exists, whereas a defendant who knows that he didn't commit the crime (the "good type") would disclose with lower probability. Then disclosure would cause the jury to update in the direction of the bad type, but it could go the other way if the jury thought that the good type would be more likely to disclose video evidence.

Because both the face-value and litigant-type signals relate to the underlying state of interest—here the defendant's guilt or innocence—the jury combines them when updating about the state. If the face-value signal is strong compared to the bounds of the defendant-type signal, then the jury will update toward innocence regardless of the jury's belief about the defendant's strategy. In such a case, the defendant can conclude that disclosure of the surveillance video is sure to have a positive effect (from the defendant's point of view) and both types of defendant surely prefer to disclose, ensuring that the information provided by the disclosure is exactly the face-value signal.

But if the face-value signal is relatively weak compared to the defendant-type signal, then disclosure of the surveillance video could lead the jury to update toward *either* innocence or guilt, depending on the jury's belief about the defendant's disclosure strategy. And then it is possible—and consistent with rationality—for the good defendant to think the jury would interpret disclosure as a signal of guilt (the jury believing that only the guilty type of defendant would disclose), whereas the bad defendant has the opposite belief. Then the good defendant would not disclose the hard evidence and the bad defendant would disclose. Importantly, the jury could rationally think only the good defendant would disclose, which makes the jury update in precisely the wrong direction compared to what would happen if the jury knew the defendant's actual strategy. Hard evidence in this case is misleading and disadvantageous to society.[2]

To summarize, the example demonstrates that there are circumstances in which both types of defendant and the jury are sophisticated and rational Bayesians with a common

[2]The scenario can be embellished further by adding detail about the different types of potential defendants and the choices made by them, law enforcement officers, and others who influence whether the crime would be committed and whom might be charged. For example, one could imagine various types of defendant, including a sophisticated criminal who meticulously plans to visit the stadium and walk in front of a remote security camera before racing across town to rob the store. We comment on such "primary activity" in the next section and in the Conclusion. In this article we take all primary activity to be exogenous, focusing on the litigant's strategic disclosure choice and the fact-finder's interpretation.

prior belief about the fundamentals, these facts are *common knowledge* between them, and yet hard evidence is misleading to society's detriment. Notably, this is a non-equilibrium phenomenon, in that at least one player (or player-type) has an inaccurate belief about the actual strategy used by others. As we argue in the article, there is good reason to doubt that the players' beliefs in the settings studied here would happen to coincide, much less be coordinated on society's preferred equilibrium strategy profile. Therefore, our criterion for welfare analysis is *robustness*, which is the requirement that the litigation process delivers the intended (socially desirable) outcomes whether or not the fact-finder and all types of the litigant are coordinated on an equilibrium strategy profile.

To evaluate robustness, the solution concept we employ is *rationalizability*, which identifies the range of possible outcomes consistent with common knowledge of rationality. The robustness criterion evaluates policy alternatives by calculating the minimum expected welfare over all rationalizable strategy profiles of the litigation game. The main policy choice that we consider is whether to exclude hard evidence (individual pieces of evidence and/or bundles) from being presented by the litigant or being considered by the jury. We also comment on policies that would compel hard evidence to be disclosed and policies that would foster alignment of beliefs.

Our main result identifies conditions under which robust litigation justifies excluding relevant hard evidence because of its potential to mislead the jury. Specifically, the court should compare the face-value signal of evidence against the possible litigant-type signal, and exclude evidence when the latter outweighs the former. The model therefore offers support for the wording of Rule 403, with the term "probative value" taken to mean the face-value signal of hard evidence, and where "unfair prejudice," "confusing the issues," and "misleading the jury" are possible manifestations of relatively strong litigant-type signaling. So along with the theoretical results, the modeling exercise herein contributes to the literature by offering a methodology for modeling statistical hard evidence, particularly in distinguishing between the face-value signal and the litigant-type signal. The application of a robustness criterion for legal policy represents a secondary contribution, and along with it the demonstration that rationalizability is a useful concept for studying robustness. These elements are novel compared to the related literature, which we discuss at the end of the article.

Although we focus on evidence exclusion at trial, our results have other potential applications. Admissibility policies are relevant for the wide variety of settings where hard evidence is evaluated. For example, in many procurement auctions, principals review collections of materials (such as design documents, credentials, and other evidence) and award contracts based on a combination of the bid prices and a calculated score of the other elements. The selection of candidates for employment works in a similar way. To achieve robust selection processes in these settings, it may be useful to establish rules that exclude certain kinds of information from being submitted by bidders/candidates or from being considered by reviewers. Financial markets are another example. Whereas much attention has been paid to mandatory disclosure rules, such as those imposed by the U.S. Securities and Exchange Commission, there can also be restrictions on disclosure of potentially misleading information. For example, the effects of the August 7, 2018, announcement by Tesla CEO Elon Musk concerning funding to take the company private likely mixed verifiable information about the company with speculation about what Musk privately knew at the time concerning the company's financial health and ability to achieve production targets. His actions were deemed illegal and led to a costly settlement with the SEC.[3] We argue that, in developing policies to govern the disclosure of hard evidence, exclusion should be considered alongside mandatory disclosure.

The following section presents our basic model, which limits attention to a simple setting with one litigant, a jury that must decide a finding in the case, and one piece of hard evidence (which we call the "document") that the litigant may possess. In Section 3 we provide the following results: If a litigant has significant information about the state beyond what can be disclosed as hard evidence, then there is a problem of coordination of beliefs and behavior between the litigant and the jury, and hard evidence is misleading in some rationalizable outcomes. Further, the potential welfare loss of misleading evidence exceeds the potential gain of the face-value signal. In contrast, if a litigant's private information adds little to what can be disclosed as hard evidence, then there is a unique rationalizable outcome and, in a setting in which the document is positive evidence of the litigant's favored state, the litigant discloses the document whenever it exists. We provide additional details for the setting in which the litigant's type and the hard evidence are conditionally independent.

Section 4 discusses implications of the basic model for the courts, including steps that can help to align beliefs and two policy implications regarding admissibility of hard evidence. In Section 5 we extend our model to the case of two documents, which allows for an analysis

[3]See, for example, Michaels and Rapoport (2018).

of a wider range of evidentiary rules than in the basic model. We show that robust litigation requires that, under certain conditions, multiple documents should be admissible only as a bundle. A complication in the analysis is that disclosure of a single document may serve as both a signal of the litigant's private information and a signal of whether the other document exists.

A discussion of how our model and results relate to the law, including a numerical example based on a well-known case related to Rule 403, is contained in Section 6. We discuss the related literature in Section 7 and offer concluding remarks in Section 8. There we also provide some discussion of the setting with two litigants and possible extensions. Proofs of the theorems and a numerical example for Rule 404 may be found in the Appendices.

2 Basic Model

Description of the Game

We study a simple litigation game with hard evidence. There are two players: the *litigant* and the *jury*. The litigant can be a plaintiff or defendant in a legal case before the court. The jury is the court-designated fact-finder whom we model as a single agent. At issue is an underlying state of the world $\theta \in \Theta \equiv \{0, 1\}$, where $\theta = 0$ is called the "low state" and $\theta = 1$ is the "high state." The state represents whether the litigant's claim in the legal case is true ($\theta = 1$) or false ($\theta = 0$).

The litigant has two sources of information about the state. First, the litigant privately observes an unverifiable signal $x \in X$, where X is some arbitrary finite set. We occasionally refer to this signal as the litigant's "x-type." Second, the litigant may possess hard evidence, which is verifiable and can be disclosed to the jury. Suppose hard evidence takes the form of a single document d. We let $e \in E \equiv \{d, \emptyset\}$ denote the *evidentiary state*, where $e = d$ means that the document exists (and the litigant possesses it) and $e = \emptyset$ means the document does not exist.

The players interact as follows. At the beginning of the game, the underlying state θ, the evidentiary state e, and the private signal x are determined exogenously by nature according to the joint probability distribution f.[4] Neither player observes θ. The litigant privately

[4]So $f(\theta, e, x)$ is the probability that (θ, e, x) is realized. Although it would be appropriate to call the entire vector (θ, e, x) the *state*, for simplicity we shall sometimes refer to θ as the state (without the "underlying"

observes the signal x and the evidentiary state e.

Next, the litigant has an opportunity to present hard evidence to the jury, and there is a chance that the document, when it exists, will be disclosed exogenously. If $e = d$ then with probability ψ the document is disclosed exogenously, and with probability $1 - \psi$ the litigant has the choice of whether to disclose it or disclose nothing. In the event that $e = \emptyset$, there is no disclosure and the litigant has no choice; this is the defining characteristic of hard evidence. The parameter ψ captures the idea that the litigant may be influenced by a lawyer or other party who induces disclosure of available hard evidence, or that the judge may intervene to force disclosure. In the latter scenario, the judge would need to *independently* detect whether the document exists, so ψ represents the likelihood that the court learns of the document's existence and compels disclosure under penalty of contempt.

The jury observes whether d is disclosed but does not observe e or whether disclosure was due to exogeous forces. Then the jury takes an action $a \in [0, 1]$, which in practical terms is a finding in the case. This action can represent, for instance, the degree to which the litigant is held responsible for a crime or the amount of monetary damages to award the litigant. Findings closer to 1 are more in the litigant's favor relative to findings closer to 0.

We assume that the jury has society's preferences, given by the payoff function $u_{\mathrm{J}}(a, \theta) = -(a - \theta)^2$. This implies that the jury's optimal action is equal to its posterior probability of the high state, after updating on the basis of whatever hard evidence was disclosed. We assume that the litigant's payoff is given by $u_{\mathrm{L}}(a, \theta) = a$, so the litigant's interest is to act in whatever fashion will maximize the jury's posterior probability of $\theta = 1$.[5]

A mixed (behavior) strategy for the litigant is given by a function $\beta\colon X \to [0, 1]$, where for each $x \in X$, $\beta(x)$ is the probability that the litigant chooses to disclose the document in the event that $e = d$ and his private signal is x. Due to the possibility of exogenous disclosure, the total probability of disclosure is thus given by $\sigma(x) \equiv \psi + (1 - \psi)\beta(x)$. Note that choosing β is equivalent to choosing $\sigma\colon X \to [\psi, 1]$. Because it will be convenient in our analysis to deal directly with σ, we shall therefore refer to σ as the litigant's strategy,

qualifier) because this is what is of direct interest to the jury.

[5]For simplicity, we assume that the litigant cares only about the jury's action. We could allow u_{L} to be a function of the state, the litigant's private signal, the realization of hard evidence, and whether the litigant discloses the document, but such ingredients are not necessary for the logical connections that we focus on. We could also model the jury as having only two available actions, such as finding the litigant guilty or not guilty, and obtain much the same analysis if the jury is influenced by a noisy information source in addition to the information received from the litigant's dosclosure choice.

with the understanding that $\sigma(x) \geq \psi$ for every x. We assume that $\psi \in [0,1)$, so $\psi = 0$ is allowed.

Assume that the foregoing description is common knowledge between the players. To recap, in this incomplete-information game an exogenous random draw determines (θ, e, x). The litigant obtains e and also observes x. Then, if $e = d$, disclosure of d occurs exogenously with probability ψ and otherwise the litigant decides whether to disclose d. Finally, the jury observes only whether d is disclosed, forms its posterior belief about the state θ, and selects its action a.

Regarding notation, we will sometimes evaluate f over sets. For example, we write expressions such as $f(\theta, e, K)$ for $K \subset X$, which is the probability that the underlying state is θ, the evidentiary state is e, and the private signal is an element of K. Let $r \equiv f(1, E, X)$ denote the marginal probability that the underlying state is high and assume that $r \in (0,1)$. It will sometimes be useful to write the probability of e and x conditional on θ, which is given by the standard conditional-probability formula:

$$f(e, x \mid \theta) \equiv \frac{f(\theta, e, x)}{f(\theta, E, X)}.$$

We shall assume that $f(\Theta, d, X) \in (0,1)$ and that $f(\Theta, d, x) > 0$ for all $x \in X$. We also assume that $f(d, x \mid 1) \neq f(d, x \mid 0)$ for some x, for otherwise hard evidence conveys no information.

A Note on Litigant Types and Primary Activity

Our model describes a strategic situation between a litigant and jury, conditional on the case being in court. To analyze a real-world application, it can be helpful to describe how events that would lead to a court case imply the distribution f. Developing the context, we see that different types of litigant in our model are typically different people in the real world. For instance, consider the following simple story about the events leading to a court case.

There are a variety of individuals in society, differing in their propensity to commit a crime and, if so, how to go about it. Their behavior and some exogenous random forces lead to an outcome of preliminary activity, which includes whether and how a crime is committed, evidence relevant to the crime, and the detainment by the police of an individual who is brought to trial. It is possible that this defendant—the litigant in our model—is a legitimate suspect but actually did not commit the crime, just as it is possible that the defendant did in fact commit the crime. These two types of defendant are different people in the society and their personal backgrounds are, to the extent not observable to law enforcement, captured in the x variable.

If the question before the jury is whether the defendant's culpability exceeds a particular evidentiary standard, and if the defendant has some understanding about whether he performed a criminal act, then a component of x is correlated with θ but is not necessarily perfectly correlated. That is, the defendant may have information about whether he is culpable but not know precisely whether his behavior exceeds the cutoff for a guilty verdict or for a particular sentence. For instance, the litigant may lack an understanding of the law or be unsure of whether his behavior was criminal. If, in this example, the defendant knows precisely whether he or she committed the crime and this is the question that the jury considers, then a component of x would be perfectly correlated with θ.

Consideration of the social backdrop also demonstrates how natural it is for there to be correlation between e and x, conditional on the underlying state θ. Take the store robbery example in the Introduction and suppose d denotes the litigant (defendant) being on the recording of the stadium security camera at a time that would make it challenging for him to have traveled to the store and committed the robbery. Suppose that X is partitioned into four subsets representing four different groups of people in the society: I, I', G, and G'. Types in I and I' would never commit a criminal act and those in I happen to be on the stadium video, types in G are sophisticated criminals who plan to make an appearance in front of the camera at the stadium before racing to the store to commit the crime, and types in G' are naive criminals who would commit the crime on the spur of the moment and would not be on the stadium video. Assume that x is randomly drawn, the crime occurs if $x \in G \cup G'$, and there is some randomness in police work so that with some probability an innocent person is the one brought to trial. Then, along with the x already defined, we have $\theta = 1$ if $x \in I \cup I'$ and $\theta = 0$ otherwise. The video evidence exists, so that $e = d$, if $x \in I \cup G$. The distribution f is then defined from the distribution of x and the randomness induced by the police work to identify a suspect, *conditional on a crime occurring*. In this example, x and e are correlated overall and they are correlated conditional on θ.

Solution Concept

We shall analyze the litigation game using the solution concept of *rationalizability*, which assumes it is common knowledge that the players form beliefs about each other and best respond to their beliefs. The set of rationalizable strategy profiles contains all of the profiles consistent with this assumption. For $\psi > 0$ both of the jury's information sets in the game are always reached with positive probability, so standard normal-form rationalizability suffices. In the setting of $\psi = 0$, with some strategy profiles the jury's no-disclosure information set is not reached. For this setting, technically we are characterizing extensive-form rationalizability and we add the assumption that the players' beliefs are *plainly consistent* (Watson 2017), which implies Bayes'-rule updating and some structure on the jury's belief at the no-disclosure information set as described below. Importantly, in a rationalizable outcome it is not necessarily the case that one player's beliefs are accurate about the other player's beliefs and behavior. Depending on parameters, there may be a rationalizable outcome in which the litigant has an incorrect belief about the jury's reasoning, or vice versa.

The rationalizability concept is appropriate for settings in which the litigant and jury lack experience in dealing with each other, and where the legal institution and social norms would not be expected to completely coordinate the litigant's and jury's beliefs and behavior. For example, a significant fraction of civil and criminal cases feature a litigant who has had little previous experience in court and who has not faced the same circumstances before. Most jurors also have limited experience in fact-finding. These players may be able to engage in sophisticated reasoning and understand each other's incentives and rationality, but still not be fully coordinated.

With the rationalizability concept, we do not require the beliefs of different types of litigant to be the same. Indeed, we think it is important to allow for *non-aligned types*, whereby the litigant's beliefs may depend on his private signal and hard evidence. The main justification for this is that, as noted already, the various litigant x-types typically refer to different people in a population, and there is no reason to believe that different people have exactly the same beliefs. Non-aligned types will play an important role in our theory.

Much of our analysis will be put in terms of the jury's posterior belief regarding the state, conditioned on whether the document is disclosed. Let $b(d)$ denote the posterior probability of the high state in the event that d is disclosed and let $b(\emptyset)$ be the probability of the high state in the event that d is not disclosed. These values define the jury's *interpretation* of hard evidence. It is important to recognize that $b(d)$ and $b(\emptyset)$ depend on the jury's belief about the litigant's strategy σ, as well as the jury's understanding of the information system.

The jury's initial belief (at the beginning of the game) about the litigant's strategy is given by a function $\lambda\colon X \to [\psi, 1]$, where for each $x \in X$, $\lambda(x)$ is the probability with which the jury thinks the document is disclosed in the event that $e = d$ and the litigant's private signal is x. We can determine the jury's posterior beliefs in terms of λ and the fundamentals of the model. Note that $f(1, e, x) = rf(e, x \,|\, 1)$ and $f(0, e, x) = (1-r)f(e, x \,|\, 0)$. If $\lambda(x) > 0$ for some x, then the jury's posterior belief conditional on disclosure is given by Bayes' rule:

$$b(d) = \frac{\sum_{x \in X} rf(d, x \,|\, 1)\lambda(x)}{\sum_{x \in X} \left[rf(d, x \,|\, 1) + (1-r)f(d, x \,|\, 0)\right] \lambda(x)}. \tag{1}$$

Note that the denominator is strictly positive if and only if $\lambda(x) > 0$ for some x, because we have assumed that $f(\Theta, d, x) > 0$ for all $x \in X$. Clearly $\psi > 0$ is a sufficient condition.

In the case of $\lambda(x) = 0$ for all x, which is possible only in the setting where $\psi = 0$, the denominator is zero and Bayes' rule overall does not apply. However, as we show in Appendix A, plain consistency implies that Equation 1 must still hold with λ replaced by another function λ' that describes the jury's *updated* belief about the litigant's strategy. Thus, Equation 1 shall always be valid, with the understanding that λ may refer to an updated belief about σ.

We have a similar Bayes' rule expression for the jury's posterior belief conditional on nondisclosure. It is always valid for the jury's initial belief about the litigant's strategy, because we have assumed that with positive probability the document does not exist.

As noted already, the jury's optimal action is to match its posterior belief about the state, so it selects $a = b(d)$ in the event that the document is disclosed and $a = b(\emptyset)$ if the document is not disclosed. Likewise, the litigant's optimal strategy depends on each type's belief about the jury's strategy. For any given type of the litigant, this type optimally discloses the document if, according to his beliefs, the expected action of the jury conditional on disclosure weakly exceeds the expected action of the jury conditional on nondisclosure. To express this condition in terms of the jury's beliefs, a litigant type weakly prefers disclosure if, according to his belief, the expected $b(d)$ is weakly greater than the expected $b(\emptyset)$. A litigant type optimally withholds the document if his expected $b(d)$ is weakly below his expected $b(\emptyset)$.

Welfare and Admissibility

Social welfare is measured by the jury's actual payoff. We will take the expectation with respect to the distribution f, calling this the *expected actual payoff of the jury.* Note that this may differ from the expected payoff in the mind of the jury, because what the jury expects and what actually happens may differ in a rationalizable outcome.

There are two key elements of welfare analysis. First, we want to identify whether hard evidence is useful in the litigation process, in terms of raising the jury's actual payoff, and we want to quantify the extent to which hard evidence can function in a misleading way that lowers welfare. Second, we investigate the design of admissibility rules with a goal of robust litigation, which is to find rules that ensure hard evidence plays a constructive role in the litigation process.

The *face-value signal* of the hard evidence is the marginal signal provided by the existence or nonexistence of the document, averaging over the litigant's private signal x. If $f(d, X \mid 1) > f(d, X \mid 0)$ then we say that the document is *positive evidence* of the high state and the absence of the document is *negative evidence* of the low state (Bull and Watson, 2004). If $f(d, X \mid 1) < f(d, X \mid 0)$ then we say the opposite—the document is *positive evidence* of the low state.[6]

On the welfare front, we use the following notation. Let U_{J}^{*} denote the jury's expected payoff in an artificial setting in which the jury directly observes whether the document exists (but does not observe x), forms the proper posterior belief, and best responds. Let U_{J}^{0} denote the jury's expected payoff in an artificial setting without hard evidence, where the jury must choose a without interacting with the litigant. Then $U_{\mathrm{J}}^{*} - U_{\mathrm{J}}^{0}$ is the *face value of hard evidence*, in other words the welfare gain due to the face-value signal provided by hard evidence.

Let us say that *hard evidence is ineffective* if the litigant plays a strategy other than full disclosure ($\sigma = 1$), because in such a case the jury does not always see the document when it exists and therefore is unable to benefit fully from its face-value signal. Further, let $b^{\sigma}(d)$ and $b^{\sigma}(\emptyset)$ be the jury's posterior probabilities under the assumption that the jury knows the litigant's actual strategy (and so $\lambda = \sigma$), where we recognize that the former is defined

[6] Extreme cases of *absolute proof* are given by $f(d, X \mid 1) = 1$ and $f(d, X \mid 0) = 0$, where disclosure of d proves that the state is high, and $f(d, X \mid 1) = 0$ and $f(d, X \mid 0) = 1$, where disclosure of d proves that the state is low. Another extreme has $f(d, X \mid 1) = f(d, X \mid 0) = 1$, which is the case of a *cheap document*, but the assumptions made already rule this out.

only if $\sigma(x) > 0$ for some x. Then we say that *hard evidence is misleading* if $b^\sigma(d)$ is well defined and yet $b(d) \neq b^\sigma(d)$ and/or $b(\emptyset) \neq b^\sigma(\emptyset)$, so that the jury's posterior beliefs are not consistent with the litigant's actual strategy.

Finally, let $\underline{U}_{\mathrm{J}}$ denote the lowest expected actual payoff of the jury over the rationalizable outcomes of our litigation game, so that $U^*_{\mathrm{J}} - \underline{U}_{\mathrm{J}}$ is the potential welfare loss due to hard evidence being ineffective or misleading. The *potential loss ratio* is defined to be the ratio

$$L = \frac{U^*_{\mathrm{J}} - \underline{U}_{\mathrm{J}}}{U^*_{\mathrm{J}} - U^0_{\mathrm{J}}}.$$

Note that $L \geq 0$. If we replace $\underline{U}_{\mathrm{J}}$ with the jury's expected actual payoff, then L becomes the *actual loss ratio*.

3 Conditions For Misleading Hard Evidence

Whether hard evidence can turn out to be ineffective or misleading depends on the strength of the litigant's private signal relative to the strength of the hard evidence. To explore the connection, let us examine the possibilities for $b(d)$ relative to r. Define:

$$K^+ \equiv \{x \in X \,|\, f(d,x\,|\,1) \geq f(d,x\,|\,0)\} \quad \text{and} \quad K^- \equiv \{x \in X \,|\, f(d,x\,|\,1) \leq f(d,x\,|\,0)\}.$$

Note that K^+ is the set of litigant x-types for which the *combination* of the occurance of x and existence of the document is weakly positive evidence of $\theta = 1$, whereas K^- is the set of litigant x-types for which the combination of the occurrence of the type and existence of the document is weakly positive evidence of $\theta = 0$. The following conditions compare these combination signals to the face-value signal of the hard evidence:

(D) $\quad K^- \neq \emptyset$ and $\psi\,[f(d,X\,|\,1) - f(d,X\,|\,0)] \leq (1-\psi)\,[f(d,K^-\,|\,0) - f(d,K^-\,|\,1)]$

(U) $\quad K^+ \neq \emptyset$ and $\psi\,[f(d,X\,|\,1) - f(d,X\,|\,0)] \geq (1-\psi)\,[f(d,K^+\,|\,0) - f(d,K^+\,|\,1)]$.

The bracketed term on the left side of each of these inequalities is the strength of the document's face-value signal; this term is greater than zero if d is positive evidence of the high state and less than zero if it is positive evidence of the low state. Similarly, the bracketed terms on the right side of these inequalities give the strength of the combination signal for the x-types in K^- and K^+, where the K^- term is nonnegative and the K^+ term is nonpositive.

We use the letters "D" and "U" in reference to these conditions because, as the following lemma shows, the first relates to whether the jury can revise its belief about the state downward conditional on disclosure of d, and the second relates to upward revision. Let us call λ "feasible" if it is a function $\lambda : X \to [\psi, 1]$ satisfying $\lambda(x) > 0$ for some x.

Lemma 1: *Let $b(d)$ be given by Equation 1. Condition D holds if and only if there is a feasible λ such that $b(d) \leq r$. Likewise, Condition U holds if and only if there is a feasible λ such that $b(d) \geq r$.*

Proof: Let us define beliefs λ^+ and λ^- as follows:

$$\lambda^+(x) = \begin{cases} 1 & \text{if } x \in K^+ \\ \psi & \text{if } x \in X \backslash K^+ \end{cases} \quad \text{and} \quad \lambda^-(x) = \begin{cases} 1 & \text{if } x \in K^- \\ \psi & \text{if } x \in X \backslash K^- \end{cases} .$$

Let us write $b(d) \leq r$ using Equation 1 to substitute for $b(d)$, and then rewrite the expression by substituting $f(d, x \,|\, \theta)\lambda(x) = f(d, x \,|\, \theta)\psi + f(d, x \,|\, \theta)(\lambda(x) - \psi)$. Factoring and dividing by $(1 - r)$ produces the following inequality:

$$\psi \left[f(d, X \,|\, 1) - f(d, X \,|\, 0) \right] \leq \sum_{x \in X} \left[f(d, x \,|\, 0) - f(d, x \,|\, 1) \right] (\lambda(x) - \psi) . \tag{2}$$

Clearly λ^- maximizes the right side of Inequality 2 by choice of $\lambda : X \to [\psi, 1]$. Plugging in λ^- yields the inequality of Condition D.

Therefore, if for some feasible λ we have $b(d) \leq r$, then λ satisfies Inequality 2 and this further implies the inequality in Condition D. The existence of such a λ also implies $K^- \neq \emptyset$. We verify this using a contradiction argument. If $K^- = \emptyset$ and $\psi > 0$ then the left side of Inequality 2 is strictly positive, so the right side must also be strictly positive, implying that $f(d, x \,|\, 0) > f(d, x \,|\, 1)$ for some x (contradicting $K^- = \emptyset$). If $K^- = \emptyset$ and $\psi = 0$ then the left side of Inequality 2 is zero, and so the right side is nonnegative. But then by feasibility of λ we must have $\lambda(x) > 0$ for some x, and this implies that $f(d, x \,|\, 0) \geq f(d, x \,|\, 1)$ for some x (again contradicting $K^- = \emptyset$). To prove the converse relation, observe that if Condition D holds, then λ^- is feasible and satisfies Inequality 2.

The relation $b(d) \geq r$ is equivalent to the reverse of Inequality 2. Repeating the argument above for K^+ and λ^+ establishes the second claim of the lemma. $\square$

For intuition, observe that if the document is positive evidence of the low state, so that $f(d, X \mid 1) \leq f(d, X \mid 0)$, then Condition D trivially holds. If the document is strictly positive evidence of the high state, where $f(d, X \mid 1) > f(d, X \mid 0)$, then the inequality for Condition D requires that this face-value signal can be offset by evidence of the low state in the form of the existence of the document in combination with an x-type in K^-. Likewise, if $f(d, X \mid 1) \geq f(d, X \mid 0)$ then Condition U trivially holds. Otherwise, the inequality of Condition U requires that the x-types in K^+, in combination with the document existing, provide sufficiently strong positive evidence of the high state.

It is not difficult to verify that Conditions D and U are both satisfied if (i) the document exists with strictly positive probability in both underlying states, (ii) ψ is sufficiently small, and (iii) the litigant's private signal provides a sufficiently accurate indication of the underlying state. Suppose, for instance, that the litigant precisely knows the state. In the example of a criminal trial, this would mean that the defendant knows perfectly whether his culpability exceeds the threshold for guilt. Then K^+ is the set of x-types that know the state is high and K^- are the x-types that know the state is low, so $f(d, K^+ \mid 1) > 0$, $f(d, K^+ \mid 0) = 0$, $f(d, K^- \mid 0) > 0$, and $f(d, K^- \mid 1) = 0$. In this case, the inequalities of Conditions D and U trivially follow for a small enough value of ψ. This is an extreme example. Real litigants would generally not know the underlying state perfectly because of uncertainty regarding standards of proof and the law.

It is exactly when Conditions D and U are both satisfied that there are rationalizable outcomes in which hard evidence is misleading. The root cause is miscoordination between the players' beliefs and behavior. For example, take the setting in which the litigant is a defendant who knows his culpability. Suppose a guilty litigant (with $x \in K^-$) believes that the jury's posteriors satisfy $b(d) \geq r$ as Condition U allows; we show in the proof of Theorem 1 below that this implies $r \geq b(\emptyset)$ as well. Suppose further that an innocent litigant (with $x \in K^+$) has the opposite belief, thinking that $b(d) \leq b(\emptyset)$ as Condition D allows. Both kinds of beliefs are rational because they are consistent with feasible beliefs and proper updating by the jury. These beliefs would lead the guilty litigant to disclose the document and the innocent litigant to not disclose. Further suppose that the jury actually believes that the innocent x-type would disclose and the guilty x-type would not disclose. Then the jury gets the hard evidence signal backward and evidence is misleading. Importantly, every x-type of litigant and the jury are behaving rationally and fully incorporate the rationality of the other player-types, so the outcome is rationalizable.

Lemma 1 leads to the following general result on conditions for misleading or ineffective hard evidence.

Theorem 1: *If Condition U fails, then the unique rationalizable outcome has every litigant type disclosing d at minimum probability ψ, the actual loss ratio is $L \in (0, 1]$, and hard evidence is ineffective. If Condition D fails, then the unique rationalizable outcome has the litigant disclosing d whenever he has it, the actual loss ratio is $L = 0$, and hard evidence is effective.*[7] *If Conditions D and U are both satisfied, then the potential loss ratio is $L > 1$ and there are rationalizable outcomes in which hard evidence is misleading.*

To explore the scope of the third case, where Conditions D and U both hold, consider equilibrium outcomes as a benchmark. Under Conditions D and U, if additionally $f(d, X \mid 1) < f(d, X \mid 0)$, then every perfect Bayesian equilibrium of the litigation game features minimal disclosure, so hard evidence is ineffective and the actual loss ratio is in the interval $(0, 1]$. In the subcase where $f(d, X \,|\, 1) > f(d, X \,|\, 0)$ and $\psi > 0$, the unique perfect Bayesian equilibrium features full disclosure and the actual loss ratio is 0. Finally, if $f(d, X \mid 1) > f(d, X \mid 0)$ and $\psi = 0$, and assuming $K^- \neq \emptyset$, then there are multiple perfect Bayesian equilibria, including ones in which the document is never disclosed (so the actual loss ratio is 1) and one in which the document is always disclosed (with an actual loss ratio of 0). Overall, in an equilibrium there is no misleading evidence and the loss ratio is always between 0 and 1. But regardless of the subcase of f, there are rationalizable outcomes in which hard evidence is misleading and the loss level strictly exceeds 1, so welfare falls strictly below what can be achieved in any equilibrium.[8]

The conditions for misleading evidence are particularly simple in the setting in which e and x are conditionally independent. For this setting, let q_θ denote the probability that $e = d$ conditional on θ, and let $p_\theta(x)$ denote the probability that the litigant's private

[7] Although rationalizability assumes common knowledge of rational behavior, to prove the first two claims we need only that players are rational (best responding to their beliefs) and know this about each other.

[8] See Watson (2017) for the perfect Bayesian equilibrium definition, which is equivalent to sequential equilibrium for the game studied here. The equilibrium benchmark also is helpful in motivating the prospect of miscoordinated beliefs and behavior. Suppose that $\psi = 0$ and $f(d, X \,|\, 1) > f(d, X \,|\, 0)$. Then the various types of litigant and the jury could all believe that they are playing a perfect Bayesian equilibrium of the litigation game, but they may not have the same equilibrium in mind. For instance, one type of litigant may behave according to an equilibrium in which the document is never disclosed, while another behaves according to an equilibrium in which the document is always disclosed. Whatever equilibrium the jury thinks is being played, evidence will turn out to be misleading.

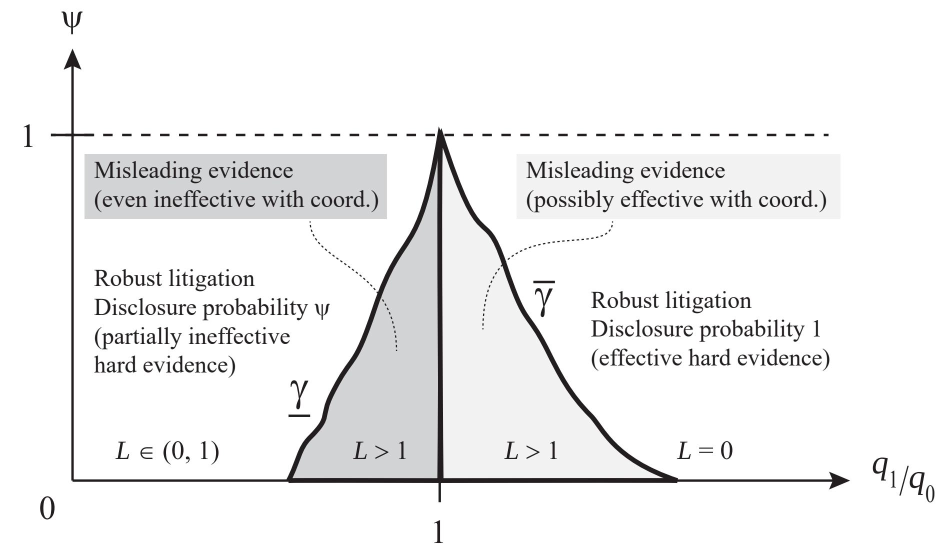

Figure 1: Parameter regions for misleading evidence.

signal is x conditional on θ. Thus $f(d, x \mid \theta) = q_\theta p_\theta(x)$ and $q_\theta = f(d, X \mid \theta)$. Recall that $q_1/q_0 < 1$ means that d is positive evidence of the low state, the opposite inequality means d is positive evidence of the high state, and pushing this ratio farther from 1 represents a stronger face-value signal.

Theorem 2: *Suppose that hard evidence e and private signal x are independent conditional on the underlying state θ. There are functions $\underline{\gamma} : [0,1] \to (0,1]$ and $\overline{\gamma} : [0,1] \to [1,\infty)$, which depend on the distributions $p_1(\cdot)$ and $p_0(\cdot)$, with the following properties. For any given ψ, Condition D is equivalent to $q_1/q_0 \leq \overline{\gamma}(\psi)$ and Condition U is equivalent to $q_1/q_0 \geq \underline{\gamma}(\psi)$. Furthermore, $\underline{\gamma}$ is increasing in ψ, $\overline{\gamma}$ is decreasing in ψ, $\underline{\gamma}(1) = \overline{\gamma}(1) = 1$, $\overline{\gamma}(0) \equiv \max_{x \in X} p_0(x)/p_1(x)$, and $\underline{\gamma}(0) \equiv \min_{x \in X} p_0(x)/p_1(x)$.*

Note that $\overline{\gamma}(0)$ and $\underline{\gamma}(0)$ are the maximum and minimum of the likelihood ratio (low state to high state) of the litigant's private signal realizations. They provide bounds on the highest and lowest posterior probability that the litigant can have about the low state based on only his private signal. Thus, in the case of $\psi = 0$, the conditions for misleading evidence come down to a straightforward comparison between the likelihood ratio of the face-value signal and the maximum and minimum likelihood ratios of the litigant's private signal.

Overall, when the face-value signal of hard evidence is strong relative to the litigant's private signal, there is no concern of misleading evidence. If in addition the document is positive evidence of the high state, then hard evidence is effective because the litigant always discloses it. However, when the information provided by the private signal is strong relative to the face-value signal of hard evidence, then there is scope for the evidence to be misleading, with the welfare loss exceeding the welfare gain of the face-value signal. The scope for misleading evidence shrinks as ψ increases. These findings are illustrated in Figure 1 for the setting in which e and x are conditionally independent.

Although it may be useful to determine how changes in parameter values affect L, unfortunately calculating the potential loss is difficult and we cannot provide a clean comparative-statics result. However, one can uncover a property of L for parameter values near the cutoff for robust litigation, and the news is not so good. Specifically, for parameters close to the region of robust litigation, the potential loss ratio is bounded away from 1.

Theorem 3: *Fixing ψ, suppose that hard evidence e and private signal x are independent conditional on the underlying state θ and that $\underline{\gamma}(\psi) < \overline{\gamma}(\psi)$. Let q_1 be bounded away from 0 by a fixed number. Then there is a number $\underline{L} > 1$ so that, for all q_0 and q_1 satisfying $\underline{\gamma}(\psi) \leq q_1/q_0 \leq \overline{\gamma}(\psi)$, the potential loss ratio satisfies $L \geq \underline{L}$.*

4 Implications for Courts

The prospect of misleading evidence is, in our view, an appreciable problem that the legal system should address. Litigants and juries arrive in court typically having little experience with similar settings and with each other; so, lacking an institution that would coordinate them, there is little reason to expect that their beliefs and behavior would be aligned.[9] Even if some measure of coordination could be achieved, theory does not provide much hope that the parties will coordinate on society's preferred outcome.[10] Importantly, the legal system actually does address the issue of misleading evidence, as we noted in the Introduction and discuss further in Section 6.

[9] In fact, the jury-selection process typically ensures that jurors have not had prior interaction with the litigants.

[10] Standard equilibrium refinements do not restrict the equilibrium outcomes. The intuitive criterion (Cho and Kreps 1987) has no bite in the basic model. The divinity criterion (Banks and Sobel 1987) has no bite in the slightly more general model in which the litigant's preferences can depend on the state and there is a small cost of disclosure; for instance, the litigant's marginal value of the court's action is slightly higher in the low state than in the high state.

Our modeling exercise suggests that, when developing evidentiary policies, the legal system should strive to achieve *robust litigation*. Robustness means looking at all rationalizable outcomes, in particular the worst-case scenarios for society, with the recognition that parties' beliefs and behavior may not be coordinated even though it is common knowledge that they are rational. We next discuss three options for policy, all having some bearing on reality.

The first policy option is for the court to be transparent regarding its interpretive rules, meaning that it articulates standards for how some common evidentiary actions should be interpreted. This may help to align the beliefs of the various litigant x-types and also put them into alignment with the jury's beliefs. However, projecting complete transparency will typically be impossible because of the great many contingencies that the court would have to explain before litigation commences in a given case. Further, such an intervention assumes that the court fully understands the background of each case and has the jury's knowledge about the statistical details of evidence, which voids the function of the jury as fact-finder.

In a related vein, experienced attorneys may help align the beliefs of the litigant types, in particular if they represent a variety of types in similar cases and give consistent advice to these clients. One may even argue that some decisions regarding evidence disclosure are effectively made by the attorneys. But however assertive attorneys may be, litigants still have strategic choices about whether to make their attorneys aware of hard evidence and other information. Therefore, we would not expect beliefs to be completely aligned across litigant types. Further, an attorney's plan for trial may be influenced by the details of the case and the composition of the jury, and this plan could be correlated with the litigant type.

Although an experienced attorney may push litigant types toward alignment, it is much less likely that a jury, whose members rarely hear cases, would have coordinated beliefs. In fact, even if the beliefs of the litigant types are actually aligned, evidence can still be misleading so long as the jury may entertain the possibility of non-aligned types. To be precise, under Conditions D and U when hard evidence is allowed, there are rationalizable outcomes in which the litigant types have aligned beliefs, evidence is misleading, and the loss ratio L strictly exceeds 1.[11]

[11] For instance, suppose that the document is positive evidence of $\theta = 1$. If the jury believes that types in K^+ would not disclose the document, whereas types in $X \setminus K^+$ would disclose, the jury's posterior beliefs satisfy $b(d) < b(\emptyset)$. If, with aligned beliefs, all litigant types disclose the document, then hard evidence is providing its full face-value signal and yet the jury updates in the opposite direction.

The second policy option is to compel the litigant to disclose hard evidence under some conditions. In terms of the model, such a legal rule would have the effect of raising the parameter ψ, which is unambiguously good for society given the robustness criterion.[12] However, recall that enforcing mandatory disclosure requires the court to directly determine whether the document exists, without relying on the litigant for this information. For many kinds of evidence, the judge would have no way to do this at low cost, even in an inquisitorial system. Still, there are cases in which it is possible to enforce mandatory disclosure. For example, mandatory disclosure may be possible for a standard document that the court is familiar with.[13]

Policy Implication 1: *The court should impose a mandatory disclosure policy for a document that the court knows to be relevant, is free to produce, and whose existence or nonexistence the court can independently verify.*

The third policy option is to establish *admissibility rules*, which explicitly *disallow* disclosure of certain kinds of hard evidence. These rules may be imposed in two ways. The court could simply refuse to allow the document to be presented in some cases. Alternatively, the court could instruct the jury to treat disclosure or lack thereof as providing no information whatsoever.

Our model provides a justification for excluding evidence under some conditions. The criteria for admissibility can be expressed in terms of the potential loss measure and the nature of the document (whether it is positive evidence of the high state or the low state). Let us assume that ψ is optimally set, as high as possible, for a given case. Then, seeking robust litigation, the court optimally makes the document d inadmissible if L is large, in

[12] Any welfare level that can be achieved for some value ψ must also be achievable for a lower value ψ'. For instance, suppose that in the setting of ψ, both disclosure and nondisclosure can be rationalized for two rational beliefs of the litigant. Then these two beliefs still rationalize disclosure and nondisclosure under ψ', and this implies that the corresponding set of rational beliefs for the jury includes all of the beliefs that were rational under ψ. Intuitively, raising ψ decreases the potential loss ratio by reducing the range of other parameters under which evidence may be misleading (tightening the inequalities for Conditions D and U). Raising ψ also directly increases the minimum probability on evidence disclosure, which is particularly beneficial if the document is positive evidence of the low state (where in equilibrium, for example, the litigant discloses at the minimum probability).

[13] In some cases it is known that documentation exists to verify or disprove a claim, such as the amount τ that a person paid in income taxes. The relevant issue may be whether $\tau \geq m$ for a specific value m, and so the "document" in our model would a tax receipt that shows $\tau \geq m$. But it may be known that there always exists a tax receipt, which implies that if the document showing $\tau \geq m$ does not exist, then there exists a document showing $\tau < m$. By requiring the litigant to disclose a tax receipt, the court would be making disclosure of the "$\tau \geq m$" document mandatory.

particular greater than 1. If d is positive evidence of the high state (the case in which effective evidence may be achieved with coordinated beliefs), then the threshold on L could be increased to the extent that society can tolerate the risk of misleading evidence or the judge predicts that beliefs and behavior will be well coordinated. This reasoning does not extend to the setting in which d is positive evidence of the low state, because there evidence is ineffective with coordinated beliefs.

The admissibility test can be restated in terms of the standard notion of *relevance*, defined as the strength of the document's face-value signal. Consider the setting in which e and x are conditionally independent. Here, a change in parameter values that causes q_1/q_0 to move closer to 1 represents a decrease in relevance. Evidence may be misleading if $\underline{\gamma}(\psi) \leq q_1/q_0 \leq \overline{\gamma}(\psi)$, which is precisely when the evidence is mildly relevant in relation to the litigant's private signal and ψ.

Policy Implication 2: *To ensure robust use of evidence in litigation, the court must make hard evidence inadmissible when it is mildly relevant in relation to the litigant's private, unverifiable information. The court may relax the margin for admissibility if evidence would tend to favor the litigant's interests and if there is reason to believe that the litigant's and jury's beliefs will be in concert.*

5 A Setting with Multiple Documents

Extending the model to include multiple documents adds another layer of possible inference, as disclosure of one document or a set of documents could be interpreted as providing information about the existence or nonexistence of documents that were not disclosed. In this section we consider a setting with two documents and we explore the implications of restricting the litigant's disclosure options. We show that, for some parameter values, requiring the litigant to bundle documents leads to robust litigation whereas litigation is not robust without this constraint.

Consider an extension of our model with documents d_1 and d_2. There are four evidentiary states, which we represent in terms of subsets of documents: $E = \{\emptyset, \{d_1\}, \{d_2\}, \{d_1, d_2\}\}$, where $e = \emptyset$ means no documents exist, $e = \{d_k\}$ means only document d_k exists ($k = 1$ or $k = 2$), and $e = \{d_1, d_2\}$ means documents d_1 and d_2 both exist. The litigant is able to disclose only documents that exist; for instance, in the event that $e = \{d_1\}$, the litigant can only disclose d_1 or disclose nothing. Thus, in evidentiary state e, the litigant can select any $e' \subset e$ to disclose. In some expressions, instead of writing "$e = \{d_k\}$" and "$e = \{d_1, d_2\}$," we will write "$e = d_k$" and "$e = d_1 d_2$" for convenience.

Let us assume, for simplicity, that existence of the documents and the realization of the litigant's private signal x are all conditionally independent given the state θ. Further assume that $\psi = 0$. Let $p_\theta(x)$ be the probability that the private signal is x given the state θ, let $q_{1\theta}$ be the probability that d_1 exists given θ, and let $q_{2\theta}$ be the probability that d_2 exists given θ. Then, for instance, we have $f(d_1, x \,|\, \theta) = q_{1\theta}(1 - q_{2\theta})p_1(x)$ and $f(d_1 d_2, x \,|\, \theta) = q_{1\theta} q_{2\theta} p_1(x)$. We shall focus on the setting in which the two documents are positive evidence of the high state, so throughout this section we maintain the assumption that $q_{11} > q_{10}$ and $q_{21} > q_{20}$.

Let $\lambda_{e'}(x, e)$ be the probability that the jury assigns to the litigant disclosing $e' \subset e$ in the event that the evidentiary state is e and the litigant's private signal is x. The jury's posterior belief about the state conditional on disclosure of e', which we denote $b(e')$, is given by:

$$b(e') = \frac{\displaystyle\sum_{x \in X,\ e \supset e'} r f(e, x \,|\, 1) \lambda_{e'}(x, e)}{\displaystyle\sum_{x \in X,\ e \supset e'} r f(e, x \,|\, 1) \lambda_{e'}(x, e) + \sum_{x \in X,\ e \supset e'} (1 - r) f(e, x \,|\, 0) \lambda_{e'}(x, e)}. \tag{3}$$

Here λ is the jury's initial belief about the litigant's strategy or, in the case that this belief makes the denominator zero, λ is an arbitrary updated belief (as plain consistency requires).

As before, let σ denote the litigant's strategy. For each evidentiary state $e \in E$, every private signal $x \in X$, and each $e' \subset e$, we define $\sigma_{e'}(x, e)$ to be the probability that the litigant discloses e' when he possesses evidence e and receives private signal x. We use the term "full disclosure" to describe the litigant's strategy that always discloses every existing document. Let $b^\sigma(e')$ be the posterior probability from Equation 3 when we set $\lambda = \sigma$, where we recognize that this is defined only if $\sigma_{e'}(x, e) > 0$ for some x and e satisfying $f(\Theta, e, x) > 0$. The notions of misleading hard evidence and effective hard evidence carry over from the basic model. For instance, hard evidence is misleading if there exists a disclosure $e' \in E$ such that $b^\sigma(e')$ is well defined and yet $b(e') \neq b^\sigma(e')$.

We generalize the definitions of potential and actual loss ratios here, to allow various comparisons with regard to restrictions on the submission of evidence. Let δ denote a disclosure policy, which specifies the sets of documents that are allowed to be disclosed, with the empty set assumed to be included. For instance, one possible policy is $\delta^{\text{Bundle}} =$

$\{\emptyset, \{d_1, d_2\}\}$, which means that the litigant is allowed to disclose only nothing ($\emptyset$) or both documents ($d_1 d_2$). A disclosure policy δ implies a restricted game, in which the litigant's actions are limited to the options in δ.

Let $U_{\mathrm{J}}^*(\delta)$ denote the jury's expected payoff in an artificial setting in which the jury directly observes the maximal allowed disclosure of available documents (but does not observe x), forms the proper posterior belief, and best responds.[14] Let $\underline{U}_{\mathrm{J}}(\delta)$ denote the lowest expected actual payoff of the jury in a rationalizable outcome of the litigation game in which the litigant is restricted to disclosures in the set δ.

We can compare two disclosure policies, δ and δ', where δ' is a subset of δ. Note that $U_{\mathrm{J}}^*(\delta) - U_{\mathrm{J}}^*(\delta')$ is the *face value of the additional hard evidence allowed in* δ, which is nonnegative because policy δ allows more disclosure sets. The loss possible under δ is $U_{\mathrm{J}}^*(\delta) - \underline{U}_{\mathrm{J}}(\delta)$ and the loss possible under δ' is $U_{\mathrm{J}}^*(\delta') - \underline{U}_{\mathrm{J}}(\delta')$, so the difference is the potential *additional welfare loss* due to the allowance of more disclosure sets under δ. The *potential loss ratio* is defined as

$$L(\delta, \delta') = \frac{U_{\mathrm{J}}^*(\delta) - \underline{U}_{\mathrm{J}}(\delta) - [U_{\mathrm{J}}^*(\delta') - \underline{U}_{\mathrm{J}}(\delta')]}{U_{\mathrm{J}}^*(\delta) - U_{\mathrm{J}}^*(\delta')}.$$

When applied to the basic, single-document model with $\delta = \{\emptyset, \{d\}\}$ and $\delta' = \{\emptyset\}$, this measure is equivalent to the loss ratio L defined earlier.

Conditions for Misleading Hard Evidence

The following results provide conditions for effective and misleading hard evidence. First, we have conditions guaranteeing full disclosure, so that all hard evidence is effective. Let $\delta^{\mathrm{Full}} = \{\emptyset, \{d_1\}, \{d_2\}, \{d_1 d_2\}\}$ and let $\delta^0 = \{\emptyset\}$. The definition of $\overline{\gamma}$ is unchanged from the basic model with $\psi = 0$; that is, $\overline{\gamma}(0) = \max_{x \in X} p_0(x)/p_1(x)$.

Theorem 4: *Suppose that there are no limits on disclosure (the δ^{Full} disclosure policy) and*

$$\frac{q_{11}(1 - q_{21})}{q_{10}(1 - q_{20})} > \overline{\gamma} \quad \textit{and} \quad \frac{q_{21}(1 - q_{11})}{q_{20}(1 - q_{10})} > \overline{\gamma}.$$

[14]For the policy $\delta = \{\emptyset, \{d_1\}, \{d_2\}\}$, the maximal disclosure is not well defined in the event that $e = d_1 d_2$. We will not consider this policy here. In a general model, one might make the assumption that all policies under consideration are closed with respect to unions, so that if d_1 and d_2 can be disclosed individually, then disclosure of $d_1 d_2$ is also allowed. An interesting question is whether such a restriction has merit on efficiency grounds.

Then the only rationalizable outcome entails full disclosure, the actual loss ratio comparing δ^{Full} *to* δ^0 *is* $L(\delta^{\text{Full}}, \delta^0) = 0$*, and hard evidence is effective.*

Note that the condition of Theorem 4, which can also be written as $f(d_k, x\,|\,1) > f(d_k, x\,|\,0)$ for all $x \in X$ and $k = 1, 2$, means that existence of exactly one document is sufficiently strong positive evidence to outweigh any negative inference from the private signal x. It also implies that existence of both documents is likewise strong. The proof of this theorem requires several rounds of the iterated-elimination procedure.

We next show that policy δ^{Bundle}, which makes each document inadmissible on its own but allows the two documents to be admitted as a package, has advantages if the face-value signal provided by the documents individually is not so strong as to satisfy the presumption of Theorem 4. We do this in two steps, first comparing δ^{Bundle} to δ^0 and then comparing δ^{Full} to δ^{Bundle}. Note that

$$\frac{q_{11}q_{21}}{q_{10}q_{20}} > \frac{q_{11}(1-q_{21})}{q_{10}(1-q_{20})} \quad \text{and} \quad \frac{q_{11}q_{21}}{q_{10}q_{20}} > \frac{q_{21}(1-q_{11})}{q_{20}(1-q_{10})}.$$

We shall focus on the range of parameter values such that

$$\frac{q_{11}q_{21}}{q_{10}q_{20}} > \overline{\gamma} > \frac{q_{11}(1-q_{21})}{q_{10}(1-q_{20})} \quad \text{and} \quad \frac{q_{11}q_{21}}{q_{10}q_{20}} > \overline{\gamma} > \frac{q_{21}(1-q_{11})}{q_{20}(1-q_{10})}. \tag{4}$$

Theorem 5: *Suppose that*

$$\frac{q_{11}q_{21}}{q_{10}q_{20}} > \overline{\gamma}.$$

In every rationalizable strategy profile of the litigation game in which the litigant is restricted to disclosures in the set δ^{Bundle}*, the litigant always discloses both documents when they both exist. Hard evidence is effective and the actual loss ratio for* δ^{Bundle} *compared to the no-evidence policy* δ^0 *is* $L(\delta^{\text{Bundle}}, \delta^0) = 0$.

Theorem 6: *Suppose that Inequalities 4 hold. In the setting with no limits on disclosure (the* δ^{Full} *disclosure policy), there are rationalizable outcomes in which disclosure of a single document is misleading but disclosure of both documents is effective, and there are also rationalizable outcomes in which all hard evidence is effective. The potential loss ratio comparing* δ^{Bundle} *to* δ^{Full} *is* $L(\delta^{\text{Full}}, \delta^{\text{Bundle}}) > 1$.

For intuition, one can see that there are multiple perfect Bayesian equilibria of the game under policy δ^{Full}. There are an infinite number of equilibria in which the litigant discloses both documents when they both exist and otherwise discloses nothing. In such an equilibrium, the jury's beliefs satisfy $b(d_1), b(d_2) < r < b(d_1 d_2)$ because the jury associates disclosure of one document with an adverse (for the litigant) realization of x. There is also an equilibrium featuring full disclosure and $r < b(d_1), b(d_2) < b(d_1 d_2)$. If, for instance, a litigant with a single document believes that the latter equilibrium is being played, whereas the jury believes that the litigant behaves according to the former equilibrium, then disclosure of a single document is misleading to society's detriment (lowering social welfare compared to the outcome under policy δ^{Bundle}).

Implications for Courts

Analysis of the two-document model allows us to expand on the policy implications developed in the basic model by examining a wider range of admissibility rules. Theorems 5 and 6 yield the following general conclusion.

Policy Implication 3: *To ensure robust use of evidence in litigation, the court sometimes must require that multiple documents be admissible only as a bundle (not individually).*

This conclusion brings to mind a common occurrence in the court: When an attorney attempts to present marginally relevant evidence, the judge asks where the presentation is headed and what the attorney intends to establish, and the judge allows the evidence under the expectation of additional complementary evidence to follow. Cleary (1972) notes that the judge has the discretion to "ask the proponent what additional circumstances he expects to prove."

6 Legal Rules

In this section, we bolster the connection between our model and legal policy by commenting on how U.S. law defines "relevance" and how it justifies exclusions according to FRE Rules 403 and 404. We include a brief account of a prominent legal case associated with Rule 403 and we present a numerical example based on this case.

Modern reform of the law of evidence relied heavily on the ideas of James Bradley Thayer, collected in Thayer's *A Preliminary Treatise on Evidence at the Common Law* (1898); see Anderson, Schum, and Twining (2005) for a recent analysis. Thayer viewed evidence law in the typical trial process as focusing on the following four issues: 1) materiality, meaning the facts to be proved; 2) relevance; 3) admissibility; and 4) the evaluation of weight or probative force of evidence. The first three are matters for the judge. The fourth is for the jury or other fact-finder.

Definitions of the legal terms *probative* (providing proof regarding a claim) and *relevance* (tending to strengthen the particular claim being assessed) indicate that the interpretation of evidence is often a matter of establishing a degree of confidence rather than reaching a conclusion with certainty.[15] FRE Rule 401 states: "Evidence is relevant if (a) it has any tendency to make a fact more or less probable than it would be without the evidence; and (b) the fact is of consequence in determining the action." So clearly the law embraces the notion that evidence is statistical, as in our model, and relevance is the determination of whether the evidence provides a signal of the claim being evaluated.

Because juries are responsible for weighing the evidence, with judges providing only guidance, the potential for evidence to be misleading is a real concern for courts, and we assert that robustness is a natural objective. Thayer described the analysis of evidence as being governed by "logic and experience;" Anderson, Schum, and Twining (2005) suggest that "an alternative interpretation is that the criteria for the weight of evidence are provided by probability theory, of which there are many versions."

FRE Rules 403 and 404

The idea that exclusion of evidence can improve fact-finding is unique to common law systems (see Damaska 1997). Thayer argued for judicial discretion regarding admissibility. He suggested that the judge should have discretion to exclude evidence that is "slightly" or "remotely" relevant, and may "complicate the case" or "confuse, mislead, or tire the minds" of the jury. This principle is refined in FRE Rule 403, quoted at the beginning of this article, which is sometimes referred to as the "prejudice rule." Additionally, the notes of the Advisory Committee for Rule 403 state that "The case law recognizes that certain circumstances call for the exclusion of evidence which is of unquestioned relevance." Swift (2000) describes Rule 403 as "the primary example of guided discretion in modem evidence law."

[15] This is not to say that near certainty is never achieved. For example, murder charges against a man named Juan Catalan were eventually dismissed following submission of a video showing him at a Los Angeles Dodgers game at exactly the time of the crime (a time pinpointed by witnesses). For details of the case, see Rubin and Bloomkatz (2008).

Our modeling exercise provides a rationale for Thayer's recommendation and Rule 403, which both focus on the potential for—not certainty of—misleading the jury.[16] Our results provide some guidance for striking a balance between relevance and the potential for misleading the jury. One can view a potential loss ratio L that exceeds 1 as corresponding to the "substantially outweighed by a danger of" condition of Rule 403.[17] Theorems 1-2 and Policy Implications 1-2 indicate the conditions under which judges should consider excluding hard evidence. We note also that the law acknowledges the value of incremental evidence—which we can call *marginally relevant*—especially if several such pieces can be combined to make a stronger signal (for example, see the United States v. Pugliese, 2d 1946). Thus the law recognizes the trade-offs evaluated by our Theorem 6 and Policy Implication 3.

We assert that "unfair prejudice," "confusing the issues," and "misleading the jury" all may arise as manifestations of what in our model is called "misleading evidence," because they are problems that can occur when updating beliefs. They also have some nuanced differences. For instance, the Advisory Committee notes on 403 states that "'Unfair prejudice' within its context means an undue tendency to suggest decision on an improper basis, commonly, though not necessarily, an emotional one." Dolan (1976) notes that courts do not often make distinctions between prejudice and confusing the issues, and offers the following guidance: "Confusion of the issues is not always the same as prejudice. The offered evidence may be noninflammatory, encouraging not so much an irrational or emotional result as simply an incorrect one." In discussing misleading the jury, Dolan writes: "Generally the problem is with evidence that will, in the court's view, be given too much weight by the jury, although neither prejudicial nor involving ancillary issues."

Rule 404, also stated in the Introduction, documents courts' general reluctance to allow character evidence or evidence of a prior conviction about a defendant in a criminal case. Essentially Rule 404 isolates a special circumstance that, due to its commonality, warrants its own rule. Although there are some nuances related to a criminal defendant who chooses to testify at his own trial, overall Rule 404 states that, in the language of our model, the

[16] Interestingly, it seems, in practice, misleading evidence is viewed as less of an issue for bench trials. See Capra (2000), which cites Gulf States Utilities Company v. Ecodyne Corporation, 635 F.2d 517 (5th Cir. 1981). There the Fifth Circuit described exclusion for unfair prejudice in a bench trial as "a useless procedure."

[17] In our model, there can be rationalizable outcomes of the litigation game that give the jury an expected actual payoff in excess of U_{J}^{*}, which means that ineffective evidence provides a welfare gain, but we presume that the law would not bank on such an outcome.

face-value signal of character evidence is so outweighed by the potential litigant-type signal as to make this kind of evidence unreliable for robust litigation. Appendix B picks up the discussion with a numerical example.

There are other avenues for exclusion in the common law, some that the model here may not be equipped to rationalize. In the U.S., evidence obtained by an illegal search by police is excluded. In addition to differences in some exclusion laws across countries, there are interesting differences across states. In California, the "Pitchess law" provides many restrictions on defendants' access to records of a police officer's previous discipline for activities such as evidence tampering. See Rubin (2018) for a discussion of these laws.

Numerical Illustration of a Case

Consider the classic case of Robitaille v. Netoco Community Theatre Inc., 305 Mass. 265. Robitaille was injured after falling on a stairway at the Netoco Community Theatre and sued for damages. There was a thick carpet on the stairway and a critical issue in the case was whether the carpet was loose, which could have occurred due to the tacks holding it in place having come out. A few weeks before the accident, two girls fell on the stairway when the carpet was loose. At trial, Robitaille was allowed to present evidence that the two girls fell although there was not evidence that the condition of the stairway at the time of Robitaille's fall was the same as at the time of the earlier fall. Robitaille did not present evidence that spoke to whether the carpet was loose at the time of her accident. On appeal, it was ruled that evidence of the prior accident with loose carpet should not have been allowed.

Based on this case, we construct a stylized example in the context of the events that potentially lead to the relevant likelihood ratios. The following time line describes the sequence of events.

1. The following exogenous random draws occur at the time of a possible prior accident:
   - The rug is loose (ℓ) with probability 1/2 and not loose (n) with probability 1/2.
   - The "prior accident" (a) occurs with probability 1/2 if the carpet is loose and with probability 1/8 if the carpet is not loose.
2. If the carpet was loose and there was a prior accident, the theater owner becomes aware of the carpet being loose. Otherwise, the owner is not aware of the carpet's condition and has no action to take.

When aware of the loose carpet, a "good" owner, occurring with probability 3/4, repairs the loose carpet. A "bad" owner does not repair the loose carpet.[18] Following this stage, we describe the carpet as being loose by L and not loose by N, where L is the case if and only if the rug was previously loose (ℓ) and either no prior accident occurred or it occurred and the owner is bad (so the carpet is not repaired).

3. A random patron visits the theatre. If L, an accident occurs with probability 1/2. If N, an accident occurs with probability 1/8.

4. If an accident occurred at Date 3, a lawsuit is initiated and the patron (the litigant in our model) receives a private signal $x \in \{\underline{x}, \overline{x}\} = X$ about the condition of the carpet. This private signal represents the patron's noisy observation of the carpet just before or after the accident. Assume that conditional on N, $\overline{x}$ occurs with probability 1/4 and $\underline{x}$ occurs with probability 3/4. Conditional on L, $\overline{x}$ occurs with probability w and $\underline{x}$ occurs with probability $1 - w$.

   Additionally, the litigant possesses document d if and only if there was a prior accident that occurred *with* loose carpet (that is, when a and ℓ occur).

Our model picks up this story at Date 4, conditional on the accident occurring. The state θ refers to whether the owner is liable, in which case the jury and society would like to reach a judgment of liability. Liability ($\theta = 1$) is the event in which, conditional on the accident occurring, the carpet is loose (L), the prior accident occurred with loose carpet, and the owner did not repair the carpet. That is, conditional on the accident occurring, a, ℓ, and L together imply $\theta = 1$. Non-liability ($\theta = 0$) is the complement event, conditional on the accident occurring. Note that we assume that in order for the case to go to trial, there must have been an accident.[19]

By constructing an event tree and calculating the probabilities of the various paths, we obtain the following conditional probabilities:

$$f(d, \overline{x} \mid 1) = w, \;\; f(d, \overline{x} \mid 0) = \frac{1}{36}, \;\; f(d, \underline{x} \mid 1) = 1 - w, \;\; \text{and} \;\; f(d, \underline{x} \mid 0) = \frac{3}{36}.$$

[18] The assumption that the owner learns of the loose carpet only when there is an accident is for simplicity. A richer example that has the good owner more likely to become aware of the carpet being loose has similar qualitative results. One can motivate the two types of owner on the basis of the cost of repairing the carpet, with the good owner having a small cost and the bad owner facing a large cost. Also, if the "good" owner is more likely to become aware of the loose carpet, the implications do not change.

[19] This does not change the qualitative results for the example. A compelling motivation for this is that if there were no accident then the case would likely be dismissed. Of course, in practice not all frivolous lawsuits are dismissed. In that direction, we also note that, given our simplifying assumptions for this example, when the litigant does not possess d, she knows that $\theta = 0$, but it's also possible that the litigant possesses d when $\theta = 0$. So there is concern about evidence potentially being misleading here.

See Appendix B for some of the calculations. Note there are two possibilities for evidence to be misleading here. Naturally, this depends on how informative the litigant's private signal x is, which depends on the value of w.[20]

For simplicity, assume $\psi = 0$. Then the inequality in Condition D reduces to $f(d, K^- \,|\, 1) \leq f(d, K^- \,|\, 0)$ and the inequality in Condition U reduces to $f(d, K^+ \,|\, 1) \geq f(d, K^+ \,|\, 0)$. By Theorem 1, there is scope for evidence to be misleading when Conditions D and U are both satisfied. Because there are only two x-types of litigant in this example, the key question is whether either $f(d, \underline{x} \,|\, 1) \leq f(d, \underline{x} \,|\, 0)$ and $f(d, \overline{x} \,|\, 1) \geq f(d, \overline{x} \,|\, 0)$ or $f(d, \overline{x} \,|\, 1) \leq f(d, \overline{x} \,|\, 0)$ and $f(d, \underline{x} \,|\, 1) \geq f(d, \underline{x} \,|\, 0)$.

Consider first the case in which $f(d, \underline{x} \,|\, 1) = 1 - w \leq f(d, \underline{x} \,|\, 0) = 3/36$ and $f(d, \overline{x} \,|\, 1) = w \geq f(d, \overline{x} \,|\, 0) = 1/36$. That is, $K^- = \{\underline{x}\}$ and $K^+ = \{\overline{x}\}$. This requires $w \geq 33/36$. So for large values of w, meaning that the patron has an accurate private signal of whether the carpet was loose at the time of the litigant's accident and $\overline{x}$ suggests it was loose, the conditions for evidence to be misleading are satisfied. Next consider the case in which $f(d, \overline{x} \,|\, 1) = w \leq f(d, \overline{x} \,|\, 0) = 1/36$ and $f(d, \underline{x} \,|\, 1) = 1 - w \geq f(d, \underline{x} \,|\, 0) = 3/36$ so that $K^- = \{\overline{x}\}$ and $K^+ = \{\underline{x}\}$. This requires $w \leq 1/36$, implying the litigant has accurate private information about the carpet. Thus, for extreme values of w, where the litigant's private signal is informative relative to the hard evidence, there is a rationalizable outcome in which evidence d is misleading.

7 Related Literature

We first discuss models of evidence in the law-and-economics and the greater economics literature, and we then discuss work aimed at explaining exclusion. On the former, two main approaches to modeling evidence stand out. The first treats evidence as statistical in nature but views evidence as arriving exogenously. These models utilize Bayes' rule updating but they address neither the parties' incentives to disclose evidence nor the fact-finder's evaluation of these incentives. The second approach focuses on litigants' incentives to produce evidence but treats the adjudicator as a mechanistic system whose judgment is an exogenous function of the quantities of evidence that the two sides produce. Evidence

[20]Note that since the probabilities of $\underline{x}$ and $\overline{x}$ are fixed following N, the informativeness of the private signal only depends on w. Because of this simplification, the conditions under which evidence can be misleading below may be more extreme than in practice.

production is costly, and each party's marginal cost is higher in the state of the world that favors the other party. These models treat evidence as an abstract quantity and they conclude that the litigant types will be separated in equilibrium (although the resulting signal is not utilized by the adjudicator).[21]

In the latter set of articles, Daughety and Reinganum (2000a,b) are perhaps the most related to our modeling exercise. They develop an axiomatic approach with a strategic-search model of evidence. The second of these articles incorporates a non-Bayesian assessment of credible evidence by the trial court, suggesting that "Whether one models this as 'mostly Bayesian with a few constraints' or 'mostly constrained with a few opportunities for Bayesian updating' is a judgment call." Our exercise, in relation, is fully Bayesian but without an equilibrium assumption.

Other prominent entries in the literature feature both the litigants' incentive to disclose evidence and a Bayesian decision maker, but they assume that hard evidence provides definitive proof of the state or some subset of possible states.[22] There are also mechanism-design models that seek to find the optimal judgment rule (a mechanism that maps feasible evidence sets to judgments) under the assumption that the litigants will find their way to an equilibrium in the induced evidence-production game.[23]

In the direction of multiple documents, Fluet and Lanzi (2018) and Guttman, Kremer, and Skrzypacz (2014) study settings in which a sender may have more than one signal that can be verifiably disclosed, so there are multiple channels of information. Both assume equilibrium, and they focus on different issues than we address. Guttman, Kremer, and Skrzypacz analyze a setting of financial disclosure where the manager of a firm makes disclosure decisions over two periods. Fluet and Lanzi explore cross-examination. In their model,

[21] We have here summarized Talley's (2013) characterization of the law-and-economics literature on evidence. See Talley's chapter, and also Sanchirico (2010), for citations.

[22] Milgrom (1981) and Shin (1994) are examples. Che and Severinov (2017) focus on the role for attorneys to suppress evidence in a model in which litigants probabilistically possess evidence and an additional judgment-relevant piece of information is observed by attorneys and the court. Both of these are from a continuum and satisfy a monotone likelihood ratio property. They find that attorneys are helpful in that they can suppress favorable evidence in equilibria with play of weakly-dominated strategies.

[23] Bull and Watson (2004, 2007), Green and Laffont (1986), and Kartik and Tercieux (2012) are examples in this category. Bull (2012) studies a model in which a piece of evidence can exist both when an accused is guilty and when he is innocent, but this focuses on the different issues of police interrogation and incentives for evidence fabrication. Tangentially related "Bayesian-persuasion" models involve a sender committing to an informative experiment to influence a receiver; Watson (1996) is an early entry here. Those models typically assume that the sender and receiver have shared prior information. Hedlund (2017) studies Bayesian persuasion with a privately-informed sender.

each litigant invests in gathering evidence either to directly support the litigant's own case or to counter the other side's argument in cross-examination. Incorporating this sort of costly evidence acquisition into our modeling approach may make for some interesting future work.

Turning to possible rationales for excluding hard evidence, Lester, Persico, and Visschers (2009) provide an explanation that focuses on cognitive limitations of the jury. In their model, the jury finds it costly to evaluate evidence and thus faces trade-offs in selecting lines of evidence to invest time and effort to assess. The jury's incentives may not be aligned perfectly with society's incentives, in which case the judge, by barring some kinds of evidence, may be able to distort the jury's choices in a direction that benefits society. These authors speak of their analysis as providing "several results pointing to the difficulty of eliciting general principles that can inform the exclusion of specific pieces of evidence as a general rule." Our approach proposes a different mechanism that leads to socially undesirable outcomes (based on misleading evidence) and, in comparison, provides a modest foundation for some general principles. Other examples of a cognitive-limitation approach include Gold (1986) and Langevoort (1998).[24]

Two other related articles that develop theory on exclusion of character evidence are Sanchirico (2001) and Schrag and Scotchmer (1994). These articles suggests that a potential wrongdoer's choice of action does not influence character evidence, and so character evidence should not be used to provide incentives.[25] Lester, Persico, and Visschers (2009) note that the argument may not fit with Rule 403, which gives the judge a great degree of latitude "to exclude evidence on a case-by-case basis," and this seems at odds with the incentive-providing argument.

8 Conclusion

In this article, we have developed a methodology for analyzing statistical hard evidence, particularly in distinguishing between the face-value signal and the litigant-type signal. We

[24]Cognitive limitations also underlie the "story view" of information processing, which postulates that fact-finders are best able to process evidence that is woven into a coherent and simple story—especially if it has a linear logical progression and is supported by analogous experience—and then compare alternative stories using available data. See Pennington and Hastie (1986, 1991, 1992) and Hastie (1999).

[25]Sanchirico's main example is a bar patron who is deciding whether to assault another patron who is annoying him. The argument is that if the trial decision focuses on trace evidence of whether assault occurred, even a defendant who has a history of assault in bars would have appropriate incentives.

have argued for application of the rationalizability concept to study robustness of reasoning about hard evidence and the decisions of litigants and fact-finders. Our results provide conditions under which hard evidence may be misleading, and they also provide guidance for balancing relevance against the potential to mislead the jury. Overall, we advocate for policy makers in the law and elsewhere to consider how the interpretation of hard evidence relates to the incentives to disclose it.

Our modeling exercise is based on the position that it is practically impossible for the law to commit courts to an optimal mechanism that dictates how evidence is interpreted. There are too many idiosyncrasies in individual cases for overarching rules to be useful, and the law would not be able to describe exactly what the interpretation should be for every specific case. In other words, fact-finders such as juries are in the business of processing information and interpreting evidence, and this is an essential exercise in the pursuit of society's objectives. Thus, fact-finders are ideally Bayesian and the legal system must recognize this, but the law (and courts) may optimally put some restrictions on how they interpret evidence and make judgments. Realistic instruments in this regard are admissibility rules and associated standards of relevance, suitably described in statistical terms. It is the fact-finder's job to evaluate the idiosyncrasies of individual cases, weigh the evidence, and conduct a Bayesian analysis.

Regarding promising lines for future research, it will be useful to explore more deeply the incentives of players at the "primary activity" stage (before litigation). In the present model, the prospect of misleading evidence disappears if an institution coordinates the players on an equilibrium. Adding primary activity to the model, and in particular actions that are payoff relevant and affect the availability of evidence (such as effort to get on the surveillance video in the example in the Introduction), yields a rich environment in which it is optimal to exclude certain kinds of hard evidence even in equilibrium without it being misleading. Insights may be generated about incentives in primary activity and the cost of production or suppression of evidence.[26] For instance, depending on the anticipated interpretation, a potential litigant may choose to take actions that influence the probability with which hard evidence will exist.

One could also enrich the current model to allow the litigant's payoff to depend on the

[26]Bull (2009) studies legal institutions when there is scope for suppressing evidence, which is modeled as reducing the probability evidence is available.

litigant's type, whether a document exists, and whether the litigant discloses the document. There could, for example, be a cost of disclosure or a cost of withholding evidence. If costs such as these are small in relation to the other parameters of the model, then our results will still hold in a modified form. Likewise, very large costs in some state and for some litigant type would be tantamount to nonexistence of the document in this contingency. Moderate costs, whether or not depending on the litigant's type or the state, would introduce a new element in the model, as types may separate on the basis of a cost differential.

Another avenue for future research is to model the jury as a group of individual strategic actors, and to explore the scope for miscoordination and disagreement among them on the interpretation and meaning of evidence. This is a critical issue given that jurors come from varied backgrounds and typically have not interacted with each other before, and few have much experience as fact-finders. Further, it would be interesting to pursue applications of our model outside the legal realm, such as in finance and marketing.

Finally, let us comment on a setting with two strategic litigants. The modeling exercise here was meant to demonstrate the problem of misleading evidence in the interaction between a single litigant and fact-finder, honing in the coordination problems in the interpretation of hard evidence. Our analysis trivially extends to the two-litigant setting if one holds fixed the behavior of the previously unmodeled opposing litigant. Further, we expect that the problem of miscoordinated beliefs and behavior would be magnified by the presence of an additional strategic agent. For instance, the defendant and plaintiff may have different beliefs about how the jury will respond to bundles of hard evidence, and an individual litigant may incorrectly predict how the other will respond to this litigant's initial disclosure of hard evidence. Two-sided private information would further complicate the matter. We hesitate to make any claims about insights that a modeling exercise with two strategic litigants can generate, except to say that this seems to be a fruitful direction for future work.

Additionally, we expect that analysis of the kind introduced here may be useful in exploring the justification for, and implications of, the Brady Rule. Lempert, Gross, and Liebman (2000) describe Brady v. Maryland (1963) and United States v. Bagley (1985) as establishing that "prosecutors have a constitutional duty to disclose exculpatory evidence—including evidence that could be used to impeach government witnesses—if there is a reasonable probability that disclosure would produce a different outcome." The "reasonable probability" requirement provides latitude for prosecutors to withhold exculpatory evidence. In terms of our model, for evidence possessed by the prosecutor that has a strong face-value signal in favor of the defendant, this amounts to increasing ψ. However, for other such evidence, the analysis would be more complicated because there are several layers of interpretation at issue, including by the prosecutor, an appeals court, the defense, and the jury. Holding aside the Brady Rule, even if both litigants had access to exactly the same evidence (as in Bull and Watson's (2004) *enforced discovery requests* setting), there is still the issue of interpretation. An additional element, but not the main reason for the interpretation problem, is that the jury observes by whom evidence is disclosed.

A Proofs

This appendix contains proofs of the theorems and supporting analysis. We start with details about how plain consistency (Watson 2017) implies the Bayes'-rule expression even in the case in which the jury's initial belief puts zero probability on the document being disclosed.

We claimed in the text that the jury's posterior belief conditional on disclosure is characterized by Bayes' rule (Equation 1) for the jury's initial belief λ or, if Bayes' rule does not apply for this belief, then for the jury's updated belief about the litigant's strategy. The latter condition arises in the setting of $\psi = 0$. Here is the justification, which is based on the assumption that the jury's belief system satisfies plain consistency.

Theorem 7: *The following holds for every belief system satisfying plain consistency. The jury's posterior belief $b(\emptyset)$ satisfies Bayes' rule, using the jury's initial belief λ about the litigant's strategy. The jury's posterior belief $b(d)$ satisfies Equation 1, where λ is the jury's initial belief about the litigant's strategy if it satisfies $\sum_{x \in X} f(\Theta, d, x)\lambda(x) > 0$ and otherwise λ is an arbitrary (updated) belief about the litigant's strategy.*[27]

Proof: The proof features similar steps as in Watson (2017), Section 4. First, that $b(\emptyset)$ satisfies Bayes' rule is due to that the prior probability that the jury puts on the document not being disclosed is strictly positive, and that plain consistency implies proper conditional-probability updating when possible.

We must also show that $b(d)$ satisfies Equation 1. As before, if $\sum_{x \in X} f(\Theta, d, x)\, \lambda(x) > 0$, then plain consistency implies that the jury uses proper conditional-probability updating in order to arrive at Equation 1 upon observing document disclosure. We will now consider the case when $\sum_{x \in X} f(\Theta, d, x)\, \lambda(x) = 0$.

Note that Nature's strategy in the litigation game is a selection (θ, e, x), and Nature mixes according to the distribution f. Let S_{J} denote the jury's strategy space and let S_{L} denote the litigant's strategy space. For every $x \in X$, let S_{L}^x be the subset of S_{L} that specifies disclosing d in the event that x is the private signal and the document exists. Observe that

[27] This means that the beliefs are structurally consistent (Kreps and Ramey 1987). Note that it must be that $\lambda(x) > 0$ for some x; otherwise, Equation 1 could not hold.

the sets $\{S_{\mathrm{L}}^x\}$ are distinct and not disjoint; for instance, for any given x and x', the strategy that always discloses belongs to both S_{L}^x and $S_{\mathrm{L}}^{x'}$. Define:

$$\begin{aligned} W_{\mathrm{N}}(x) &\equiv \{(1,d,x)\} & Y_{\mathrm{N}}(x) &\equiv \{(\theta,d,x) \mid \theta \in \{0,1\}\} \\ Y_{-\mathrm{N}}(x) &\equiv S_{\mathrm{L}}^x \times S_{\mathrm{J}} & Z(x) &\equiv Y_{\mathrm{N}}(x) \times S_{\mathrm{L}} \times S_{\mathrm{J}}. \end{aligned}$$

Let $Y(x) \equiv Y_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x)$. Note that $\{Z(x)\}_{x\in X}$ is a disjoint collection of sets. Because $W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x) \subset Y(x) \subset Z(x)$, clearly $\{Y(x)\}_{x\in X}$ and $\{W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x)\}_{x\in X}$ are both disjoint collections of sets as well.

Let h_{J}^d denote the information set in which the jury has observed disclosure of the document and must select the judgment a. Let $S(h_{\mathrm{J}}^d)$ denote the strategy profiles that are consistent with the game reaching information set h_{J}^d. By the above definitions of the various sets, we have that $S(h_{\mathrm{J}}^d) = \cup_{x\in X} Y(x)$. Further, the subset in which $\theta = 1$ is given by $T \equiv \cup_{x\in X} W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x)$. Let p^h denote the belief at information set h by the player on the move there about the strategy profile being played; this is called an appraisal. We want an expression for $p^{h_{\mathrm{J}}^d}(T)$, which is the definition of $b(d)$, the probability that the jury puts on $\theta = 1$ in the event that the document is disclosed.

We can apply the plain consistency condition (see Watson 2017 for definitions) to get an expression for the jury's belief about θ at information set h_{J}^d. Specifically, we relate player 2's appraisal at h_{J}^d to her appraisal at $\underline{h}_{\mathrm{J}}$, the artificial information set for the jury that signifies the beginning of the game. The required conditions to apply plain consistency on each set $Z(x)$ are satisfied. Sets $Y(x)$ and $Z(x)$ are *comparable* relative to Nature's information set, $\underline{h}_{\mathrm{J}}$, and h_{J}^d. Further, $W_{\mathrm{N}}(x) \subset Y_{\mathrm{N}}(x)$. Thus, plain consistency requires the jury's appraisal at h_{J}^d to be the product of a distribution over Nature's strategy and distributions over the strategies of the litigant and jury. Further, the appraisal must preserve probability ratios in that

$$\frac{p^{h_{\mathrm{J}}^d}(W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x))}{p^{h_{\mathrm{J}}^d}(Y_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x))} = \frac{p^{\underline{h}_{\mathrm{J}}}(W_{\mathrm{N}}(x) \times S_{\mathrm{L}} \times S_{\mathrm{J}})}{p^{\underline{h}_{\mathrm{J}}}(Y_{\mathrm{N}}(x) \times S_{\mathrm{L}} \times S_{\mathrm{J}})}, \tag{5}$$

if both denominators are strictly positive.

Note that the terms on the right side of Equality 5 refer to the jury's belief at the beginning of the game about Nature's strategy and, since this belief is initially accurate because the jury knows the game being played, the right side is simply $f(1,d,x)/f(\Theta,d,x)$. The denominator $f(\Theta,d,x)$ is strictly positive for all $x \in X$ (recall that we assumed this in the description of the model). Let us write $\mu(x) \equiv p^{h_{\mathrm{J}}^d}(Y_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x))$ for each x. Multiplying both sides of Equation 5 by $\mu(x)$ and summing, we get

$$\sum_{x\in X} p^{h_{\mathrm{J}}^d}(W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x)) = \sum_{x\in X} \frac{f(1,d,x)}{f(\Theta,d,x)} \mu(x). \tag{6}$$

Recall that the collection of sets $\{W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x)\}_{x\in X}$ is disjoint and the union is T. Thus, $p^{h_{\mathrm{J}}^d}(T) = \sum_{x\in X} p^{h_{\mathrm{J}}^d}(W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x))$. Also, $p^{h_{\mathrm{J}}^d}(T)$ defines $b(d)$, so the left side of Equation 6 is simply $b(d)$. Because $\{Y(x)\}_{x\in X}$ is a disjoint collection and the union is $S(h_{\mathrm{J}}^d)$, we know that $\sum_{x\in X} \mu(x) = p^{h_{\mathrm{J}}^d}(S(h_{\mathrm{J}}^d)) = 1$. Defining $\pi(x) \equiv \mu(x)/f(\Theta,d,x)$ for all x, Equation 6 becomes

$$b(d) = \sum_{x \in X} f(1, d, x)\pi(x). \tag{7}$$

Because $\sum_{x \in X} \mu(x) = 1$, we have $\sum_{x \in X} f(\Theta, d, x)\pi(x) = 1$ and Equation 7 is equivalent to

$$b(d) = \frac{\sum\limits_{x \in X} f(1, d, x)\pi(x)}{\sum\limits_{x \in X} f(\Theta, d, x)\pi(x)}.$$

By construction, $\pi(x) \geq 0$ and $\sum_{x \in X} \pi(x) > 0$. For a small enough strictly positive number ϕ and letting $\lambda(x) \equiv \phi\pi(x)$, we have that $\lambda(x) \in [0, 1]$ for every $x \in X$. Furthermore, because ϕ cancels in the fraction, we have

$$b(d) = \frac{\sum\limits_{x \in X} f(1, d, x)\lambda(x)}{\sum\limits_{x \in X} f(\Theta, d, x)\lambda(x)},$$

which is exactly Equation 1 from the text and $\lambda(x)$ may be interpreted as a probability. $\square$

The rest of this appendix deals with the theorems in the main text. For each theorem, we restate the result and then provide a proof.

Theorem 1: *If Condition U fails, then the unique rationalizable outcome has the litigant disclosing d always at minimum probability ψ, the actual loss ratio is $L \in (0, 1]$, and hard evidence is ineffective. If Condition D fails, then the unique rationalizable outcome has the litigant disclosing d whenever he has it, the actual loss ratio is $L = 0$, and hard evidence is effective.*[28] *If Conditions D and U are both satisfied, then the potential loss ratio is $L > 1$ and there are rationalizable outcomes in which hard evidence is misleading.*

Proof: Let $(a(d), a(\emptyset))$ denote the jury's strategy, the first term being the jury's action in the event the document is disclosed and the second term being the action in the event the document is not disclosed. Remember that sequential rationality requires $a(d) = b(d)$ and $a(\emptyset) = b(\emptyset)$.

Suppose first that Condition U fails. Then from Lemma 1 and Theorem 7 the jury's posterior beliefs must satisfy $b(d) < r \leq b(\emptyset)$ regardless of the jury's initial belief about the litigant's strategy.[29] Therefore the jury's strategy must satisfy $a(d) < a(\emptyset)$. In other words, every strategy of the jury that specifies $a(d) \geq a(\emptyset)$ cannot be rationalized and so

[28] Although rationalizability assumes common knowledge of rational behavior, to prove the first two claims we need only that players are rational (best responding to their beliefs) and know this about each other.

[29] If the jury's initial belief λ implies strictly positive probability of both disclosure and non-disclosure, then $b(d)$ is calculated using this function λ. Letting ϕ be the probability that d is disclosed when the litigant behaves according to λ, we have $r = \phi b(d) + (1 - \phi)b(\emptyset)$ by the law of iterated expectations, and so $b(d) < r$ implies $b(\emptyset) > r$. If the jury's initial belief satisfies $\lambda(x) = 0$ for all x, then $b(\emptyset) = r$ because non-disclosure conveys no information.

is removed from consideration. Because the litigant knows this (rationalizability assumes common knowledge of sequential rationality), every x-type of litigant strictly prefers to not disclose the document, regardless of the exact belief about the jury's strategy. This means that there is a single rationalizable strategy for the litigant: $\sigma(x) = \psi$ for all $x \in X$. The jury knows this and therefore has the belief $\lambda(x) = \psi$ for all $x \in X$, and hence the jury accurately updates conditional on disclosure and conditional on no disclosure. There is only partial disclosure and so hard evidence provides less than its face-value signal, and we have $L \in (0, 1]$.

The same steps establish that if Condition D fails then the jury's strategy must satisfy $a(d) > a(\emptyset)$ and the litigant rationally must always disclose: $\sigma(x) = 1$ for all $x \in X$. The jury knows this and therefore has the belief $\lambda(x) = 1$ for all $x \in X$, and hence the jury accurately updates conditional on disclosure and no disclosure. There is full disclosure, hard evidence provides its full face-value signal, and $L = 0$.

Finally, consider the case in which Conditions D and U both hold. Define beliefs λ^+ and λ^- as in the proof of Lemma 1 in Section 3. Let $(b^+(d), b^+(\emptyset))$ denote the jury's posterior beliefs derived from initial belief λ^+ and let $(b^-(d), b^-(\emptyset))$ denote the jury's posterior beliefs derived from initial belief λ^-. Using the steps in the proof of Lemma 1 and the law of iterated expectations (as above), we see that $b^+(d) \geq b^+(\emptyset)$ and $b^-(d) \leq b^-(\emptyset)$. Therefore, if the jury's initial belief is λ^+ then the jury's sequentially rational strategy satisfies $a(d) \geq a(\emptyset)$ and if the jury's initial belief is λ^- then the jury's sequentially rational strategy satisfies $a(d) \leq a(\emptyset)$. Any type of litigant can rationalize disclosing with probability 1 in response to the former strategy and disclosing with the minimum probability ψ in response to the latter strategy. Thus, no strategies are removed from consideration in the rationalizability procedure.

From our assumption that $f(d, x \mid 1) \neq f(d, x \mid 0)$ for some x, we know that $f(d, K^- \mid 0) - f(d, K^- \mid 1) > f(d, K^+ \mid 0) - f(d, K^+ \mid 1)$, which means that at least one of the inequalities in Conditions D and U holds strictly. Using the construction in the proof of Lemma 1, this implies that either $b^-(d) < r$ or $b^+(d) > r$.

Take the case in which $b^-(d) < r$. Then we have $b^-(d) < r < b^-(\emptyset)$ by the law of iterated expectations. We can then find a rationalizable outcome that entails misleading hard evidence and a welfare loss greater than 1. Suppose the litigant's actual strategy is defined by $\sigma = \lambda^+$, so the optimal strategy of the jury (if the jury knew that σ was being played) would specify $a(d) = b^+(d)$ and $a(\emptyset) = b^+(\emptyset)$, where $b^+(d) \geq r \geq b^+(\emptyset)$. And any departure from these actions would strictly reduce the jury's expected payoff. Specifically, lowering $a(d)$ and/or raising $a(\emptyset)$ uniformly reduces the jury's actual expected payoff. Suppose that the jury's initial belief is actually λ^-, to which the jury rationally responds by selecting $a(d) = b^-(d)$ and $a(\emptyset) = b^-(\emptyset)$. Because of quadratic-loss proferences and that $b^-(d) < r < b^-(\emptyset)$, the jury's actual expected payoff is strictly less than it would be if the jury set $a = r$ regardless of whether the document is presented, which is the expected payoff in the setting without hard evidence.

The same argument works for the case in which $b^+(d) > r$, by supposing that the litigant's actual strategy is $\sigma = \lambda^-$ whereas the jury's initial belief is λ^+. We have therefore proved that $\underline{U}_{\mathrm{J}} < U^0_{\mathrm{J}}$ and so $L > 1$. □

Theorem 2: *Suppose that hard evidence e and private signal x are independent conditional* *on the underlying state θ. There are functions $\underline{\gamma} : [0,1] \to (0,1]$ and $\overline{\gamma} : [0,1] \to [1,\infty)$, which depend on the distributions $p_1(\cdot)$ and $p_0(\cdot)$, with the following properties. For any given ψ, Condition D is equivalent to $q_1/q_0 \le \overline{\gamma}(\psi)$ and Condition U is equivalent to $q_1/q_0 \ge \underline{\gamma}(\psi)$. Furthermore, $\underline{\gamma}$ is increasing in ψ, $\overline{\gamma}$ is decreasing in ψ, $\underline{\gamma}(1) = \overline{\gamma}(1) = 1$, $\overline{\gamma}(0) \equiv \max_{x\in X} p_0(x)/p_1(x)$, and $\underline{\gamma}(0) \equiv \min_{x\in X} p_0(x)/p_1(x)$.*

Proof: Because $f(d, x\,|\,\theta) = q_\theta p_\theta(x)$, we have

$$K^+ \equiv \{x \in X \mid q_1/q_0 \ge p_0(x)/p_1(x)\} \quad\text{and}\quad K^- \equiv \{x \in X \mid q_1/q_0 \le p_0(x)/p_1(x)\}.$$

Because these sets depend on q_1/q_0, let us write K^+ and K^- as functions of q_1/q_0. Note that K^+ is increasing in q_1/q_0 and K^- is decreasing in q_1/q_0. The inequalities for Conditions D and U simplify to, respectively:

$$\frac{q_1}{q_0} \le \frac{\psi + (1-\psi)p_0(K^-(q_1/q_0))}{\psi + (1-\psi)p_1(K^-(q_1/q_0))} \equiv \overline{y}(q_1/q_0)$$
$$\frac{q_1}{q_0} \ge \frac{\psi + (1-\psi)p_0(K^+(q_1/q_0))}{\psi + (1-\psi)p_1(K^+(q_1/q_0))} \equiv \underline{y}(q_1/q_0).$$

Let us fix ψ. Functions $\underline{y}$ and $\overline{y}$ are piecewise constant (their values change only as the sets $K^+(q_1/q_0)$ and $K^-(q_1/q_0)$ gain or lose elements as q_1/q_0 changes), $\overline{y}$ is left-continuous, and $\underline{y}$ is right-continuous. It is clear that $\phi \le \overline{y}(\phi)$ for all $\phi \le 1$ and $\phi \ge \underline{y}(\phi)$ for all $\phi \ge 1$. Further, we can show, for any $\phi' > \phi$, that $\phi > \overline{y}(\phi)$ implies that $\phi' > \overline{y}(\phi')$. Likewise for any $\phi' < \phi$, inequality $\phi < \underline{y}(\phi)$ implies that $\phi' < \underline{y}(\phi')$. In the case of $\overline{y}$, for instance, raising ϕ to ϕ' causes $\overline{y}$ to change by removing each term x for which $p_0(x)/p_1(x) \in [\phi, \phi')$. Therefore, there are numbers A, B, C, D such that

$$\phi' > \frac{A}{C} \ge \phi > \overline{y}(\phi) = \frac{1+A+B}{1+C+D}, \text{ and } \overline{y}(\phi') = \frac{1+B}{1+D}.$$

Combining the second and third inequalities and rearranging terms yields $A/C > (1 + B)/(1+D)$, which means $\phi' > \overline{y}(\phi')$. Similar calculations can be done for $\underline{y}$.

These properties imply that there exists a value $\overline{\gamma}(\psi) \ge 1$ such that $\phi \le \overline{y}(\phi)$ if and only if $\phi \le \overline{\gamma}(\psi)$, and there exists a value $\underline{\gamma}(\psi) \le 1$ such that $\phi \ge \underline{y}(\phi)$ if and only if $\phi \ge \underline{\gamma}(\psi)$, which proves the first claim of the theorem. Because $p_0(K^-) \ge p_1(K^-)$ and $p_0(K^+) \le p_1(K^+)$, we know that $\overline{y}$ is decreasing in ψ and $\underline{y}$ is increasing in ψ. The limits as ψ approaches 1 and 0 are obvious, in the latter case by recalling the definitions of K^- and K^+. $\square$

Theorem 3: *Fixing ψ, suppose that hard evidence e and private signal x are independent conditional on the underlying state θ and that $\underline{\gamma}(\psi) < \overline{\gamma}(\psi)$. Let q_1 be bounded away from 0 by a fixed number. Then there is a number $\underline{L} > 1$ so that, for all q_0 and q_1 satisfying $\underline{\gamma}(\psi) \le q_1/q_0 \le \overline{\gamma}(\psi)$, the potential loss ratio satisfies $L \ge \underline{L}$.*

Proof: This result follows directly from the analysis used to prove Theorem 1. We can construct the rationalizable outcome described at the end of the proof. Because $\underline{\gamma}(\psi) < \overline{\gamma}(\psi)$, the actual loss L is strictly greater than 1 for all q_0 and q_1 for which $\underline{\gamma}(\psi) \leq q_1/q_0 \leq \overline{\gamma}(\psi)$. The actual loss is a continuous function of q_1 and q_0, which are in a compact set given the lower bound on q_1, and so the actual loss is minimized over this set. The value $\underline{L}$ can be taken to be the minimum loss. □

Theorem 4: *Suppose that there are no limits on disclosure (the δ^{Full} disclosure policy) and*

$$\frac{q_{11}(1-q_{21})}{q_{10}(1-q_{20})} > \overline{\gamma} \quad \textit{and} \quad \frac{q_{21}(1-q_{11})}{q_{20}(1-q_{10})} > \overline{\gamma}. \tag{8}$$

Then the only rationalizable outcome entails full disclosure, the actual loss ratio comparing δ^{Full} to δ^0 is $L(\delta^{\text{Full}}, \delta^0) = 0$, and hard evidence is effective.

Proof: Our analysis proceeds with a series of claims that we prove in turn. These identify strategies that cannot be rationalized as sequential best responses and are thus removed in the iterated procedure for rationalizability.

Claim 1: Whatever is the jury's belief system, the posteriors satisfy $b(d_1d_2) > r$, $b(d_1) > r$, $b(d_2) > r$, and $b(\emptyset) \leq r$.

Using Equation 3 in the text and simplifying, we find that $b(e') > r$ is equivalent to

$$\sum_{x \in X,\ e \supset e'} \left[f(e, x \,|\, 1) - f(e, x \,|\, 0)\right] \lambda_{e'}(x, e) > 0. \tag{9}$$

For $e' = d_1d_2$, this becomes

$$\sum_{x \in X} \left[q_{11}q_{21}p_1(x) - q_{10}q_{20}p_0(x)\right] \lambda_{d_1d_2}(x, d_1d_2) > 0. \tag{10}$$

Recall that we assume $q_{11} > q_{10}$ and $q_{21} > q_{20}$. Along with the assumption for the theorem—that Inequalities 8 hold—these imply that $q_{11}/q_{10} > \overline{\gamma}$ and $q_{21}/q_{20} > \overline{\gamma}$. Also, we know that $\overline{\gamma} \geq 1$. So we have

$$\frac{q_{11}q_{21}}{q_{10}q_{20}} > \overline{\gamma}^2 \geq \overline{\gamma} \geq \frac{p_0(x)}{p_1(x)}$$

for all $x \in X$, which means that

$$q_{11}q_{21}p_1(x) - q_{10}q_{20}p_0(x) > 0$$

for all x. This proves that Inequality 10 holds regardless of $\lambda_{d_1d_2}(\cdot, d_1d_2)$, which implies $b(d_1d_2) > r$.

For $e = d_1$, Inequality 9 becomes

$$\sum_{x \in X} \left[q_{11}q_{21}p_1(x) - q_{10}q_{20}p_0(x)\right] \lambda_{d_1}(x, d_1d_2) \\ + \sum_{x \in X} \left[q_{11}(1-q_{21})p_1(x) - q_{10}(1-q_{20})p_0(x)\right] \lambda_{d_1}(x, d_1) \ > \ 0.$$

The first summation expression on the left covers the case in which $e = d_1d_2$ and the second one covers the case in which $e = d_1$. The analysis above proves that the first summation expression is positive. The second is also positive due to the first of Inequalities 8 and because $\overline{\gamma} \geq p_0(x)/p_1(x)$ for all x. At least one of the terms must be strictly positive for some x, for otherwise $\lambda_{d_1}(x, e) = 0$ for all x and $e \in \{d_1, d_{12}\}$, in which case the posterior belief is not defined. This proves that $b(d_1) > r$. The same steps establish that $b(d_2) > r$.

To prove the last inequality of Claim 1, note that if the jury's initial belief is that $\lambda_\emptyset(x, e) = 1$ for all x and e, meaning that the jury expects no documents to ever be disclosed, then $b(\emptyset) = r$. Otherwise, by the law of iterated expectation and the fact that $b(d_1d_2) > r$, $b(d_1) > r$, and $b(d_2) > r$, it must be that $b(\emptyset) < r$.

Claim 2: Every strategy of the litigant with the property $\sigma_\emptyset(x, e) > 0$ for some $x \in X$ and some $e \neq \emptyset$ is not sequentially rational and therefore removed from consideration in the iterative-elimination rationalizability procedure.

This claim follows immediately from the previous claim. From Claim 1, we know that, for every nonempty disclosure, the jury rationally must choose a higher action than it would if nothing is disclosed. Thus any strategy of the jury that would select a weakly higher action following disclosure of $\emptyset$ cannot be sequentially rational and is removed in the first round of iterated-elimination procedure. The litigant knows this and so it is never sequentially rational to disclose nothing in a contingency in which the litigant possesses a document, and strategies that would disclose $\emptyset$ are removed in the second round of the iterative-elimination rationalizability procedure.

Claim 3: Given that the jury understands Claim 2, the jury's belief system must satisfy $b(d_1d_2) > b(d_1)$ and $b(d_1d_2) > b(d_2)$.

Note that we can write the jury's posterior belief about the state in this way:

$$b(e') = \left[\frac{\displaystyle\sum_{x \in X,\ e \supset e'} (1-r) f(e, x \,|\, 0) \lambda_{e'}(x, e)}{\displaystyle\sum_{x \in X,\ e \supset e'} r f(e, x \,|\, 1) \lambda_{e'}(x, e)} + 1 \right]^{-1} \tag{11}$$

To prove Claim 3, we start by noting that the jury's belief about the litigant's strategy must satisfy $\lambda_\emptyset(x, e) = 0$ for every $x \in X$ and every $e \neq \emptyset$, given Claim 2. That is, the litigant must disclose one or both documents whenever he or she possesses some hard evidence. This implies that, when the litigant possesses exactly one document, he must disclose it for sure and so $\lambda_{d_1}(x, d_1) = 1$ and $\lambda_{d_2}(x, d_2) = 1$ for all x. Thus, the only behavior not pinned down is what the litigant would disclose when possessing both documents.

Let us compare $b(d_1)$ with $b(d_1d_2)$. To ease notation, define $y(x) \equiv \lambda_{d_1}(x, d_1d_2)$ and $z(x) \equiv \lambda_{d_1d_2}(x, d_1d_2)$. Using Equation 11, we have

$$b(d_1) = \left[\left(\frac{1-r}{r} \right) \left(\frac{q_{10}}{q_{11}} \right) \left(\frac{1 - q_{20} + q_{20} \displaystyle\sum_{x \in X} p_0(x) y(x)}{1 - q_{21} + q_{21} \displaystyle\sum_{x \in X} p_1(x) y(x)} \right) + 1 \right]^{-1}$$

and

$$b(d_1d_2) = \left[\left(\frac{1-r}{r} \right) \left(\frac{q_{10}}{q_{11}} \right) \left(\frac{q_{20} \sum_{x \in X} p_0(x) z(x)}{q_{21} \sum_{x \in X} p_1(x) z(x)} \right) + 1 \right]^{-1}.$$

Thus, $b(d_1) \geq b(d_1d_2)$ is equivalent to

$$\frac{1 - q_{20} + q_{20} \sum_{x \in X} p_0(x) y(x)}{1 - q_{21} + q_{21} \sum_{x \in X} p_1(x) y(x)} \leq \frac{q_{20} \sum_{x \in X} p_0(x) z(x)}{q_{21} \sum_{x \in X} p_1(x) z(x)}. \tag{12}$$

From the definition of $\overline{\gamma}$, we know that the right side is maximized at the value $\overline{\gamma} q_{20}/q_{21}$ by setting $z(\overline{x}) = 1$ for the draw $\overline{x} \in X$ that identifies $\overline{\gamma}$ and setting $z(x) = 0$ for all other x values. Here we are using the fact that $A/B \geq C/D > 0$ implies $(A + C)/(B + D) \leq A/B$. That is, including additional terms in the summations by raising $z(x)$ above zero would only lower the fraction.

We next show that the left side of Inequlity 12 is bounded below by 1. To demonstrate this, let us write $\rho_0 = \sum_{x \in X} p_0(x) y(x)$ and $\rho_1 = \sum_{x \in X} p_1(x) y(x)$, so the left side can be written as

$$\frac{1 - q_{20} + q_{20}\rho_0}{1 - q_{21} + q_{21}\rho_1} = \frac{1 - q_{20}(1 - \rho_0)}{1 - q_{21}(1 - \rho_1)}. \tag{13}$$

Suppose $\rho_1 < 1$. As ρ_1 and ρ_0 are total probabilities over the same weighted fraction of the space X, there must be an x-type $x' \in X$ such that $p_0(x')/p_1(x') \geq (1 - \rho_0)/(1 - \rho_1)$. Otherwise we would have a contradiction in summing over the complementary weighted fraction of X. This means that $\overline{\gamma} \geq (1 - \rho_0)(1 - \rho_1)$. From the assumption of the theorem, we also know that $q_{21}/q_{20} > \overline{\gamma}$. Putting this together with the previous inequality implies that $q_{21}(1 - \rho_1) > q_{20}(1 - \rho_0)$, and thus the value in Expression 13 is at least 1. In the case of $\rho_1 = 1$, similar reasoning establishes that we must also have $\rho_0 = 1$ and the value of Expression 13 is 1.

In summary, we have shown that the right side of Inequality 12 is bounded above by $\overline{\gamma} q_{20}/q_{21}$ and the left side is bounded below by 1. Because $q_{21}/q_{20} > \overline{\gamma}$, we therefore know that Inequality 12 cannot hold, and therefore $b(d_1) < b(d_1d_2)$. The same steps establish that $b(d_2) < b(d_1d_2)$.

Claim 4: Every strategy of the litigant with the property $\sigma_{d_1d_2}(x, d_1d_2) < 1$ for some $x \in X$ is not sequentially rational and therefore removed from consideration in the iterative-elimination rationalizability procedure.

From Claim 3, we know that the jury rationally must choose a strictly higher action when both documents are disclosed that if exactly one document is disclosed. Thus any strategy of the jury that would select a weakly higher action following disclosure of a single document cannot be sequentially rational and is removed in the third round of iterated-elimination procedure. As a result, it is never sequentially rational for the litigant to disclose a single document when he possesses both documents, and such strategies are removed in the fourth round of the iterative-elimination rationalizability procedure.

Combining Claims 2 and 4, we conclude that every rationalizable outcome entails full disclosure. Because of this, the jury optimally responds as though the jury directly observed the existing documents. Any non-optimal action choice is removed in the fifth round of the iterative-elimination rationalizability procedure. □

Theorem 5: *Suppose that*

$$\frac{q_{11}q_{21}}{q_{10}q_{20}} > \overline{\gamma}.$$

In every rationalizable strategy profile of the litigation game in which the litigant is restricted to disclosures in the set δ^{Bundle}, the litigant always discloses both documents when they both exist. Hard evidence is effective and the actual loss ratio for δ^{Bundle} compared to the no-evidence policy δ^0 is $L(\delta^{\text{Bundle}}, \delta^0) = 0$.

Proof: The proof of this theorem follows the same steps as in the proof of Theorem 1 (and Theorem 2), where we now interpret $d_1 d_2$ as a single document. □

Theorem 6: *Suppose that Inequalities 4 hold. In the setting with no limits on disclosure (the δ^{Full} disclosure policy), there are rationalizable outcomes in which disclosure of a single document is misleading but disclosure of both documents is effective, and there are also rationalizable outcomes in which all hard evidence is effective. The potential loss ratio comparing δ^{Bundle} to δ^{Full} is $L(\delta^{\text{Full}}, \delta^{\text{Bundle}}) > 1$.*

Proof: The proof is along the same lines as that of Theorem 1 using the description of beliefs and behavior that follow the statement of Theorem 6 in the text. □

B Examples and Calculations

This appendix contains calculations for the Rule 403 example in the text, and an additional example for Rule 404.

Rule 403 Numerical Case Study Calculations

Based on the timeline and probabilities with which events occur specified in the example, we can calculate joint probabilities $f(e, x, \theta, A)$, where A denotes that the accident occurred. These are:

$$f(d, \overline{x}, 1, A) = \frac{w}{32},\ f(d, \overline{x}, 0, A) = \frac{3}{512},\ f(d, \underline{x}, 1, A) = \frac{1-w}{32},\ f(d, \underline{x}, 0, A) = \frac{9}{512},$$
$$f(\emptyset, \overline{x}, 1, A) = 0,\ f(\emptyset, \overline{x}, 0, A) = \frac{w}{8} + \frac{1}{64},\ f(\emptyset, \underline{x}, 1, A) = 0,\ \text{and}\ f(\emptyset, \underline{x}, 0, A) = \frac{1-w}{8} + \frac{3}{64}.$$

From these, one can calculate the conditional probabilities, which suppress the A, found in the text.

Rule 404 Numerical Case Study – Character Evidence

Judges in the U.S. are generally reluctant to allow character evidence or evidence of a prior conviction about defendants in criminal cases, as Rule 404 forbids under some conditions. Here, we do not explore the more nuanced issues concerning a criminal defendant who chooses to testify at his trial.[30] Instead, we consider general ideas such as those found in People v. Beagle, 6 Cal. 3d 441, a well-known case in which the defendant testified, and we focus on evidence of a prior conviction presented in this case.

Harvey Lynn Beagle II appealed his conviction by a jury of attempted arson and arson.[31] Here is a brief summary of the relevant facts of the case. On May 25, 1969, Beagle was kicked out of Rudy's Keg because he "became intoxicated and obnoxious while a patron in the bar." At the time, Beagle made comments, about hiring someone to "fire bomb" the bar. Then in the afternoon of July 1st, Beagle asked the owner of the bar if he could return to the bar and was told no. Later that night, the roof of the building that housed the bar caught fire after what seemed to be an explosion. The bar owner put out the fire and discovered a soda bottle containing gasoline and a wick. Shortly after that, Beagle was arrested at his nearby apartment. He smelled of gasoline and had several books of matches in his pockets.

Beagle appealed on several grounds. One of these was that evidence of his having a prior conviction for writing a bad check was allowed. The Supreme Court of California found this inappropriate.[32] The general idea is that a judge should have discretion to exclude some prior convictions from being admitted as evidence.[33]

Consider a numerical example, motivated by Beagle, in which a prosecutor may choose whether to disclose d, which represents evidence of a prior conviction of the defendant for writing a bad check. A time line of the events follows.

1. A random draw determines whether there is a bad check prior conviction (c) or not (n). Evidence d exists if and only if c is realized. Assume c occurs with probability 1/2 and n occurs with probability 1/2.
2. If c is realized, the person either reforms, which we denote by g, or has a higher propensity for criminal behavior, which we denote by b. Assume g occurs with probability 1/2 so that b also occurs with probability 1/2.
3. The defendant is matched with a bar/situation. The type with no prior conviction

[30] There is a bit more judicial discretion when the defendant testifies as a witness because there is scope for using character evidence or evidence of prior convictions to impeach a witness. Potentially different rules apply to civil cases.

[31] The Supreme Court of California heard the case on January 5, 1972.

[32] It noted: "Although we reject all of the many contentions presented by defendant on appeal from the judgment, we nevertheless conclude, inter alia, that a trial judge must exercise his discretion to prevent impeachment of a witness by the introduction of evidence of a prior felony conviction when the probative value of such evidence is substantially outweighed by the risk of undue prejudice. (See Evid. Code, §352.)"

[33] The court noted that the nature of the prior conviction and whether it reflects badly on the defendant's honesty or integrity is a factor in determining whether it should be allowed to impeach the defendant as a witness. How recent the prior conviction was is also a factor. Additionally, it's noted that a prior conviction for a similar crime should be "admitted sparingly." The court cites Judge (later Chief Justice) Burger in Gordon v. United States (1967) 383 F.2d 936, 940-941 [127 App.D.C. 343] on these. The idea is that prior convictions for similar crimes may put significant pressure on a jury to convict. The court also suggested that the effects on the incentive for the defendant to testify should also be considered. We suggest that all of theses issues fit with our model.

commits the crime with probability $1/5$, type g with probability $1/10$, and type b with probability $1/2$.[34] If the defendant does not commit the crime, with probability $1/5$ someone else does.
4. Following commission of the crime and the defendant's arrest, the prosecutor's private information x is realized as follows.[35] If the defendant has no prior conviction, $x = g$. For type g, $\overline{x}$ is realized with probability $1/4$ and $\underline{x}$ is realized with probability $3/4$. For type b, $\overline{x}$ is realized with probability w and $\underline{x}$ is realized with probability $1 - w$.

Our model picks up the story at Date 4. As we are considering the prosecutor's disclosure decision, $\theta = 1$ corresponds to the defendant being guilty. We assume that the case is not brought when the crime is not committed.

By constructing an event tree and calculating the probabilities of the various paths, we obtain the following joint probabilities of the form $f(e, x, \theta, A)$, where A denotes that the crime was committed:

$$f(d,a,1,A) = 0,\ f(d,a,0,A) = 0,\ f(d,\overline{x},1,A) = \frac{1}{160} + \frac{w}{8},\ f(d,\overline{x},0,A) = \frac{9}{800} + \frac{w}{40},$$
$$f(d,\underline{x},1,A) = \frac{3}{160} + \frac{1-w}{8},\ f(d,\underline{x},0,A) = \frac{27}{800} + \frac{1-w}{40},\ f(\emptyset,a,1,A) = \frac{1}{10},$$
$$f(\emptyset,a,0,A) = \frac{2}{25},\ f(\emptyset,\overline{x},1,A) = 0,\ f(\emptyset,\overline{x},0,A) = 0,\ f(\emptyset,\underline{x},1,A) = 0,\ \text{and}\ f(\emptyset,\underline{x},0,A) = 0.$$

From these, we obtain the following conditional probabilities (suppressing A):

$$f(d,\overline{x} \mid 1) = \frac{1}{40} + \frac{w}{2},\ f(d,\overline{x} \mid 0) = \frac{3}{40} + \frac{w}{6},$$
$$f(d,\underline{x} \mid 1) = \frac{3}{40} + \frac{1-w}{2},\ \text{and}\ f(d,\underline{x} \mid 0) = \frac{9}{40} + \frac{1-w}{6}.$$

As in the numerical example shown in the text, there are two possibilities for evidence to be misleading here. Naturally, this depends on how informative the litigant's private signal is, which here depends on the value of w.

Consider first the case in which $f(d,\underline{x}\,|\,1) = 3/40 + [1/2][1-w] \leq f(d,\underline{x}\,|\,0) = 9/40 + [1/6][1-w]$ and $f(d,\overline{x}\,|\,1) = 1/40 + w/2 \geq f(d,\overline{x}\,|\,0) = 3/40 + w/6$. That is, $K^- = \{\underline{x}\}$ and $K^+ = \{\overline{x}\}$. This requires $w \geq 11/20$. So for large values of w, meaning that the prosecutor has an accurate private signal of whether the previously convicted defendant has reformed and $\overline{x}$ suggests b, the conditions for evidence to be misleading are satisfied. Next consider the case in which $f(d,\overline{x}\,|\,1) = 1/40 + w/2 \leq f(d,\overline{x}\,|\,0) = 3/40 + w/6$ and $f(d,\underline{x} \mid 1) = 3/40 + [1/2][1-w] \geq f(d,\underline{x} \mid 0) = 9/40 + [1/6][1-w]$ so that $K^- = \{\overline{x}\}$ and $K^+ = \{\underline{x}\}$. This requires $w \leq 3/20$, implying that the prosecutor has accurate private

[34] Regarding the probabilities with which each type commits the crime, we note the following. Someone who has been convicted of writing a bad check might learn from the experience, and wish to avoid any further legal problems. For that type of defendant, the prior conviction for writing a bad check may actually make it less likely that he would commit a crime like arson. On the other hand, it's possible the conviction caused the defendant to feel animosity towards those in authority, including someone who owns a bar and can prevent him from drinking there. In this case, the prior conviction may make it more likely that the defendant would commit arson.

[35] The prosecutor may be better informed than the fact-finder regarding the defendant's type.

information about whether the previously convicted defendant has reformed. We conclude that for extreme values of w, where the litigant's private signal is informative relative to the hard evidence, there is a rationalizable outcome in which evidence d is misleading.

References

Anderson, T., Schum, D., and Twining, W. *Analysis of Evidence*, 2nd ed., Cambridge: Cambridge University Press, 2005.

Banks, J. S. and Sobel, J. "Equilibrium Selection in Signaling Games," *Econometrica* Vol. 55 (1987), pp. 647–661.

Brady v. Maryland, 373 U.S. 83 (1963).

Bull, J. "Mechanism Design with Moderate Evidence Cost," *The B.E. Journal of Theoretical Economics*, (Contributions), Vol. 8 (2008), Article 15.

Bull, J. "Costly Evidence and Systems of Fact-Finding," *Bulletin of Economic Research*, Vol. 61 (2009), pp. 103–125.

Bull, J. "Interrogation and Evidence Fabrication," working paper, Florida International University, 2012.

Bull, J. and Watson, J. "Evidence Disclosure and Verifiability," *Journal of Economic Theory*, Vol. 118 (2004), pp. 1–31.

Bull, J. and Watson, J. "Hard Evidence and Mechanism Design," *Games and Economic Behavior*, Vol. 58 (2007), pp. 75–93.

Capra, D. J. *Case Law Divergence from the Federal Rules of Evidence*. Washington, D.C.: Federal Judicial Center, 2000

Che, Y. and Severinov, S. "Disclosure and Legal Advice," *American Economic Journal: Microeconomics*, Vol. 9 (2017), pp.188–225.

Cho, I.-K. and Kreps, D.M. "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, Vol. 102 (1987), pp. 179–221.

Cleary, E. W., ed., *McCormick's Handbook on the Law of Evidence*, 2nd ed., St. Paul, MN, West Publishing (1972).

Crawford, V. P. and Sobel, J. "Strategic Information Transmission," *Econometrica*, Vol. 50 (1982), pp.1431–1451.

Damaska, M.R. *Evidence Law Adrift*. New Haven, CT: Yale University Press. 1997.

Daughety, A. and Reinganum, J. "Appealing Judgments." *RAND Journal of Economics*, Vol. 31 (2000a), pp. 502–525.

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

Dolan, A. K. "Rule 403: The Prejudice Rule in Evidence," *Southern California Law Review*, Vol. 49 (1976), pp. 220–85.

Federal Rules of Evidence.

Fluet, C. and Lanzi, C. "Adversarial Persuasion with Cross-Examination," manuscript, Université Laval (2018).

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

Guttman, I., Kremer, I., and Skrzypacz, A. "Not Only What but Also When: A Theory of Dynamic Voluntary Disclosure," *American Economic Review*, Vol. 104 (2018), pp. 2400–2420

Gold, V. J. "Jury Wobble: Judicial Tolerance of Jury Inferential Error," *Southern California Law Review*, Vol. 59 (1986), pp. 391–406.

Gulf States Utilities Company v. Ecodyne Corporation, 635 F.2d 517 (5th Cir. 1981)

Hastie, R. "The role of 'stories' in civil jury judgments," *Michigan Journal of Law Reform*, Vol. 32 (1999), pp. 1–13.

Hedlund, J. "Bayesian Persuasion by a Privately Informed Sender." *Journal of Economic Theory*, Vol. 167 (2017), pp. 229–268.

Kartik, N, and Tercieux, O. "Implementation with Evidence," *Theoretical Economics*, Vol 7 (2012), pp. 323–355.

Keefe, J., Landis, Jr., W. B., and Shad, R.B. "Sense and Nonsense about Judicial Notice," *Stanford Law Review*, Vol. 2 (1950), pp. 664–690.

Langevoort, D. C. "Behavioral Theories of Judgment and Decision Making in Legal Scholarship: A Literature Review," *Vanderbilt Law Review*, Vol. 51 (1998), 1499–1540.

Lempert, R., Gross, S., and Liebman, J. *A Modern Approach to Evidence*. 3rd ed., St. Paul, MN: West Group, 2000.

Lester, B., Persico, N., and Visschers, L. "Information Acquisition and the Exclusion of Evidence in Trials," *Journal of Law, Economics, and Organization*, Vol. 28 (2009), pp. 163–182.

Michaels, D. and Rapoport, M. "SEC Probes Tesla CEO Musk's Tweets," *The Wall Street Journal*, August 8, 2018.

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

Mueller, C. B. and Kirkpatrick, L.C. *Evidence*, 5th ed., New York: Wolters and Kluwer, 2012.

Pennington, N., and Hastie, R. "Evidence evaluation in complex decision making," *Journal of Personality and Social Psychology*, Vol. 51 (1986), pp. 242–258.

Pennington, N., and Hastie, R. “A cognitive theory of juror decision making: The Story Model,” *Cardozo Law Review*, Vol. 13 (1991), pp. 519–557.

Pennington, N., and Hastie, R. “Explaining the evidence: Tests of the story model for juror decision making,” *Journal of Personality and Social Psychology*, Vol. 62 (1992), pp. 189–206.

Rubin, J., and Bloomkatz, A.B. “Interrogation, then Revenge.” *Los Angeles Times*, July 2, 2008.

Rubin, J. “This L.A. sheriff’s deputy was a pariah in federal court. But his secrets were safe with the state,” *Los Angeles Times*, August 10, 2018.

Sanchirico, C. W. “Character Evidence and the Object of Trial,” *Columbia Law Review*, Vol. 101 (2001), 1227–311.

Sanchirico, C. W. “Models of Evidence: Survey and Assessment,” working paper, University of Pennsylvania, 2010.

Schrag, J., and Scotchmer, S. “Crime and Prejudice: The Use of Character Evidence in Criminal Trials,” *Journal of Law, Economics, and Organization*, Vol. 10 (1994), pp. 319–42.

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

Swift, E. “One Hundred Years of Evidence Law Reform: Thayer’s Triumph,” *California Law Review*, Vol. 88 (2000), pp. 2437–2476.

Talley, E. “Law, Economics, and the Burden(s) of Proof.” In J. Arlen, ed., *Research Handbook on the Economics of Torts*. Northampton, MA: Edward Elgar, 2013.

Thayer, J.B. *A Preliminary Treatise on Evidence at the Common Law*. Boston: Little Brown, 1898.

Twining, W. *Rethinking Evidence: Exploratory Essays*. Evanston, IL: Northwestern University Press, 1994.

United States v. Bagley, 473 U.S. 667 (1985).

United States v. Pugliese, 153 F.2d 497 (2d Cir. 1945).

Watson, J. “Information Transmission when the Informed Party is Confused,” *Games and Economic Behavior*, Vol. 12 (1996), pp. 143–161.

Watson, J. “A General, Practicable Definition of Perfect Bayesian Equilibrium,” manuscript, University of California, San Diego, 2017.