12-2017

Statistical Evidence and the Problem of Robust Litigation

Jesse Bull
Florida International University

Joel Watson
University of California - San Diego

Statistical Evidence and the Problem of Robust Litigation

Jesse Bull and Joel Watson*

December 2017

Abstract

We develop a model of statistical evidence with a sophisticated Bayesian fact-finder. The context is litigation, where a litigant (defendant or plaintiff) may disclose hard evidence and a jury (the fact-finder) interprets it. In addition to hard evidence, the litigant has private unverifiable information. We study the robustness of the parties' reasoning regarding the legal fundamentals and the litigant's strategic behavior. The litigant's choice of whether to disclose hard evidence entails two channels of information: the face-value signal of the hard evidence disclosure (relating to the probabilities that the hard evidence exists in different states of the world) and as a possible signal of the litigant's private information. Our results suggest that in some situations, a desire for robust reasoning about evidence would lead the court to restrict the admissibility of some relevant evidence. The modeling exercise provides support for the Federal Rules of Evidence Rules 403 and 404, along with general conclusions about evidence policy.

*Bull: Florida International University; Watson: UC San Diego. The authors thanks Alex Weiss for superb research assistance. They are also grateful to the following people for providing input: Dan Klerman, Aaron Kolb, Alex Lee, Herb Newhouse, Eric Talley, Simon Wilkie, and seminar participants at Indiana University, USC, UC Riverside, UCF, the 2015 Law and Economic Theory conference, the 2017 American Law and Economics Association conference, the 2016 Canadian Law and Economics Association conference, and the Fall 2017 Midwest Economic Theory conference.

1 Introduction

The traditional "information theory" of evidence in litigation holds that, assuming fact-finders can process and interpret information well, evidence is always beneficial because it provides information about the underlying matter to be resolved. Thus, courts should allow a wide range of evidence, however complicated and without regard to weight. An alternative view maintains that fact-finders face cognitive limitations and may have difficulty dealing with complicated or varied information. Courts then have an interest in constraining evidence in ways that would help fact-finders avoid inferential error, although the exact policy conclusions are not so clear.[1]

We develop an alternative theory of evidence, distinct from the information and cognitive-limitation views, that has direct implications for the design of evidentiary rules. Our model examines the interaction between a litigant and a fact-finder whom we generally take to be a jury. Retaining the information theory's assumption of reasoned fact-finders, we show that constructive evidence requires *alignment* of the litigant's and fact-finder's beliefs about the meaning of evidence. This meaning is generally not fixed because it relates to the circumstances under which the litigant would choose to disclose available evidence, and the parties may have different beliefs about the litigant's behavior. Absent coordination of beliefs and behavior, some types of evidence may be misleading in that they cause the fact-finder to update her belief in the opposite direction than she would if she knew the litigant's actual disclosure strategy.

Our theory provides a different behavioral foundation than does the cognitive-limitation view. Essentially, it is not so much a problem of fact-finders lacking the ability to process information but rather a problem of fact-finders having great latitude in how they might interpret evidence due to the presence of private information on the part of the litigant. Where the interpretive latitude is wide, fact-finders may not accurately predict a litigant's disclosure strategy, resulting in misleading evidence to society's detriment. The prospect of misleading evidence is inherent in the nature of hard evidence and the severity of the problem depends on the statistical parameters. Therefore, courts may do well to make some evidence inadmissible, as in Rules 403 and 404 of the Federal Rules of Evidence and in common law,

[1] Gold (1986) discusses how fact-finders may err in determining how probative evidence is or may demonstrate systematically biased reasoning in evaluating the evidence presented to them. Langevoort (1998) surveys behavioral theories of judgment and decision making in legal scholarship. A prominent theory is the "story view" of jury decision making, 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).

or to provide guidelines for its interpretation.

To understand the scope for interpretation, 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.

For example, on the question of whether the defendant in a trial 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. This piece of 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.[2] 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. Therefore, disclosure of the surveillance video (the hard evidence in this illustration) may lead a sophisticated jury (the fact-finder in this scenario) to update its belief about the defendant's involvement in the crime, raising the probability that the defendant is innocent but not raising it to 1.

By adding some details to the robbery sketch just described, we can illustrate the main idea of our modeling exercise. One might have imagined that the hard-evidence video would be disclosed if and only if it exists, so that the jury extracts from its disclosure exactly the *face-value signal* that its existence or nonexistence provides. That is, the face-value signal is the information that would be transmitted if the jury were to directly observe whether the video exists, so that it performs a Bayes'-rule update about the defendant's culpability based on the probabilities that the video would exist conditional on guilt and conditional on innocence.

However, a second channel of information operates with the surveillance video: It may signal the defendant's private information (the defendant's *type*), in particular if the jury thinks that different types of defendant would be inclined to disclose the video evidence with different probabilities. For example, suppose the jury believes that a defendant who knows he committed the crime (the "bad type") is likely to disclose the video evidence when it exists, whereas a defendant who knows that he didn't commit the crime (the "good type") is less likely to do so. Then disclosure of the video evidence provides a signal of the defendant's

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

type by way of the two types' different disclosure probabilities. In this case 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.

Thus, disclosure of surveillance video showing the defendant at 9:20 p.m. provides information through two channels: the face-value signal relating to the existence of the video and the signal of the defendant's type implied by the types' different disclosure probabilities. Because both 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 defendant-type signal, then the jury will update toward innocence regardless of the jury's belief about the defendant's behavior. 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. Hard evidence in this case is effective.

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 behavior. 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. 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.

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 instance, 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. In general and in reality, litigants have private unverifiable information in addition to hard evidence, and so a piece of hard evidence can provide information to fact-finders both through its face-value signal and through its signal of the litigant's private information.

To summarize, the example demonstrates that there are circumstances in which both types of defendant and the jury are sophisticated Bayesians, they rationally best respond to their beliefs, these facts are *common knowledge* between them, and yet hard evidence is misleading to society's detriment. Notably, this is a non-equilibrium phenomenon. As we argue in the paper, there is good reason to doubt that players in the settings studied here would coordinate on an equilibrium, much less society's preferred equilibrium.[3] 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 litigant and fact-finder are coordinated on a desired equilibrium.

To evaluate robustness, the solution concept we employ is *rationalizability*, which identifies the range of possible outcomes consistent with common knowledge of rationality (and not necessarily beliefs and behavior that are coordinated across players). Thus, a technical sidepoint of this paper is to suggest robustness as an important criterion for legal institutions and rationalizability as a useful concept for studying it. Our robustness criterion evaluates whether evidence is meaningful and constructive in all rationalizable strategy profiles of our litigation game. Our main point is a simple one: Robustness is difficult to achieve without imposing some restrictions on admissibility of evidence and/or on how the fact-finder may interpret evidence. In other words, to prevent misleading evidence, the court may optimally restrict evidence. Our results provide justification for Rules of Evidence 403 and 404.

Before going on to the model, let us make a few comments regarding the related literature. In the law-and-economics literature, two main approaches to modeling evidence stand out.[4] The first treats evidence as statistical in nature, as just described, but it views evidence as arriving exogenously. These models are exercises in Bayes' rule but they address neither the parties' incentives to disclose evidence nor the fact-finder's evaluation of these incentives. The second modeling approach focuses on the incentives of the litigants to produce evidence, but it views the adjudicator as a mechanistic system whose judgment is an exogenous function of the quantities of evidence that the two sides produce. Evidence 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 in an abstract way and they conclude that the types of litigants

[3] While experienced judges and attorneys may be more likely to think and behave in coordinated ways, it is less likely that fact-finders (typically juries, whose members do not routinely hear cases) and litigants (who may not have much experience in court) are so coordinated. In practice, the Rule 403 exclusion for potentially misleading evidence is not viewed as necessary for bench trials, which have a judge in a fact-finding role. See Capra (2001).

[4] Here we are summarizing Talley's (2013) characterization of the law-and-economics literature on evidence. See also Sanchirico (2010).

will be separated in equilibrium, yet the adjudicator does not utilize this signal.[5]

Some prominent entries in the literature feature both the litigants' incentive to disclose evidence and a Bayesian decision maker, but they assume an extreme view of hard evidence as definitive proof of the state or some subset of possible states.[6] 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.[7] Tangentially related "Bayesian-persuasion" models involve a sender committing to an informative experiment to influence a receiver.[8]

In reality, the litigants control most pieces of evidence and disclosure is subject to their individual incentives. Evidence is discrete and statistical; a piece of evidence either exists or doesn't exist, and the chance of existence depends on the state of the world. Producing evidence may be costly, but the cost differential between states is typically small (for instance, if the surveillance video exists, then a culpable defendant can just as easily obtain and present it to the court as can a non-culpable defendant) or very large (such as the culpable defendant having to fabricate it).[9] Juries and other fact-finders are typically sophisticated enough to assess the litigants' incentives and recognize the signal inherent in the disclosure, or lack of disclosure, of evidence.

The following section presents our basic model, which limits attention to a simple setting with one litigant, a jury that will impose a judgment, and one "document" (the piece of hard evidence) 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

[5]Daughety and Reinganum (2000a) use an axiomatic method to study the processing of information by trial courts in a model in which evidence is the result of strategic search and the court observes only that evidence presented by the litigants. Daughety and Reinganum (2000b) study bias, in an axiomatic approach with a strategic search model of evidence, due to differences in the litigant's sampling cost or sampling distribution.

[6]Milgrom (1981) and Shin (1994) are classic 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.

[7]Bull and Watson (2004 and 2007), Green and Laffont (1986), and Kartik and Tercieux (2012) are examples of 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.

[8]See Watson (1996) for an early version of this type of model. Those models typically assume that the sender and receiver have shared prior information. Hedlund (2017) studies Bayesian persuasion with a privately-informed sender.

[9]In real settings, parties also may invest resources to gather evidence. We do not consider this strategic variable in the present paper.

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. In this case we say that hard evidence is effective.

Section 4 briefly discusses implications of the basic model for the courts. For the sake of robust usage of hard evidence in litigation, under some conditions it is optimal for the court to make the document inadmissible. This is essentially an exercise in mechanism design, where the objective is to implement a mapping from the realized evidence to the judgment, but we restrict ourselves to the nearly trivial but realistic design element of whether to allow the single document to be admitted into evidence.

In Section 5 we extend our model to the case of two documents, which allows for an analysis of a wider range of evidentiary rules than in the basic model. For instance, the court can make a single document inadmissible but allow the two documents to be disclosed together. We show that such a rule is optimal for robustness in settings where the face-value signal of a single document is relatively weak and yet the face-value signal of two documents is strong. 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 examples based on some well-known cases, is contained in Section 6. We conclude in Section 7. Proofs of the theorems may be found in the Appendix.

2 Basic Model

2.1 Description of the Game

We study a simple two-player game with hard evidence. The first player has information about an *underlying state of the world* and may be able to disclose hard evidence about it. The second player observes whatever evidence is presented and then takes an action that affects both players. The model portrays litigation in court, where the first player is a *litigant* (plaintiff or defendant) and the second player is the *fact-finder* (typically a jury).[10] To keep

[10]The model has applications outside of the legal realm but we focus on the legal application here.

things simple, we call the players the "litigant" and "jury," and we think of the jury as a single agent. The jury and society care about the state and also about the jury's decision, which in practical terms is a finding in the case. Let θ denote the state and suppose that $\theta \in \Theta \equiv \{0,1\}$. We will speak of $\theta = 0$ as the "low state" and $\theta = 1$ as the "high state." Later we shall round out our description of the court by including a judge in the story.

Whereas the jury does not observe the state, the litigant has two sources of information about it. 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 obtain *hard evidence*, which is verifiable and can be disclosed to the jury. Suppose hard evidence takes the following binary form: The litigant may or may not possess a single document d. We represent the hard evidence that the litigant possesses by the *evidentiary state* $e \in E \equiv \{d, \emptyset\}$, where $e = d$ means the litigant possesses the document (and we say "the document exists") and $e = \emptyset$ means that the litigant does not possess the document ("the document does not exist"). If $e = d$ then the litigant can choose to either disclose the document or disclose nothing. If $e = \emptyset$, then he must disclose nothing; this is the defining characteristic of hard evidence. We sometimes describe disclosing nothing as "disclosing $\emptyset$." The jury observes only whether d is disclosed, not whether d exists. That is, the jury does not observe e.

The underlying state θ, the evidentiary state e, and the private signal x are determined exogenously by nature and in general are correlated.[11] Let f denote the joint distribution, so that $f(\theta, e, x)$ is the probability that (θ, e, x) is realized. Defining f over sets, we also 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 $r \equiv f(1, E, X)$ be 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 $x \in X$.

Consider the *face-value signal* of the hard evidence, which 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

[11] While it would be appropriate to call the entire vector (θ, e, x) the *state* (with no qualifier "underlying" or "evidentiary"), we sometimes call θ the state because this is what is of direct interest to the jury.

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. 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.[12]

The inference that the jury can draw from the litigant's disclosure (or lack of disclosure) depends not only on the properties of the document but also on the incentives of the litigant to disclose it, and the litigant's behavior may be conditioned on his private signal x. Therefore, any inference to be made from the disclosure or nondisclosure should incorporate both that the document cannot be disclosed if it does not exist—this is the direct information from hard evidence—and how x influences the decision to disclose, which we refer to as the soft signaling role.

After seeing whether the litigant discloses d, the jury updates its belief about the underlying state and selects its action. To keep things simple, suppose that the jury's action is a selection $a \in [0, 1]$ representing, for instance, the degree to which the litigant is held responsible for a crime or the amount of monetary damages to award the litigant. Assume that the jury's (and society's) payoff is decreasing in the square of the difference between the action a and the state θ, so the jury's payoff is $u_{\mathrm{J}}(a, \theta) = -(a - \theta)^2$. This implies that the jury's optimal action is equal to its posterior probability of the high state.

Assume that the litigant's payoff is strictly increasing in the jury's expected action, regardless of the underlying state. The simplest such payoff function is $u_{\mathrm{L}}(a, \theta) = a$. It is thus in the litigant's interest to act in whatever fashion will maximize the jury's posterior probability of $\theta = 1$.[13] More generally, we could allow u_{L} to be a function of the state or even of the litigant's private signal and realization of hard evidence, but this will not be necessary for the logical connections that we focus on.

For most of our analysis, we will not need to examine the jury's action directly. Rather, we can formulate the analysis in terms of the jury's posterior belief. We assume that the jury is Bayesian in that its posterior belief results from a proper application of the conditional probability formula, given the jury's belief about the litigant's strategy. For now, we also

[12] 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.

[13] Another setting that leads to the same analysis and results is one in which the jury has only two actions available, such as finding the litigant guilty or not guilty, and the jury prefers or is instructed to choose not guilty in the high state and guilty in the low state. However, in addition to the information received from the litigant's evidence choice, the jury is influenced by a separate, independent noisy information source. Therefore, the jury's judgment is random and increasing in the posterior conditional on the litigant's evidence choice.

assume that the court cannot force the jury to commit to a decision rule in advance of the litigant's disclosure choice.

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 $e = d$, the litigant decides whether to disclose d. Finally, the jury observes whether d is disclosed, forms its posterior belief about the state θ, and selects its action a.

2.2 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 can actually be different people in the real world. For instance, consider the following simple story about the events leading to a court case.

There is a variety of individuals in society and they may differ 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.[14] 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. 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

[14] The litigant may lack an understanding of the law or is unsure of whether his behavior was criminal.

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.

2.3 Strategies and Beliefs

A pure strategy for the litigant is a function mapping X to the choice of whether to disclose, in the event that $e = d$.[15] A mixed (behavior) strategy for the litigant is given by a function $\sigma : X \to [0, 1]$, where for each $x \in X$, $\sigma(x)$ is the probability that the litigant discloses the document in the event that $e = d$ and his private signal is x. "Full disclosure" refers to the strategy that always discloses the document, so $\sigma(x) = 1$ for every $x \in X$.

To enrich the model a bit, we will assume that the litigant's disclosure probabilities are bounded below by a number $\psi \in [0, 1]$. That is, the litigant is constrained to choose $\sigma(x) \geq \psi$ for every x. This lower bound captures the idea that the litigant may be influenced by a lawyer or other party who induces the litigant to disclose available hard evidence, or that the document is automatically disclosed with some probability. We make no assumption about ψ, so $\psi = 0$ is allowed and in this case the litigant is unconstrained.

The jury's posterior belief regarding the state is 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.

[15] Recall that the litigant has no choice in the event that $e = \emptyset$.

It is important to recognize that $b(d)$ and $b(\emptyset)$ depend on the jury's belief about the litigant's strategy, as well as the jury's understanding of the information system. Likewise, the litigant's optimal strategy depends on the litigant's belief about the jury's updating rule. Let $\overline{b}(d)$ denote the mean of the litigant's belief about $b(d)$, and let $\overline{b}(\emptyset)$ denote the mean of the litigant's belief about $b(\emptyset)$.

In the event that the document exists, it is clearly optimal for the litigant to disclose it if $\overline{b}(d) > \overline{b}(\emptyset)$ and to not disclose if $\overline{b}(d) < \overline{b}(\emptyset)$. The litigant is indifferent if $\overline{b}(d) = \overline{b}(\emptyset)$. These incentives do not depend on the litigant's private signal x because the litigant cares only about increasing the jury's posterior belief, which is a function of evidence only.

Let us represent the jury's belief about the litigant's strategy as a function $\lambda\colon X \to [0,1]$, where for each $x \in X$, $\lambda(x)$ is the probability that the jury thinks the litigant discloses the document 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 $\sum_{x\in X} f(\Theta,d,x)\lambda(x) > 0$, then the jury's posterior belief conditional on disclosure is given by Bayes' rule:

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

Likewise, the Bayes' rule expression for the jury's posterior belief conditional on nondisclosure is:

$$b(\emptyset) = \frac{r\left[1 - \sum_{x\in X} f(d,x \,|\, 1)\lambda(x)\right]}{r\left[1 - \sum_{x\in X} f(d,x \,|\, 1)\lambda(x)\right] + (1-r)\left[1 - \sum_{x\in X} f(d,x \,|\, 0)\lambda(x)\right]}, \tag{2}$$

where the denominator is always strictly positive because of our assumption that the document exists with a probability strictly less than one. However, the denominator in the equation for $b(d)$ is zero if the jury's initial belief about the litigant is $\lambda(x) = 0$ for all x, and in this case the expression is not valid (Bayes' rule overall does not apply). The solution concepts we study impose constraints on the jury's belief in such a situation, as we demonstrate in the next subsection.

2.4 Solution Concept, Welfare, and Admissibility

We shall analyze the game using mainly 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 instance, the jury will not put positive probability on a strategy for the litigant that itself cannot be rationalized. 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, so that $\bar{b}(d) \neq b(d)$ and $\bar{b}(\emptyset) \neq b(\emptyset)$. We shall adopt a version of rationalizability in which the players' beliefs are assumed to be *plainly consistent* (Watson 2015); the implications and motivation are discussed below.

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.[16]

With the rationalizability concept, our model does 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.

In our model, plain consistency puts some structure on the jury's beliefs. It implies the standard use of Bayes' rule to calculate the posteriors $b(d)$ and $b(\emptyset)$ if the jury initially put positive probability on disclosure. That is, if λ gives the jury's initial belief about the litigant's strategy, then Equations 1 and 2 hold if the denominators are strictly positive. In fact, plain consistency further implies that the posterior beliefs have this structure even in the case in which the jury initially put zero probability on the document being disclosed.

Theorem 1: *The following holds for every belief system satisfying plain consistency. The jury's posterior belief $b(\emptyset)$ satisfies Equation 2, where λ is the jury's initial belief about the*

[16]In the Appendix we also look at the stronger solution concept of perfect Bayesian equilibrium (PBE), which for our model is equivalent to sequential equilibrium. In a PBE, the players are rational and their beliefs and behavior are aligned, so that $\bar{b}(d) = b(d)$, $\bar{b}(\emptyset) = b(\emptyset)$, and $\lambda = \sigma$. Furthermore, the players' beliefs satisfy plain consistency (See Watson 2015). The main reason to look at PBE is that it provides intuition for the rationalizability construction. It also may be an appropriate solution concept for settings in which the legal institution or some other institution serves to align beliefs and behavior. We discuss later the role of transparency and interpretive guidance in the legal system.

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.*[17]

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 jury's expected payoff—that is, 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.

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.[18]

Let us say that *hard evidence is (partially or completely) ineffective* if the litigant plays a strategy other than full disclosure, 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. We'll say that *hard evidence is useless* if the litigant never discloses it. Further, let $b^{\sigma}(d)$ and $b^{\sigma}(\emptyset)$ be the posterior probabilities from Equations 1 and 2 under the assumption that $\lambda = \sigma$, where we recognize that the former is defined only if $\sigma(x) > 0$ for some x satisfying $f(\Theta, d, x) > 0$. These would be the jury's posterior beliefs if the jury knew the litigant's actual strategy. 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 $b(\emptyset) \neq b^{\sigma}(\emptyset)$. That is, the jury's posterior beliefs are not consistent

[17] This means that the beliefs are structurally consistent (Kreps and Ramey 1987).

[18] Let $b^{*}(d) = \frac{f(1,d,X)}{f(\Theta,d,X)}$ and $b^{*}(\emptyset) = \frac{f(1,\emptyset,X)}{f(\Theta,\emptyset,X)}$ be the updated probabilities of the underlying state conditional on $e = d$ and $e = \emptyset$. Then, because the jury optimally sets a equal to the posterior probability of the high underlying state, we have $U_{\mathrm{J}}^{*} = -f(1,d,X)\left(b^{*}(d)-1\right)^{2} - f(0,d,X)\left(b^{*}(d)-0\right)^{2} - f(1,\emptyset,X)\left(b^{*}(\emptyset)-1\right)^{2} - f(0,\emptyset,X)\left(b^{*}(\emptyset)-0\right)^{2}$ and $U_{\mathrm{J}}^{0} = -r(r-1)^{2} - (1-r)(r-0)^{2}$.

with the litigant's actual strategy.

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

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

Note that $L \geq 0$. If we replace $\underline{U}_{\text{J}}$ with the jury's expected actual payoff, then L becomes the *actual loss ratio*. We will find that rationalizable outcomes under some conditions lead to values of L strictly greater than 1, which means that hard evidence can be so misleading as to more than offset the potential welfare gain that its face-value signal can provide.[19]

The second key element of the welfare analysis is to 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. Consistent with the notion of robust mechanism design, our objective for robustness is to maximize the minimum rationalizable expected actual payoff of the jury. The optimization exercise is by the law's choice of admissibility rule. In the basic model, an admissibility rule simply declares whether the single document is admissible, and clearly the social objective requires making the document inadmissible if $L > 1$. If it is not admissible, then the litigation game becomes a restricted game in which the litigant can never disclose, and so the litigant has no action. In the two-document model analyzed later in this paper, there are several feasible levels of admissibility.

3 Conditions For Misleading Hard Evidence

Whether hard evidence can turn out to be ineffective or misleading depends critically 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 $b(d)$. Define:

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

Thus, K^+ is the set of litigant x-types for which the *combination* of the occurance of x and existence of the document is positive evidence of $\theta = 1$, whereas K^- is the set of litigant x-types for which the combination of the occurance of the type and existence of the document

[19]In an equilibrium of our litigation game, the actual loss ratio would always be in the interval $[0, 1]$. Incidentally, 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.

is positive evidence of $\theta = 0$. Note that $X = K^+ \cup K^-$. The key conditions compare these combination signals to the face-value signal of the hard evidence d.

Lemma 1: *Let $b(d)$ be given by Equation 1. There is a function $\lambda : X \to [0, 1]$ satisfying $\lambda(x) \geq \psi$ for every $x \in X$, and such that $b(d) < r$, if and only if*

$$\psi\,[f(d, X\,|\,1) - f(d, X\,|\,0)] < \left[f(d, K^-\,|\,0) - f(d, K^-\,|\,1)\right]. \tag{3}$$

Likewise, there is a function $\lambda : X \to [0, 1]$ satisfying $\lambda(x) \geq \psi$ for every $x \in X$, and such that $b(d) > r$, if and only if

$$\psi\,[f(d, X\,|\,1) - f(d, X\,|\,0)] > \left[f(d, K^+\,|\,0) - f(d, K^+\,|\,1)\right]. \tag{4}$$

Proof: We write $b(d) < r$ using Equation 1 to substitute for $b(d)$, and then rewrite the expression using $f(d, x\,|\,\theta)\lambda(x) = f(d, x\,|\,\theta)\psi + f(d, x\,|\,\theta)(\lambda(x) - \psi)$. Factoring and dividing by $(1 - r)$, we obtain the following inequality:

$$\psi\,[f(d, X\,|\,1) - f(d, X\,|\,0)] < \sum_{x \in X} [f(d, x\,|\,0) - f(d, x\,|\,1)]\,(\lambda(x) - \psi)\,. \tag{5}$$

Because $\lambda(x) \geq \psi$ is required, the right side is maximized by setting $\lambda(x) = 1$ for all $x \in K^-$ and $\lambda(x) = \psi$ for all $x \in K^+$, which yields the conclusion for Inequality 3. In words, the jury believes that the litigant would disclose the document for sure if $x \in K^-$ and otherwise would disclose the document with just the minimal probability ψ. This leads the jury to update the state downward following disclosure, and so $b(d) < r$. A further implication is that $b(\emptyset) \geq r$, whatever is the jury's initial belief about the litigant's strategy.[20]

The relation $b(d) > r$ is equivalent to the reverse of Inequality 5, and so we want to minimize the right side. This is accomplished by setting $\lambda(x) = 1$ for all $x \in K^+$ and $\lambda(x) = \psi$ for all $x \in K^-$, which yields the conclusion for Inequality 4. Here the jury believes that x-types in K^+ would disclose for sure and the others would disclose only minimally, so that disclosure induces the jury to update downward. □

If $f(d, X \mid 1) < f(d, X \mid 0)$ then Inequality 3 trivially holds. This is the case in which the document is positive evidence of the low state overall, so the face-value signal of hard

[20] 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, and $b(d) < r$ implies $b(\emptyset) > r$ by the law of iterated expectations. If the jury's initial belief satisfies $\lambda(x) = 0$ for all x, then $b(\emptyset) = r$ because non-disclosure conveys no information.

evidence favors the low state. If the face-value signal of hard evidence favors the high state, then Inequality 3 requires K^- to be nonempty, so that there are x-types that, in combination with existence of the document, provide positive evidence of the low state. Further, the magnitude of this positive evidence of the low state is required to exceed the magnitude of the document's face-value signal as positive evidence of the high state, by an amount that is high enough in relation to ψ. Likewise, if $f(d, X \mid 1) > f(d, X \mid 0)$ then Inequality 4 trivially holds. Otherwise, Inequality 4 requires x-types that, in combination with the document existing, provide sufficiently strong positive evidence of the high state.

It is not difficult to verify that Inequalities 3 and 4 are both satisfied if the litigant's private signal provides a sufficiently accurate indication of the underlying state, if ψ is sufficiently small, and if the document exists with strictly positive probability in both underlying states. 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 $f(d, K^- \mid 1) = 0$. Then Inequalities 3 and 4 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 precisely when Inequalities 3 and 4 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. Realistically, different x-types may have different beliefs about the jury's reasoning and interpretation. Consider, for example, the setting in which the litigant is a defendant who knows his culpability and suppose that a guilty litigant (with $x \in K^-$) believes the jury will look favorably on disclosure of the document, so that this x-type's beliefs satisfy $\bar{b}(d) > \bar{b}(\emptyset)$. Suppose that the beliefs of an innocent litigant (with $x \in K^+$) are opposite: $\bar{b}(d) < \bar{b}(\emptyset)$. 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 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 then leads to the conditions for misleading or ineffective hard evidence, which are summarized as follows.

Theorem 2: *If Inequality 3 holds and Inequality 4 is reversed, 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 Inequality 3 is reversed and Inequality 4 holds, 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. If Inequalities 3 and 4 are both satisfied, then there are rationalizable outcomes in which hard evidence is misleading and the potential loss ratio is $L > 1$.*

We call the second case, where the actual loss ratio is $L = 0$ and hard evidence is effective, the condition of *robust litigation*, because it is in this case that rational behavior always leads to effective hard evidence. Incidentally, although rationalizability assumes common knowledge of rational behavior, our result follows from only that players are rational (best responding to their beliefs) and know this about each other.

To interpret the third case described in Theorem 2, it is useful to consider equilibrium outcomes as a benchmark. We can divide the third case into two subcases on the basis of whether $f(d, X \mid 1) \leq f(d, X \mid 0)$. If this inequality holds, then every perfect Bayesian equilibrium of the litigation game is uninformative, meaning that hard evidence is completely ineffective. If $f(d, X \mid 1) > f(d, X \mid 0)$, then there are multiple perfect Bayesian equilibria, including ones in which the document is never disclosed (so hard evidence is completely ineffective) and one in which the document is always disclosed (so hard evidence is effective). It is easy to verify that the loss level in any equilibrium is between 0 and 1. But in both subcases, there are rationalizable outcomes in which hard evidence is misleading and the loss level strictly exceeds 1, so welfare falls strictly below what could be achieved in any equilibrium.

The equilibrium benchmark also is helpful in motivating the rationalizability concept and, in particular, the prospect of miscoordinated beliefs and behavior. Suppose that $f(d, X \mid 1) > f(d, X \mid 0)$. 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.

To see how the prospect of misleading evidence depends on the relation between the informativeness of the litigant's private signal and the face-value signal of hard evidence, consider the case in which e and x are conditionally independent and $\psi = 0$. 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 signal is x conditional on θ. Thus $f(d, x \,|\, \theta) = q_\theta p_\theta(x)$ and $q_\theta = f(d, X \,|\, \theta)$. Define:

$$\overline{\gamma} \equiv \max_{x \in X} \frac{p_0(x)}{p_1(x)} \text{ and } \underline{\gamma} \equiv \min_{x \in X} \frac{p_0(x)}{p_1(x)}.$$

These are the maximum and minimum of the likelihood ratio (low state to high state) over the private signal values of the litigant. They provide bounds on the highest and lowest posterior probability that the litigant can have about the low state based only on his private signal. Because $p_0(X) = p_1(X) = 1$, we know that $\underline{\gamma} \leq 1 \leq \overline{\gamma}$.

For the conditionally independent case, we can state our previous results as follows.

Corollary 1: *Suppose that $\psi = 0$ and hard evidence e and private signal x are independent conditional on the underlying state θ. If $q_1/q_0 < \underline{\gamma}$, then the unique rationalizable outcome has the litigant never disclosing d, the actual loss ratio is $L = 1$, and hard evidence is completely ineffective. If $\overline{\gamma} < q_1/q_0$ 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. If $\underline{\gamma} \leq q_1/q_0 \leq \overline{\gamma}$ then there are rationalizable outcomes in which hard evidence is misleading and the potential loss ratio is $L > 1$.*

Thus, 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 (which the litigant prefers), then hard evidence is effective. 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 and the prospect of the welfare loss exceeding the potential welfare gain of the face-value signal.

4 Implications for Courts

Are the prospects for misleading or useless evidence a problem for the legal system to solve? Is robustness an appreciable problem? To both questions, our answer is affirmative. Actual litigants and juries typically arrive in court with little experience with similar settings and with each other, and so, lacking an institution that would coordinate them, there is little reason to expect that their beliefs and behavior would be aligned.[21] Even if some measure of coordination could be achieved, theory does not provide much hope that the parties will

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

coordinate on society's preferred outcome. For instance, standard equilibrium refinements do not restrict the equilibrium much less select the most informative equilibrium.[22]

Viewing the robustness problem as something the legal system ought to help solve, three remedies come to mind and all have some bearing on reality. The first is for the court to be transparent regarding its interpretive rules, meaning that the court should strive to articulate 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.[23] 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.

The second remedy is for the court to impose rules that reduce the litigant's discretion regarding whether to disclose hard evidence. For example, the court may penalize a litigant who failed to disclose evidence that was later found to exist, as is the case with prosecutors and the *Brady* rule in the United States. In terms of the model, this would have the effect of raising the parameter ψ which, from Lemma 1 and Theorem 2, reduces the range of other parameters under which evidence may be misleading in the case of positive evidence of the high state.

The third remedy is a more familiar and forceful instrument: *admissibility rules*. The law can provide simple rules that identify conditions under which the court should allow evidence to be admitted. The rules must be in a form that the court can readily apply, so they should be a function of the parameters of the evidentiary system. They state whether a given document (to use our model's terminology) is admissible, which can be represented in our model as whether the litigant is allowed to disclose the document.[24] The determination is made on the basis of relevance, which refers to the likelihood parameters in our model.

Let us therefore assume that the policy instruments available to the framers of the law are admissibility rules and that, otherwise, the court does not interfere with the jury's

[22]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 slightly 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.

[23]Common legal representation for various litigant x-types—as would be the case with specialist attorneys—may serve to align the beliefs of the various x-types.

[24]Admissibility rules may operate in two ways. The court could simply refuse to allow certain documents to be presented. Alternatively, the court could declare that the jury treat specific disclosures (or lack thereof) as though providing no information whatsoever.

Bayesian analysis. We look for admissibility rules that represent society's interest in avoiding misleading evidence and associated welfare loss. Suppose that the court seeks robustness in the litigation process, which we have defined to mean maximizing the minimal social welfare over all rationalizable outcomes of our litigation game. Then the court optimally makes the document d inadmissible if and only if $L > 1$, because in this case the potential welfare loss exceeds the gain of the document's face-value signal.

Policy Implication 1: *To ensure robust use of evidence in litigation, the court sometimes must make some hard evidence inadmissible, in particular when the potential welfare loss outweighs the face-value signal.*

Because misleading evidence is the product of miscoordinated beliefs, and one would not expect the parties' to always be miscoordinated, society would generally tolerate some level of potential loss above hard evidence's face value. Effective use of evidence in the litigation process relates to lowering L, so it is worth determining 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: *Suppose that hard evidence e and private signal x are independent conditional on the underlying state θ, $\psi = 0$, and $\underline{\gamma} < \overline{\gamma}$. 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_1 and q_2 satisfying $\underline{\gamma} \leq q_1/q_0 \leq \overline{\gamma}$, the potential loss ratio satisfies $L \geq \underline{L}$.*

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:

$$b(e') = \frac{\displaystyle\sum_{x \in X,\, e \supset e'} r f(e, x \,|\, 1) \lambda_{e'}(x, e)}{\displaystyle\sum_{x \in X,\, e \supset e'} r f(e, x \,|\, 1) \lambda_{e'}(x, e) + \sum_{x \in X,\, e \supset e'} (1 - r) f(e, x \,|\, 0) \lambda_{e'}(x, e)}. \tag{6}$$

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 6 under the assumption that $\lambda = \sigma$, 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 $U^*_{\text{J}}(\delta)$ 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.[25] Let $\underline{U}_{\text{J}}(\delta)$ 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 $U^*_{\text{J}}(\delta) - U^*_{\text{J}}(\delta')$ 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 $U^*_{\text{J}}(\delta) - \underline{U}_{\text{J}}(\delta)$ and the loss possible under δ' is $U^*_{\text{J}}(\delta') - \underline{U}_{\text{J}}(\delta')$, 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

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

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.

5.1 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 $\delta^{\text{Full}} = \{\emptyset, \{d_1\}, \{d_2\}, \{d_1 d_2\}\}$ and let $\delta^0 = \{\emptyset\}$. The definition of $\overline{\gamma}$ is unchanged from the basic model.

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

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

[25] 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 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

$$\frac{q_{11}q_{21}}{q_{10}q_{20}} > \overline{\gamma} > \frac{q_{11}(1-q_{21})}{q_{10}(1-q_{20})} \quad \text{and} \quad \frac{q_{11}q_{21}}{q_{10}q_{20}} > \overline{\gamma} > \frac{q_{21}(1-q_{11})}{q_{20}(1-q_{10})}. \tag{7}$$

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 7 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 equilibria of the game under policy δ^{Full}. There are an infinite number of sequential 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}).

5.2 Implications for Courts

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

Policy Implication 2: *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. McCormick (1972) notes that the judge has the discretion to "ask the proponent what additional circumstances he expects to prove."

6 Legal Rules

We turn next to discuss, in light of our theory, two specific legal rules in the United States: Federal Rules of Evidence 403 and 404. We include brief accounts of two prominent legal cases and associated numerical examples. We also comment on some related work on exclusion in the literature.

6.1 Relevance and the Law

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.[26] The first three are matters for the judge. The fourth is for the jury or other trier of fact.

Consider first the issue of relevance. In reality, most pieces of evidence presented in support of a case are available in different states with different probabilities and thus are statistical in nature. The legal terms *probative* (providing proof regarding a claim) and *relevance* (tending to strengthen the particular claim being assessed) suggest that the interpretation of evidence is often a matter of establishing a degree of confidence rather than reaching a conclusion with certainty.[27] Rule 401 of the Federal Rules of Evidence 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."[28]

So clearly the law embraces the notion that evidence is statistical, and relevance is the determination of whether the evidence provides a signal of the claim being evaluated. We note that the law also appreciates incremental evidence—which we might call *marginally relevant*—at least to the extent that several such pieces can be combined to make a stronger signal.[29]

On the issue of weighing the evidence (Thayer's fourth item), note that although the judge can provide guidance to the jury, the jury weighs the evidence and this makes the issue of robustness and the potential for evidence to be misleading critical. Thayer described the

[26] Modern reform of the law of evidence relied heavily on the ideas of James Bradley Thayer, many of which are contained in Thayer's *A Preliminary Treatise on Evidence at the Common Law* (1898). See, for example, Anderson, Schum, and Twining (2005), Swift (2000), and Twining (1994). The discussion here draws from Anderson, Schum, and Twining (2005), which has a nice presentation of these issues.

[27] This is not to say that near certainty is never achieved. One such actual case involved a person named Juan Catalan being accused of shooting and killing 16 year-old Martha Puebla, who was to testify in a murder case against Catalan's brother and another Vineland Boyz gang member, Jose Ledesma. For details of the case, see Rubin and Bloomkatz (2008). The charges against Catalan were eventually dismissed when his attorney was able to acquire video showing him at a Los Angeles Dodgers game at the time of the shooting. In this case, Puebla was shot in front of her home, and her parents and neighbors heard the gunshots, so witnesses could identify the exact time of the crime. The video evidence pinpointed Catalan in the stadium at exactly the same time.

[28] Further, the Notes of Advisory Committee on Proposed Rules for Rule 401 include the following: "The rule summarizes this relationship as a 'tendency to make the existence' of the fact to be proved 'more probable or less probable.. . . ' The standard of probability under the rule is 'more probable than it would be without the evidence.' Any more stringent requirement is unworkable and unrealistic."

[29] In United States v. Pugliese (2d 1946), Pugliese was accused of possessing unstamped distilled spirits. Judge Learned Hand, when considering the relevancy of testimony by a former tenant of Pugliese that Pugliese had previously possessed unstamped spirits in the same location, said: "Its relevancy did not, and indeed could not, demand that it be conclusive; most convictions result from the cumulation of bits of proof which, taken singly, would not be enough in the mind of a fair minded person. All that is necessary, and all that is possible, is that each bit may have enough rational connection with the issue to be considered a factor contributing to an answer." (Wigmore §12)

analysis of evidence as being governed by "logic and experience."[30]

6.2 Exclusion – Rule 403

Thayer focused on two principles regarding admissibility: a) evidence that "is not logically probative of some matter requiring to be proved" should not be received, and b) all probative evidence should be allowed "unless a clear ground of policy of law excludes it."[31] He argued for judicial discretion for admissibility.[32] 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.[33]

In practice, admissibility may depend on the degree of relevance weighed against the potential for misleading the jury. Rule 403 of the Federal Rules of Evidence states the following:

> 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.

Further, 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."[34]

Our modeling exercise addresses why it may be useful to exclude evidence that has clear probative value and relevance. Exclusion is meant to avoid evidence being misleading, which can occur when the degree of relevance is low compared to the strength of the litigant's

[30] 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." We suggest that these views of evidence fit with our model.

[31] See Thayer, pages 206 and 530.

[32] Swift (2000) describes this as follows: "First, the exercise of judicial discretion in admitting or excluding evidence is, to use Professor Jon R. Waltz's term, 'guided' decision making. The judge's choice to admit or exclude is not unrestrained by fixed principles. Rather, the judge has some flexibility in choice of outcome, but is restrained by standards articulated in the rules of evidence law. These standards may come from a code or from appellate rulings, and the judge must apply them on the record."

[33] See Thayer, page 516. Further, in the case of United States v. Pugliese (2d 1946) mentioned earlier, Judge Hand writes: "It is true that evidence rationally probative ought sometimes to be rejected if it is likely unduly to complicate the issues, and prolong the trial. . . ."

[34] Swift describes Rule 403 as "the primary example of guided discretion in modem evidence law." It is worth noting that Swift is concerned with judges being given too much discretion with little scope for review by an appellate court. These focus on judicial decisions about admissibility of expert witness testimony, scientific proof, and character evidence.

private information. Thus, we provide a rationale for Thayer's recommendation to exclude evidence that is slightly relevant and may confuse the jury. Further, our results may provide some guidance for judicial discretion and the balancing of relevance against the potential for misleading the jury. We note that one can view the potential loss ratio $L > 1$ as corresponding to the "substantially outweighs" description of Rule 403. Interestingly, it seems, in practice, misleading evidence is viewed as less of an issue for bench trials. This may fit with the coordination issue we explore since coordination between a judge and litigant may be easier. See Capra (2001).[35]

As an example, 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.

Numerical Example

We construct a stylized example based on this case. This is done in the context of the events that potentially lead to the relevant likelihood ratios. The time line below describes the timing of events.

1. The following exogenous random draws occur at the time of a possible prior accident:
   - The rug is loose (ℓ) with probability $\frac{1}{2}$ and not loose (n) with probability $\frac{1}{2}$.

[35] In Gulf States Utilities Company v. Ecodyne Corporation, 635 F.2d 517 (5th Cir. 1981), the Fifth Circuit said:

> Excluding relevant evidence in a bench trial because it is cumulative or a waste of time is clearly a proper exercise of the judge's power, but excluding relevant evidence on the basis of "unfair prejudice" is a useless procedure. Rule 403 assumes a trial judge is able to discern and weigh the improper inferences that a jury might draw from certain evidence, and then balance those improprieties against probative value and necessity. Certainly, in a bench trial, the same judge can also exclude those improper inferences from his mind in reaching a decision.

- The "prior accident" (a) occurs with probability $\frac{1}{2}$ if the carpet is loose and with probability $\frac{1}{8}$ 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.[36]

   When aware of the loose carpet, a "good" owner, occurring with probability $\frac{3}{4}$, repairs the loose carpet. A "bad" owner does not repair the loose carpet.[37] 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 $\frac{1}{2}$. If N, an accident occurs with probability $\frac{1}{8}$.

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 observation of the carpet just before or after the accident, and the observation occurs with noise. Assume that conditional on N, $\overline{x}$ is realized with probability $\frac{1}{4}$ and $\underline{x}$ is realized with probability $\frac{3}{4}$. Conditional on L, $\overline{x}$ is realized with probability w and $\underline{x}$ is realized 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. Recall that the litigant is the patron, as in the Robitaille case. 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

[36]We assume this for simplicity. A richer example that has the good owner more likely to become aware of the carpet being loose has similar qualitative results.

[37]One can motivate the two x-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.

that we assume that in order for the case to go to trial, there must have been an accident.[38]

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

$$f(d, \overline{x} \mid 1) = w, \quad f(d, \overline{x} \mid 0) = \frac{1}{36},$$

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

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.[39]

Note that if $\psi = 0$ then Inequality 3 reduces to $f(d, K^0 \mid 0) > f(d, K^0 \mid 1)$ and $f(d, K^1 \mid 0) < f(d, K^1 \mid 1)$. Because there are just two x-types of litigant in this example, the key question is whether either $f(d, \overline{x} \mid 0) > f(d, \overline{x} \mid 1)$ and $f(d, \underline{x} \mid 0) < f(d, \underline{x} \mid 1)$ or $f(d, \overline{x} \mid 0) < f(d, \overline{x} \mid 1)$ and $f(d, \underline{x} \mid 0) > f(d, \underline{x} \mid 1)$, for in both of these cases hard evidence can be misleading.

Consider first the case in which $f(d, \overline{x} \mid 1) = w > f(d, \overline{x} \mid 0) = \frac{1}{36}$ and $f(d, \underline{x} \mid 1) = 1 - w < f(d, \underline{x} \mid 0) = \frac{3}{36}$. This requires $w > \frac{33}{36}$. 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 $f(d, \overline{x} \mid 1) = w < f(d, \overline{x} \mid 0) = \frac{1}{36}$ and $f(d, \underline{x} \mid 1) = 1 - w > f(d, \underline{x} \mid 0) = \frac{3}{36}$. This requires $w < \frac{1}{36}$. Thus, the conditions for evidence to be misleading are also satisfied for small values of w, which again means that the patron has accurate private information about the carpet (with L now being indicated by $\underline{x}$). 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.[40]

To relate these observations to the set of equilibria, note there are two perfect Bayesian equilibria (PBE): an uninformative one in which d is never disclosed and an informative one in which the litigant discloses d whenever he possesses it. Consider the case with $w > \frac{33}{36}$ so $\overline{x}$ suggests the carpet was loose at the time of the litigant's accident. The uninformative

[38]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.

[39]Note that since the probabilities with which $\underline{x}$ and $\overline{x}$ are realized is fixed following N, the informativeness of the private signal only depends on w.

[40]We note that these may be more extreme than in practice since we hold the probability of $\underline{x}$ fixed and consider only w, the probability of $\overline{x}$.

equilibrium relies on if the document is disclosed, the jury believes it was x-type $\underline{x}$ who disclosed it. Suppose that the litigant of x-type $\overline{x}$ anticipates play of the uninformative PBE and does not disclose d, the x-type $\underline{x}$ litigant anticipates play of the informative PBE and discloses d when she possesses it, and the fact-finder anticipates play of the informative PBE and updates accordingly. This results in $b(d) > r$, which is exactly the opposite of what is implied by the litigant's behavior.

6.3 Character Evidence – Rule 404

There is a general reluctance to allow character evidence or evidence of a prior conviction about a defendant in a criminal case. Rule 404 states "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." Here, we do not explore the more nuanced issues concerning a criminal defendant who chooses to testify at his trial.[41] 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 focus on evidence of a prior conviction presented in that case.

Harvey Lynn Beagle II appealed his conviction by a jury of attempted arson and arson.[42] 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.[43] The general idea is that a judge should have discretion to exclude some

[41] There is a bit more judicial discretion when the defendant testifies as a witness since there is scope for using character evidence or evidence of prior convictions to impeach a witness. These are important details and we think the general theme fits with our results. There are potentially different rules for a civil case.

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

[43] 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.)"

prior convictions from being admitted as evidence.[44]

Numerical Example

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 $\frac{1}{2}$ and n occurs with probability $\frac{1}{2}$.

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 $\frac{1}{2}$ so that b also occurs with probability $\frac{1}{2}$.

3. The defendant is matched with a bar/situation. The type with no prior conviction commits the crime with probability $\frac{1}{5}$, type g with probability $\frac{1}{10}$, and type b with probability $\frac{1}{2}$.[45] If the defendant does not commit the crime, with probability $\frac{1}{5}$ someone else does.

4. Following commission of the crime and the defendant's arrest, the prosecutor's private information x is realized as follows.[46] If the defendant has no prior conviction, $x = a$. For type g, $\overline{x}$ is realized with probability $\frac{1}{4}$ and $\underline{x}$ is realized with probability $\frac{3}{4}$. For type b, $\overline{x}$ is realized with probability w and $\underline{x}$ is realized with probability $1 - w$.

[44] 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.

[45] 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.

[46] The prosecutor may be better informed than the fact-finder regarding the defendant's type. The informativeness of $\overline{x}$ and $\underline{x}$ about the litigant's type depends on w.

Our model picks up the story at Date 4. Since we're 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 someone.

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

$$f(d, \overline{x} \mid 1) = \frac{1}{40} + \frac{w}{2}, \quad f(d, \overline{x} \mid 0) = \frac{3}{40} + \frac{w}{6},$$

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

As in the previous numerical example, there are two possibilities for evidence to be misleading here. Naturally, this depends on how informative the litigant's private signal is. Here, this depends on the value of w.[47]

Consider first the case in which $f(d, \overline{x} \mid 1) = \frac{1}{40} + \frac{w}{2} > f(d, \overline{x} \mid 0) = \frac{3}{40} + \frac{w}{6}$ and $f(d, \underline{x} \mid 1) = \frac{3}{40} + \frac{1}{2}[1 - w] < f(d, \underline{x} \mid 0) = \frac{9}{40} + \frac{1}{6}[1 - w]$. This requires $w > \frac{11}{20}$. 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 has $f(d, \overline{x} \mid 1) = \frac{1}{40} + \frac{w}{2} < f(d, \overline{x} \mid 0) = \frac{3}{40} + \frac{w}{6}$ and $f(d, \underline{x} \mid 1) = \frac{3}{40} + \frac{1}{2}[1 - w] > f(d, \underline{x} \mid 0) = \frac{9}{40} + \frac{1}{6}[1 - w]$. This requires $w < \frac{3}{20}$. So the conditions for evidence to be misleading are satisfied also for small values of w, which again means that the prosecutor has accurate private information about whether the previously convicted defendant has reformed and $\underline{x}$ suggest b. 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. The failure to coordinate here is similar to that in the previous example.

6.4 Related Work on Exclusion

The idea that exclusion of evidence can improve fact-finding is unique to common law systems.[48] Many economic models of the trial process, which typically assume that the fact-finder is able to process information without limitations, have difficulty explaining the exclusion of relevant evidence. We note that our results concerning exclusion fit very closely with those found in the Federal Rules of Evidence and the common law.

[47] Note that since the probabilities with which $\underline{x}$ and $\overline{x}$ are realized is fixed following N, the informativeness of the private signal only depends on w.

[48] See, for example, Damaska (1997).

Lester, Persico, and Visshers (2009) provide a clever explanation for the exclusion of evidence that focuses on cognitive limitations of the fact-finder. In their model, it is costly for the fact-finder to evaluate evidence. This potentially results in the fact-finder's incentives in evaluating evidence not being aligned with those of society in terms of accuracy of decision making. Thus, there is scope for the judge making some relevant evidence inadmissible to improve the decisions made by the fact-finder by preventing some socially non-optimal evidence evaluation by the fact-finder. We do not dispute this line of research and we accept that real fact-finders have cognitive limitations, not withstanding our rationality assumption here. However, Lester, Persico, and Visshers speak of their own work 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," and we believe our approach offers a foundation for general principles.

Two other related papers that also make insightful contributions to the question of exclusion are Sanchirico (2001), and Schrag and Scotchmer (1994). Sanchirico suggests that a potential wrongdoer's choice of action does not influence character evidence so character evidence should not be used to provide incentives.[49] Schrag and Scotchmer make a related argument. While we find this argument compelling, we note that it does not explain all of the issues related to exclusion. Lester, Persico, and Visshers (2009) note the following:

> This argument relies on the predictability of exclusion on the part of the potential wrong-doer. A salient feature of Rule 403 in the US Federal Rules of Evidence, in contrast, is the latitude given the judge to exclude evidence on a case-by-case basis. That latitude seems to run counter the incentive-giving argument because it makes it difficult for the potential wrongdoer to foresee what evidence might be excluded.

7 Conclusion

We have developed a model of statistical evidence with a sophisticated Bayesian fact-finder, applying the rationalizability concept to study robustness of reasoning about hard evidence and fact-finder decisions. Our model identifies the two channels of information inherent in evidence disclosure: The direct implications of the hard evidence disclosure, which we call the face-value signal, and a signal of the litigant's private information that relates to the

[49]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.

litigant's strategy. The theoretical exercise provides a simple framework to explore how these two channels of information interact. It yields conditions under which the signal of private information can outweigh the face-value signal, which leads to the possibility that evidence is misleading.

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. Indeed, just to determine whether a particular legal rule applies to an individual case requires analysis and interpretation. In other words, fact-finders 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 fact-finders interpret evidence and make judgments. Realistic instruments in this regard are admissibility rules and standards of *relevance*, which we take to relate to the likelihood that evidence exists in different states. It is the fact-finder's job to evaluate the idiosyncrasies of individual cases, weigh evidence, and conduct a Bayesian analysis.

Our results suggest that it may be optimal, given robustness concerns, to restrict the evidence that a fact-finder may consider, even disallowing some relevant evidence. The model thus provides an explanation for the Federal Rules of Evidence Rules 403 and 404. Notably, the evidence that should be excluded is that which is least relevant or would serve to confuse the fact-finder, which in our framework means would be open to opposing interpretations due to the relative weight of the litigant-type signal. The potential loss ratio provides an interpretation of the scope for misleading evidence to "substantially outweigh" the probative value of the evidence, as specified in Rule 403.

Future research may be useful to explore more deeply the incentives of players at the "primary activity" stage (before litigation) and to pursue applications outside the legal realm, such as in finance and marketing. Another avenue for future research is to model the jury as a group of individual strategic actors rather than a single player as we have in the current paper. One could then study how and whether members of a jury coordinate on interpretation and meaning of evidence—a critical issue given that jurors come from many different backgrounds and typically have not interacted with each other before (and few have much experience as fact-finders).

A Proofs and Calculations

A.1 Proofs

The theorems are restated here, along with notes and proofs.

Theorem 1: *The following holds for every belief system satisfying plain consistency. The jury's posterior belief $b(\emptyset)$ satisfies Equation 2, where λ is the jury's initial belief 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.*

Proof: The proof features similar steps as in Watson (2017), Section 4. First, that $b(\emptyset)$ satisfies Equation 2 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 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:

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

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. Since $W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x) \subset Y(x) \subset Z(x)$, 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 Watson's plain consistency condition 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.[50]

[50] See Watson (2017) for definitions used here.

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

$$\frac{p^{h_{\mathrm{J}}^{d}}(W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x))}{p^{h_{\mathrm{J}}^{d}}(Y_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x))} = \frac{p^{\underline{h}_{\mathrm{J}}}(W_{\mathrm{N}}(x) \times S_{\mathrm{L}} \times S_{\mathrm{J}})}{p^{\underline{h}_{\mathrm{J}}}(Y_{\mathrm{N}}(x) \times S_{\mathrm{L}} \times S_{\mathrm{J}})}, \tag{8}$$

if both denominators are strictly positive.

Note that the terms on the right side of Equality 8 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 8 by $\mu(x)$ and summing, we get

$$\sum_{x \in X} p^{h_{\mathrm{J}}^{d}}(W_{\mathrm{N}}(x) \times Y_{-\mathrm{N}}(x)) = \sum_{x \in X} \frac{f(1,d,x)}{f(\Theta,d,x)} \mu(x). \tag{9}$$

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 9 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 9 becomes

$$b(d) = \sum_{x \in X} f(1,d,x)\pi(x). \tag{10}$$

Since $\sum_{x \in X} \mu(x) = 1$, we have $\sum_{x \in X} f(\Theta,d,x)\pi(x) = 1$ and Equation 10 is equivalent to

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

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

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

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

Theorem 2: *If Inequality 3 holds and Inequality 4 is reversed, 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 Inequality 3 is reversed and Inequality 4 holds, 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. If Inequalities 3 and 4 are both satisfied, then there are rationalizable outcomes in which hard evidence is misleading and the potential loss ratio is $L > 1$.*

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 Inequality 3 holds and Inequality 4 is reversed. Then from Lemma 1 and Theorem 1 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, and so 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 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 no disclosure. Hard evidence provides its face value scaled down by the parameter ψ, and so $L \in (0,1]$.

The same steps establish that if Inequality 3 is reversed and Inequality 4 holds, 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. Hard evidence provides exactly its face value and $L = 0$.

Finally, consider the case in which Inequalities 3 and 4 both hold. Define beliefs λ^+ and λ^- as follows:

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

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^-. From the proof of Lemma 1, we know that $b^+(d) > b^+(\emptyset)$ and $b^-(d) < b^-(\emptyset)$. Thus, if the jury's initial belief is λ^+ then the jury's optimal strategy satisfies $a(d) > a(\emptyset)$ and if the jury's initial belief is λ^- then the optimal strategy satisfies $a(d) < a(\emptyset)$. 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.

Further, there is a rationalizable outcome that entails misleading hard evidence and a loss greater than 1. Suppose the litigant's actual strategy is σ defined by $\sigma(x) = 1$ for $x \in K^+$ and $\sigma(x) = \psi$ for $x \in K^-$. Let ϕ be the probability that d is disclosed when the litigant behaves according to σ. Then $r = \phi b^\sigma(d) + (1-\phi) b^\sigma(\emptyset)$ holds by the law of iterated expectations, and we have $b^\sigma(d) \geq r \geq b^\sigma(\emptyset)$. Due to quadratic loss utility for the jury, the optimal strategy of the jury (if the jury knew that σ was being played) specifies $a(d) = b^\sigma(d)$ and $a(\emptyset) = b^\sigma(\emptyset)$, 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 actual strategy is the optimal one for initial belief λ^-, so that in fact $a(d) < r \leq a(\emptyset)$. Then 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. Thus, $\underline{U}_\mathrm{J} < U^0_\mathrm{J}$ and so $L > 1$. □

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

Proof: This result follows directly from the analysis used to prove Theorem 2. We can construct the rationalizable outcome described at the end of the proof. Because $\underline{\gamma} <\leq \overline{\gamma}$, the actual loss L is strictly greater than 1 for all q_0 and q_1 such that $\underline{\gamma} \leq q_1/q_0 \leq \overline{\gamma}$. 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*

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

Then the only rationalizable outcome entails full disclosure, the actual loss ratio comparing δ^{Full} to δ^0 is $L(\delta^{\mathrm{Full}}, \delta^0) = 0$, 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 6 in the text and simplifying, we find that $b(e') > r$ is equivalent to

$$\sum_{x \in X,\ e \supset e'} [f(e, x \,|\, 1) - f(e, x \,|\, 0)]\, \lambda_{e'}(x, e) > 0. \tag{12}$$

For $e' = d_1d_2$, this becomes

$$\sum_{x \in X} [q_{11}q_{21}p_1(x) - q_{10}q_{20}p_0(x)]\, \lambda_{d_1d_2}(x, d_1d_2) > 0. \tag{13}$$

Recall that we assume $q_{11} > q_{10}$ and $q_{21} > q_{20}$. Along with the assumption for the theorem—that Inequalities 11 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 13 holds regardless of $\lambda_{d_1d_2}(\cdot, d_1d_2)$, which implies $b(d_1d_2) > r$.

For $e = d_1$, Inequality 12 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 11 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:

$$b(e') = \left[\frac{\displaystyle\sum_{x \in X,\ e \supset e'} (1-r) f(e, x \,|\, 0) \lambda_{e'}(x, e)}{\displaystyle\sum_{x \in X,\ e \supset e'} r f(e, x \,|\, 1) \lambda_{e'}(x, e)} + 1 \right]^{-1} \tag{14}$$

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 $b(d_1 d_2)$. To ease notation, define $y(x) \equiv \lambda_{d_1}(x, d_1 d_2)$ and $z(x) \equiv \lambda_{d_1 d_2}(x, d_1 d_2)$. Using Equation 14, 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

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

Thus, $b(d_1) \geq b(d_1 d_2)$ is equivalent to

$$\frac{1 - q_{20} + q_{20} \displaystyle\sum_{x \in X} p_0(x) y(x)}{1 - q_{21} + q_{21} \displaystyle\sum_{x \in X} p_1(x) y(x)} \leq \frac{q_{20} \displaystyle\sum_{x \in X} p_0(x) z(x)}{q_{21} \displaystyle\sum_{x \in X} p_1(x) z(x)}. \tag{15}$$

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 15 is bounded below by 1. To demonstrate this, let us write $\beta_0 = \sum_{x \in X} p_0(x) y(x)$ and $\beta_1 = \sum_{x \in X} p_1(x) y(x)$, so the left side can be written as

$$\frac{1 - q_{20} + q_{20} \beta_0}{1 - q_{21} + q_{21} \beta_1} = \frac{1 - q_{20}(1 - \beta_0)}{1 - q_{21}(1 - \beta_1)}. \tag{16}$$

Suppose $\beta_1 < 1$. Since β_1 and β_0 are total probabilities over the same weighted fraction of the space X, there must be an x-type $x' \in X$ such that $p_0(x')/p_1(x') \geq (1-\beta_0)/(1-\beta_1)$. Otherwise we would have a contradiction in summing over the complementary weighted fraction of X. This means that $\overline{\gamma} \geq (1-\beta_0)(1-\beta_1)$. 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 $q_{21}(1-\beta_1) > q_{20}(1-\beta_0)$, and thus the value in Expression 16 is at least 1. In the case of $\beta_1 = 1$, similar reasoning establishes that we must also have $\beta_0 = 1$ and the value of Expression 16 is 1.

In summary, we have shown that the right side of Inequality 15 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 15 cannot hold, and therefore $b(d_1) < b(d_1 d_2)$. The same steps establish that $b(d_2) < b(d_1 d_2)$.

Claim 4: Every strategy of the litigant with the property $\sigma_{d_1 d_2}(x, d_1 d_2) < 1$ 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 Corollary 2), where we now interpret $d_1 d_2$ as a single document. □

Theorem 6: *Suppose that Inequalities 7 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 2 using the description of beliefs and behavior that follow the statement of Theorem 6 in the text. □

A.2 Rule 403 Example 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:

$$f(d, \overline{x}, 1, A) = \frac{w}{32}, \quad f(d, \overline{x}, 0, A) = \frac{3}{512},$$

$$f(d, \underline{x}, 1, A) = \frac{1}{32}[1 - w], \quad f(d, \underline{x}, 0, A) = \frac{9}{512},$$

$$f(\emptyset, \overline{x}, 1, A) = 0, \quad f(\emptyset, \overline{x}, 0, A) = \frac{w}{8} + \frac{1}{64},$$

$$f(\emptyset, \underline{x}, 1, A) = 0, \quad \text{and} \;\; f(\emptyset, \underline{x}, 0, A) = \frac{1}{8}[1 - w] + \frac{3}{64}.$$

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

A.3 Rule 404 Example 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 crime was committed. These are:

$$f(d, a, 1, A) = 0, \quad f(d, a, 0, A) = 0,$$

$$f(d, \overline{x}, 1, A) = \frac{1}{160} + \frac{w}{8}, \quad f(d, \overline{x}, 0, A) = \frac{9}{800} + \frac{w}{40},$$

$$f(d, \underline{x}, 1, A) = \frac{3}{160} + \frac{1}{8}[1 - w], \quad f(d, \underline{x}, 0, A) = \frac{27}{800} + \frac{1}{40}[1 - w],$$

$$f(\emptyset, a, 1, A) = \frac{1}{10}, \quad f(\emptyset, a, 0, A) = \frac{2}{25},$$

$$f(\emptyset, \overline{x}, 1, A) = 0, \quad f(\emptyset, \overline{x}, 0, A) = 0,$$

$$f(\emptyset, \underline{x}, 1, A) = 0, \quad \text{and} \;\; f(\emptyset, \underline{x}, 0, A) = 0.$$

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

References

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

Banks, J. S. and J. Sobel (1987), “Equilibrium Selection in Signaling Games,” *Econometrica* 55 (3): 647–661.

Bull, J. (2008), “Mechanism Design with Moderate Evidence Cost,” *The B.E. Journal of Theoretical Economics* 8 (1) (Contributions), Article 15.

Bull, J. (2012), “Interrogation and Evidence Fabrication,” working paper.

Bull, J. and J. Watson (2004), “Evidence Disclosure and Verifiability,” *Journal of Economic Theory* 118: 1–31.

Bull, J. and J. Watson (2007), “Hard Evidence and Mechanism Design,” *Games and Economic Behavior* 58: 75–93.

Capra, D. J. (2001), “Case Law Divergence from the Federal Rules of Evidence,” 197 F.R.D. 531

Che, Y. and S. Severinov (2017), “Disclosure and Legal Advice,” *American Economic Journal: Microeconomics* 9 (2): 188–225.

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

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

Crawford, Vincent P. and J. Sobel (1982), “Strategic Information Transmission,” *Econometrica*, 50 (6): 1431–1451.

Daughety, A. and J. Reinganum (2000a), “Appealing Judgments.” *The RAND Journal of Economics* 31(3): 502–525.

Daughety, A. and J. Reinganum (2000b), “On the Economics of Trials: Adversarial Process, Evidence, and Equilibrium Bias,” *The Journal of Law, Economics, and Organization* 16 (2): 365–394.

Federal Rules of Evidence.

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

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

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

Hastie, R. (1999), “The role of ‘stories’ in civil jury judgments,” *Michigan Journal of Law Reform*, 32(2), 1–13.

Hedlund, J. (2017), “Bayesian Persuasion by a Privately Informed Sender.” *Journal of Economic Theory* 167: 229–268.

Kartik, N, and O. Tercieux (2012) “Implementation with Evidence,” *Theoretical Economics* 7: 323–355.

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

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

Lester, B., N. Persico, and L. Visshers (2009), "Information Acquisition and the Exclusion of Evidence in Trials," *Journal of Law, Economics, and Organization* 28 (1): 163–182.

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

Mueller, C. B. and L. C. Kirkpatrick, *Evidence*, 5th ed. Wolters and Kluwer (2012).

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

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

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

Rubin, Joel and A. B. Bloomkatz (2008), "Interrogation, then Revenge." *Los Angeles Times*, July 2, 2008.

Sanchirico, C. W. (2010), "Models of Evidence: Survey and Assessment," working paper.

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

Swift, E. (2000), "One Hundred Years of Evidence Law Reform: Thayer's Triumph," *California Law Review* 88: 2437–2476.

Talley, E. (2013), "Law, Economics, and the Burden(s) of Proof," in *Research Handbook on the Economic Analysis of Tort Law*, J. Arlen, ed.

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

Twining, W., *Rethinking Evidence: Exploratory Essays*. Northwestern University Press (1994).

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

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

Watson, J. (2017), "A General, Practicable Definition of Perfect Bayesian Equilibrium," manuscript.