Naive, Biased, yet Bayesian:

Can Juries Interpret Selectively Produced Evidence?*

Luke M. Froeb@

Vanderbilt University

Bruce H. Kobayashi

George Mason University
Abstract

In an idealized model of civil litigation, interested parties incur costs to produce statistical evidence. A subset of this evidence is then presented to a naive decision maker (e.g., a jury). The jury is naive in that it views evidence as a random sample when, in fact, it is selectively produced. In addition to being naive, the jury is also biased by prior beliefs that it carries into the courtroom. In spite of the jury’s naiveté and biasedness, a full information decision is reached as long as both litigants choose to produce evidence. Our results suggest that criticisms of the jury process based on jury bias or the jury’s use of simple or heuristic rules may be overstated, and underscore the potential importance of competitively-produced evidence in legal decision making.
1. Introduction

Jury decision making in civil cases, especially those involving complex or expert testimony, has been criticized by lawyers, judges, statisticians, economists, sociologists, and representatives of just about every other discipline that comes in contact with the legal system. Defenders of the jury system generally focus on its use in criminal trials, where juries are seen as a line of defense against the state’s ability to inappropriately deprive individuals of life and liberty. However, such concerns are viewed as less compelling in civil cases, suggesting to many critics, that the costs imposed by the use of juries on the adjudicative process outweigh the benefits.

Much of the debate has focused on the role of scientific or statistical evidence in the courts, and on the role of experts in the courtroom. Critics suggest that naive and potentially biased lay juries are inadequately prepared to evaluate scientific or statistical evidence. Further, critics argue that these problems are exacerbated by the adversarial presentation of evidence, a process that can yield selectively produced evidence. Such concerns have led many commentators to suggest a movement toward neutral non-adversarial proceedings, and to various proposals for reform, including the elimination of the right to a jury trial in complex civil cases, increasing the power of judges to prevent issues from reaching the jury, the appointment of court appointed experts or special masters, the constitution of court approved science panels, fee shifting, and the reform of the discovery rules to include mandatory disclosure of information.

On the other side of the debate are those who offer a “Hayekian” justification for the adversarial presentation of testimony to the jury. While both juries and the
adversarial system are inferior to an idealized centralized decision maker, such imperfect
decentralized bodies may have inherent advantages over imperfect centralized decision
makers.\textsuperscript{13} Thus, the relevant comparison for normative purposes is a comparison between
the performance of imperfect institutions. While such a comparative analysis is beyond
the scope of this article, we hope to advance the debate by explicitly modeling the
incentives of the parties to produce costly information and by examining decision making
by a naive and potentially biased decision maker faced with selectively produced
evidence.

This article focuses on one aspect of the debate—the ability of juries to
competently evaluate selectively produced statistical evidence.\textsuperscript{14} Previous attempts at
examining this issue have not considered one or more aspects of this problem. Gay,
Grace, Kale and Noe (1989) consider criminal defendant's choice of judge versus jury
trials. In their model, information processing by juries is assumed to be "noisier" than
information processing by judges. Sophisticated judges and/or juries draw inference
from both prior information and the defendant's choice of trial mode. Their analysis does
not explicitly model competition between litigants, nor the problem of evidence
production.\textsuperscript{15}

The lines of research most closely related to ours are Sobel (1985) and Milgrom
and Roberts (1986). Sobel (1985) considers the ability of litigants to withhold evidence in
a game in which it is costly to report pre-existing evidence.\textsuperscript{16} He finds a mixed strategy
equilibrium where a judge maximizes social surplus by “favoring” one of the litigants
with a lower burden of proof. His judge is a benevolent, sophisticated decision maker.
In this respect our model is more like that of Milgrom and Roberts (1986) who consider both naive decision making and the selective reporting of evidence. In their model, evidence not reported by one party because it is unfavorable, will be reported by the opposing party, and vice-versa. In equilibrium, all relevant information is reported, suggesting the following equivalence result: an unsophisticated decision maker faced with selectively produced evidence will reach a full information decision, as long as the litigants’ interests are sufficiently opposed. In their model, it is the competing incentives of the parties, rather than the sophistication of the decision maker, that permits a full information decision.

However, Milgrom and Roberts’ model ignores two salient features of the legal system, the cost of producing evidence\(^\text{17}\) and the potential for biased decision makers.\(^\text{18}\) When evidence is costly to produce, and when parties must overcome bias, not all relevant information will be reported to the decision maker. Instead, the decision of how much evidence to produce depends on the costs, as well as the benefits, of producing evidence.

In this paper, we develop a model that allows for costly evidence production and decision maker bias. Despite the added complications, we find that a naive decision maker can still reach a full information decision. There are two forces at work here: the first is the tendency of the party favored by the underlying distribution to produce more favorable evidence, due to the lower costs of doing so,\(^\text{19}\) the second is the tendency of the parties to “free ride” on the prior beliefs of the decision maker in a way that negates decision maker bias. In equilibrium, competition in the selective production of evidence
leads to a full information decision by a naive and biased decision maker. We find that the original intuition of Milgrom and Roberts about the importance of competition, rather than the sophistication of the decision maker, can be extended to situations in which evidence is costly to produce, and to situations in which the decision maker is biased.

Our results suggest that the competing incentives of litigants may work to mitigate some of the potential costs attributed to the use of juries. Consequently, criticisms of the jury process based on the observation that jurors use simplifying strategies or heuristics when assessing information may overstate the shortcomings of lay juries’ decision making abilities. Likewise, experimental research that does not subject jurors to the competing forces of adverse litigants and jury deliberations may miss the role these forces play in leading to accurate decision making—even by lay decision makers using simple rules. This finding is similar in spirit to those drawn from experiments conducted in market settings, which find that individual biases tend to disappear in market settings and is consistent with the large body of empirical research on jury competence that, in contrast to the predictions of jury critics, points to the general competence of the jury.

2. Naive and Biased Juries:

In this section, we examine the behavior of a jury in the context of an idealized model of civil litigation where parties have failed to reach out-of-court settlement and are preparing for trial. The culpability of the defendant is represented by the mean of a binomial distribution, “p,” and both parties present evidence to a jury drawn from this
distribution. At issue in the trial is the value of $p$. Using the analogy of a coin flipping experiment, let heads be a piece of evidence that is favorable to the plaintiff and tails be evidence that is favorable to the defendant. The parties each report a number of heads and tails to the jury, and the jury must infer the value of $p$ from the evidence presented in court. The total evidence reported is $\{H, T\}$, where $H$ is the number of heads and $T$ is the number of tails reported by the parties. We explicitly assume that evidence can be reported in arbitrarily small units, so that we can treat $\{H, T\}$ as continuous.

There are several ways to draw inference about the value of $p$ (Lindley and Phillips, 1976). Classical inference proceeds from an assumption about the sample space out of which the evidence is drawn. For example, if the jury assumed that each party took a fixed number of flips, and truthfully reported the outcomes, then the probability of getting the outcome $\{H, T\}$ would be $(H:T)p^H(1 - p)^T$, where $(H:T)$ denotes the choose function. The method of moments estimator is derived from this distribution function:

$$\hat{p} = \frac{H}{H+T}. \tag{1}$$

Different assumptions about the sample space lead to different estimators, but some assumptions about the sample space must be made in order to derive an estimator.

Bayesian inference proceeds from different assumptions. The jury does not need to know the sample space out of which the evidence are drawn, only that they form an “exchangeable” sequence (Heath and Sudderth, 1976). Exchangeability means that the sequence of heads and tails is generated “fairly,” by some experimental process. Unlike classical inference, the exact form of the experiment need not be known. The jury uses
evidence to update its prior beliefs about the value of \( p \). The estimator of \( p \) is the mean of the posterior distribution. We assume that the jury has a Beta\((a, b)\) prior over the unknown value of \( p \). The Beta distribution is the conjugate prior\(^{26}\) to the binomial and is a function over the interval \([0,1]\), with mean \(a/(a+b)\), that includes the uniform distribution \(\{a=1, b=1\}\) as a special case. The posterior distribution is a Beta\((a+H, b+T)\) distribution, with posterior mean:

\[
\hat{p} = \frac{a+H}{a+b+H+T}. \tag{2}
\]

Whether it uses a classical or Bayesian approach to infer the value of \( p \), the jury must make some assumption about the process used by the litigants to generate the evidence. In the classical case, it is an assumption about the sample space, and in the Bayesian case, it is the somewhat less restrictive assumption of exchangeability. The jury in our model makes one of these two assumptions, but neither are true. Thus, the jury is “naive” in the sense that it mistakenly believes or assumes that evidence presented in court is a random or exchangeable sequence when in fact it is selectively reported from each of the interested parties.\(^{27}\)

The jury’s mistaken beliefs about the randomness or exchangeability of the data means that neither the classical estimator given by (1) nor the Bayesian estimator given by (2) can be justified on statistical grounds. The two estimators discussed above are neither classical, nor Bayesian, because they are derived from mistaken assumptions.

In addition to being naive, the jury is “biased” by the prior beliefs that it carries into the courtroom. We have characterized the bias by a prior probability distribution
over the unknown value of $p$. Although juries in civil cases are supposed to choose a decision rule according to a “preponderance” of the evidence standard, there are no sanctions for behaving otherwise, and jury bias is a well known phenomenon. We explicitly allow for jury bias through the parameters \{a, b\}. A jury with a prior such that $a > b$ ($a < b$) favors the plaintiff (defendant). Large values of $a$ or $b$ mean that the jury’s prior beliefs are very strong, or that it is less likely to be swayed by evidence presented in court. For example, a jury with \{a=2, b=2\} has the same prior mean as a jury with \{a=1, b=1\}, but is less easily swayed by evidence produced in court. This is seen by examining the variance of the Beta(a,b) prior distribution, $ab/((a+b)^2(1+a+b))$. Using our examples, the variance of the Beta(2,2) is 1/20 while the variance of the Beta(1,1) is 1/12. A smaller prior variance signifies stronger prior beliefs.

The jury behavior described above can be completely characterized by the single estimator given in equation (2), $\hat{p} = (a+H)/(a+b+H+T)$. The “classical” estimator given in equation (1) is a special case of the “Bayesian” estimator, characterized by the parameters \{a=0, b=0\}. The classical estimator can also be motivated as a “split-the-difference” decision rule. The classical estimator does not require that the jury use any type of formal statistical inference--it can be generated by a jury that uses a simple weighting scheme to evaluate the competing claims of the litigants.

3. Competitively-Produced Evidence
In this section, we model the behavior of a defendant and plaintiff who are preparing for trial, facing a naive and biased jury. As mentioned above, we allow the parties to attempt to mislead the jury by selectively reporting favorable evidence. Continuing with the coin flipping analogy, the plaintiff is allowed to report a favorable subset of the flips taken, as is the defendant. As it turns out, neither party has an incentive to report unfavorable flips, i.e. the plaintiff reports only flips that come up heads, and the defendant reports only those that come up tails. This feature comports well with the observation that litigants rarely report unfavorable evidence (e.g. Rubinfeld, 1985). In addition, we assume that evidence is costly to produce, according to the number of flips that each party decides to take.

This assumption of costly evidence production differentiates our approach from that of Milgrom and Roberts (1986), who study litigation under the assumption that parties can costlessly and credibly produce and report evidence. In their model, evidence not reported by one party would be reported by the other, and vice-versa. In equilibrium, all relevant evidence would be reported, i.e. the jury would know the exact value of $p$. In contrast, we consider a different type of evidence, that which is costly to produce, and which cannot be costlessly revealed. In our model, for any party to reveal the exact value of $p$ would require an infinite number of flips, a prohibitively costly undertaking. While Milgrom and Roberts’ assumption is probably best suited to evidence which can be costlessly reported, our assumption is better suited to more costly types of evidence, the archetypal example being that of expert testimony.\textsuperscript{30}
Continuing with the coin flipping analogy, the number of flips is analogous to the effort that it takes to produce favorable evidence. For example, the plaintiff produces evidence of quantity “H” after, on average, $H/p$ flips. Similarly, the defendant produces evidence of quantity “T” after, on average, $T/(1-p)$ flips. We are implicitly assuming that the jury can verify that evidence is drawn from the distribution of interest, but cannot verify the existence of unfavorable, and unreported, evidence.31

One can immediately see that for $p > .5$, when the plaintiff is favored by the distribution, the plaintiff will have to flip fewer times, on average, than the defendant, to produce evidence of equal strength ($H=T$). If, as we postulate, evidence is costly to produce (according to the number of flips) then for $p > .5$, it is more costly, on average, for the defendant to produce evidence of the same strength as evidence produced by the plaintiff. This difference in cost turns out to be important because it induces the party favored by the distribution to produce more favorable evidence than the opposition.

Formally, suppose we interpret $p$ as the defendant’s level of fault under a comparative negligence tort standard. The plaintiff’s award equals $\hat{p} S$, where $\hat{p}$ is the jury’s estimate of the true $p$, and $S$ is the amount of harm, which is assumed to be uncontested. Given the jury’s weighting scheme, the decision of how much evidence to produce becomes an optimal stopping problem for each of the litigants. The parties stop producing evidence when they have “enough” favorable evidence.32 The effort that it takes to produce evidence is a random variable, but we assume that the parties are risk neutral, so they care only about the mean level of effort required to produce favorable
information, and not variance in the level of effort. Under these assumptions, the
litigants choose stopping values, \( \{H^*, T^*\} \), which satisfy:

\[
H^* = \text{argmax} \big[ \hat{p} S - c H/p \big] \text{ such that } H > 0 \quad (3a)
\]

\[
T^* = \text{argmax} \big[ -\hat{p} S - c T/(1-p) \big] \text{ such that } T > 0 \quad (3b)
\]

where \( \hat{p} = (a+H)/(a+b+H+T) \). For \( H > 0 \) and \( T > 0 \), this litigation game has a unique Nash equilibrium:

\[
\{H^*, T^*\} = \{-a+p^2(1-p)S/c, -b+(1-p)^2pS/c\} \quad \text{if } H^* > 0 \text{ and } T^* > 0 \quad (4a)
\]

The notable thing about this equilibrium is that when both plaintiff and defendant produce strictly positive amounts of evidence, the estimator used by the jury is exact, i.e.,

\[
\hat{p}^* = \frac{H^*}{(H^*+T^*)} = p. \quad (5)
\]

Note that there is no variance to the estimator given by (5). There is variance in the number of flips required to produce the estimator, but not in the estimator itself. Thus, when evidence is produced by both sides, the adversarial process and the naive jury produces a costly but accurate (unbiased and zero variance) estimate. We say it is “costly” because the litigants throw information away (unreported flips), but the reported flips are produced in exactly the “right” proportion.

There are two forces at work here. First is the tendency of the party favored by the distribution to produce more favorable evidence, due to the lower costs of doing so. Second, is the tendency of the parties to “free ride” on the prior beliefs of the jury. When the jury has prior beliefs that favor one party, that party produces less evidence, which exactly counteracts the effects of the prior beliefs of the jury. This happens because jury
bias lowers the marginal benefit of flipping, reducing by “a” the number of heads produced by the plaintiff, and by “b” the number of tails produced by the defendant. Effectively, the jury’s bias gives a “head start” to one side—it is as if they were endowed with some free evidence. And in equilibrium, when both parties choose to flip, jury bias disappears.

Note that this “exactness” result depends explicitly on our characterization of the unequal costs of producing evidence as a negative binomial experiment. The litigant favored by the distribution has a lower marginal cost of evidence production, and therefore flips more in equilibrium. The equal marginal costs of flipping, as opposed to the unequal costs of evidence production, can be justified by assuming a competitive market for expert coin flippers. An expert charging a supracompetitive rate would not attract any clients.

To illustrate the equilibrium, we present a numerical example. Suppose that the true $p=2/3$, so that the plaintiff is favored by the distribution, and the jury has a uniform prior over the unknown $p$, i.e. $\{a=1, b=1\}$. Suppose further that the amount at risk is $2,000,000$, and the cost of drawing is $100,000$. In equilibrium the plaintiff reports 1.96 heads, and the defendant reports 0.48 tails. On average, the plaintiff will take 2.94 flips, costing $294,000, and the defendant 1.44 flips, costing $144,000. The jury’s estimate of $p$ is exact, $(1+1.96)/(2+1.96+0.48)=.667$.

When $p$ is close to zero, or one, the marginal benefit of flipping is so small that the party favored by the distribution takes no flips. In this case, the Nash equilibrium takes the following form, depending on who decides not to flip:
\[
\{H^*, T^*\} = \{-a-b+(1/c)(-a c (1-p))^{1/2}, 0\} \quad \text{if } T^*<0 
\]
(4b)

\[
= \{0, -a-b+(b c p S)^{1/2}\} \quad \text{if } H^*<0
\]
(4c)

The stronger the beliefs of the jury, i.e. the larger are \{a, b\}, the fewer flips the respective parties take. A party produces no evidence when the bias of the jury is strong, and when \(p\) is near zero or one. In this case the marginal benefit of flipping is so small, that it is better to not flip at all. When this happens, the opposing litigant decides to flip more because the reaction functions of the litigants are downward sloping. The net effect of these changes is to bias the estimator of the jury towards the prior mean of the jury. Note that although the estimator is biased, it represents an improvement relative to the no evidence case, because it moves the jury’s estimator from the prior mean towards the true \(p\).

The Nash equilibrium is illustrated for all possible \(p\) values in three different cases for \(S=2,000,000\) and \(c=100,000\): Figure 1 \{a=1, b=1\}; Figure 2 \{a=2, b=1\}; and Figure 3 \{a=0, b=0\}. In all graphs, the true value of \(p\) is on the horizontal axis; the top graph plots the number of heads and tails reported by the litigants; and the bottom graph plots the value of the jury’s estimator relative to the true value of \(p\). In Figure 1, the jury has a uniform prior over the unknown value of \(p\). In the top graph of Figure 1, when the value of \(p\) is near zero, only the plaintiff flips, and the jury’s estimator of \(p\) is biased towards the prior mean. The jury’s prior is represented by a dashed horizontal line. The bias is represented in the bottom graph as a deviation from the 45 degree line. As \(p\) increases, when the defendant begins flipping, and the jury’s estimator becomes exact,
i.e. \( \hat{p}^* = \frac{H^*}{(H^*+T^*)} = p \). As \( p \) approaches one, the plaintiff stops flipping, and the jury’s estimator is again biased towards the prior mean.

The effects of jury bias are illustrated in Figure 2, where the jury’s prior favors the plaintiff. Here the plaintiff “free rides” on the jury’s stronger prior, and doesn’t begin flipping until a much higher value of \( p \). This means that there is a larger range over which the jury’s estimator is biased towards the prior mean 2/3, but, as above, when both parties produce evidence, bias disappears.\(^{34}\) The net effect of the adversarial system is to mitigate the prior bias of the jury. This is seen in the bottom graph in figure 2 where the jury’s estimator is closer to the true \( p \) than is the prior mean of the jury. Even though the jury’s estimator is biased, it is a better estimator than the prior mean of the jury.

The “classical” jury is illustrated in Figure 3 \( \{a=0, b=0\} \). In this case, the estimator used by the jury is exact, no matter what the value of \( p \) because there is no region where the parties do not flip.\(^{35}\)

4. Conclusions

The exactness of the jury’s estimator is sensitive to a number of assumptions, implicitly associated with the functional forms used in the model. The payoff functions\(^{36}\), the way bias is modeled, the simultaneous presentation of evidence, the specific distributional assumptions used, and the assumption of constant marginal costs of flipping\(^{37}\) are all critical to the results of the model. Though stylized, our model of litigation captures four salient features of the legal system: competition between the
litigants; selective production of favorable evidence; costly production of evidence; and a naive and biased decision maker. Our results suggest that, in equilibrium, the decision maker is able to overcome these shortcomings and reach a full information decision when both parties choose to produce evidence. This implies that competitively-produced evidence in an adversarial setting may mitigate some of the costs attributed to decision maker bias, and to the use of simplified rules or heuristics to evaluate selectively produced information.
Figure 1

[Graph showing two lines representing Heads and Tails as functions of p, with a legend indicating "Heads" and "Tails".]

[Graph showing another set of lines representing true p, jury's p, and jury's prior as functions of p, with a legend indicating "true p", "jury's p", and "jury's prior".]
References


Appendix: Characterization of Nash Equilibrium

This section characterizes the Nash equilibrium of the litigation game in the case where evidence is continuous, or measured in small enough units to be effectively continuous. In essence, we permit litigants to present “pieces” of evidence, like 2.94 Heads. In this case, the expected payoff vector to the plaintiff and defendant is

\[ \{ \hat{p} S - c \frac{H}{p}, -\hat{p} S - c \frac{T}{1-p} \} . \]

The first derivatives whose roots characterize a Nash equilibrium are:

\[ \begin{cases} \frac{-(c/p) - ((a + H)*S)/(a + b + H + T)^2 + S/(a + b + H + T),}{c/(-1 + p) + ((a + H)*S)/(a + b + H + T)^2} \end{cases} \]  \hspace{1cm} (5)

The roots of these derivatives have an explicit solution, given in equation 4a in the text which we reproduce below.

\[ \begin{cases} H*, T* \end{cases} = \begin{cases} -a + p^2(1-p)S/c, \\ -b + (1-p)^2pS/c \end{cases} \text{ if } H*>0 \text{ and } T*>0 \]  \hspace{1cm} (4a)

That this solution is unique can be found by examining the matrix of second partial derivatives of the payoff vector:

\[ \begin{cases} \{(2S*(b + T))/(a + b + H + T)^3, \\ (S*(a - b + H - T))/(a + b + H + T)^3, \\ (S*(-a + b - H + T))/(a + b + H + T)^3, \\ -2*(a + H)*S)/(a + b + H + T)^3 \} \end{cases} \]  \hspace{1cm} (6)
The determinant of this matrix is equal to \( S^2/(a + b + H + T)^4 \). For \( H > 0, T > 0 \), it is seen that the principle minors alternate in sign implying that the matrix is negative definite, which guarantees that the solution given in (4a) is unique.

When the constraints, \( H > 0 \) or \( T > 0 \), are binding, we compute the Nash equilibria by first computing the reaction functions implied by the derivatives of the payoff vector:

\[
\begin{align*}
-a - b - T + (c*p*S*(b + T))^{1/2}/c, & \quad -a - b - H + (c*(1 - p)*S*(a + H))^{1/2}/c \\
\end{align*}
\]

The reaction function of the non-constrained litigant is positively sloped in the region of interest, so for \( p \)'s near zero, where the plaintiff would like to produce negative amounts of evidence (sell evidence to the experts), but must produce more evidence (zero) this leads the defendant to produce more than he would like in an unconstrained equilibrium. Substituting the constraints into the appropriate reaction functions leads to the constrained equilibrium represented by equations (4b) and (4c) which we reproduce below:

\[
\begin{align*}
\{H**, T**\} &= \{-a-b+(1/c)(-a c(1-p))^{1/2}, 0\} \quad \text{if } T* < 0 \quad (4b) \\
&= \{0, -a-b+(b c p S)^{1/2}\} \quad \text{if } H* < 0 \quad (4c)
\end{align*}
\]
Footnotes

@ Froeb received research support from the Dean’s fund for faculty research at the Owen Graduate School of Management. Kobayashi received research support from the George Mason University Foundation. We wish to acknowledge advice about related work from Cynthia Fobian Wilhams and editorial help from Lisa Granoien. We also wish to acknowledge constructive criticism from Timothy Brennan, Stephen Fienberg, Preston McAfee and Kathryn Spier. The usual disclaimer must be strengthened here to note that some or all of those mentioned disagree with our characterization of the adversarial process.

1 For discussion of the history of the controversy over the merits of using lay juries, see Kalvin and Zeisel (1966) at 4-6, Simon (1975) at 13-18. Among the more outspoken judicial critics of the jury were Chief Justice Warren Burger and Judge Jerome Frank. For example, Frank wrote "While the jury can contribute nothing of value so far as the law is concerned, it has infinite capacity for mischief, for twelve men can easily misunderstand more law in a minute than the judge can explain in an hour." Skidmore v. Baltimore and Ohio R.R., 116 F.2d 54 (1947). Dean Griswold of the Harvard Law argued "The jury trial at best is the apotheosis of the amateur. Why should anyone think that 12 persons brought in from the street, selected in various ways, for their lack of general ability, should have any special capacity for deciding controversies between persons?" See Guinther (1988) at xiv

2 The right to a jury trial is guaranteed by the Sixth Amendment for criminal trials. See U.S. Constitution, Amendment VI. The Seventh Amendment guarantees this right for civil trials where the amount in controversy exceeds twenty dollars. See U.S. Constitution, Amendment VII. For a discussion of the history of jury trials, see Kalvin and Zeisel (1966) at 6-8.

4 In response to perceived problems faced by the courts in assessing scientific evidence, the Federal Judicial Center has undertaken and recently published a manual on Scientific Evidence. See, e.g., The Federal Judicial Center, Reference Manual on Scientific Evidence, at 1-4. The issue of admissibility of expert testimony was recently undertaken by the Supreme Court. See Daubert v. Merrill Dow Pharmaceuticals, 113 S.Ct. 2786 (1993) (Supreme Court decision rejecting the general acceptance test of the Frye standard in favor of the "more liberal" standard under the Rules 702 and 703 of Federal Rules of Civil Procedure. For a discussion of the implications of Daubert, see e.g., Parker (1994) and Cecil and Willging (1994b).

5 For proposals to introduce Bayes theorem into court proceedings to assist the trier of fact, see, e.g., Finkelstein and Farley (1970), Ellman and Kaye (1979), and Wagenaar (1988). For alternative views on this debate, see, e.g., Tribe (1971) (suggesting that statistical evidence will be given too much weight), and Saks and Kidd (1981) (suggesting that juries underuse such evidence). For evidence that supports the latter view see, e.g., Goodman (1992) Faigman and Baglioni (1988) and references cited therein. For experimental evidence suggesting experimental subjects do not conform to a Bayesian rule when integrating probabilities, see, e.g., Grether (1992); Nisbett and Ross (1980); Tversky and Kahneman (1974); the discussion in Smith (1991); and Camerer (1995) for a survey.


7 For arguments in support of a complexity exception to the Seventh Amendment, see Campbell (1980), and Campbell and LePoidevin (1988). For arguments that do not support the existence of such an exception, see, e.g., Arnold (1980). In addition, the Supreme Court has not incorporated the Seventh Amendment as applying to the States. See Colgrove v. Battin, 413 U.S. 149 (1973) (twelve person jury trials are not required at the state level). In addition, the Supreme Court has not incorporated the Seventh Amendment as applying to the States (see Stone et al. (1991) at 784), and has allowed the use of less than twelve person juries in federal
civil trials (see Colgrove v. Battin, 413 U.S. 149 (1973)) and in state criminal trials (see Williams v. Florida 399 U.S. 78 (1970)).

8 These devices include expanded use of summary judgment, summary juries and expanded use of directed verdicts. For a discussion of these issues, see Cecil, Hans and Wiggins (1991) at 736-8.

9 Court appointment of expert witnesses are provided for under Rule 706 of the Federal Rules of Evidence. See F.R.E. Rule 706. Special "reference" masters can be appointed under "exceptional conditions" under Rule 53(b) of the Federal Rules of Civil Procedure. See F.R.C.P. Rule 53(b). However, use of court-appointed experts and non pretrial special masters has been rare. See Cecil and Willging (1994a), Farrell (1994).

10 See, e.g. Meier (1986).

11 While the Supreme Court substantially limited the recoverability of expert witness fees in fee shifting cases in West Virginia University Hospital v. Casey, 111 S. Ct. 1138 (1991) (holding that a provision in 42 U.S.C Section 1988 providing statutory authority for courts to shift "reasonable attorney's fees" to the losing party does not include expert fees), Congress has recently authorized the shifting of expert fees by explicitly including such provisions in statutes. Congress chose to overrule Casey by explicitly authorizing judges to award expert fees in Civil Rights cases. See the Civil Rights Act of 1991, Pub. L. No. 102-166, Section 113(b), 105 Stat. 1071 (1991). 42 U.S.C. Sections 1988, 2000e-5(k)C. Fee shifting of expert fees are also provided in other statutes. For example, the fee shifting provision in the Clean Air Act, which became the model for nearly identical provisions in other environmental statutes, states, "The court . . . may award costs of litigation (including reasonable attorney and expert witness fees) to any party, whenever the court determines such an award is appropriate." See 42 U.S.C. Section 7604(d) (1982).
For an economic analysis of discovery, see Cooter and Rubinfeld (1994), and Schrag (1994). Dissatisfaction with discovery under the adversarial process lead to the recent amendments to Rule 26 of the Federal Rules of Civil Procedure, which controls discovery in the Federal Courts. The most controversial part of this rule was a provision that added a provision requiring early disclosure of both favorable and unfavorable information. See F.R.C.P Rule 26(a)(1). However, the rule allows local district courts to opt-out of the rule. An examination of the district courts' decisions revealed that the majority of the courts have used the local opt-out provision. Thirty-seven districts required no initial disclosures, and nine districts only require disclosure of information favorable to the disclosing party. For a discussion of the courts' decisions and the process of procedural reform, see Kobayashi, Parker and Ribstein (1994).

See, e.g. Hayek, (1945, 1948). See also Demsetz (1969). The fact that jurors decide verdicts in groups after deliberations provides protection against idiosyncratic view or biases. (see Cecil, Hans, and Wiggins (1991) at 749, Lempert (1981)). For an economic model of the effect of jury deliberations, see Kleverick and Rothschild (1979) and Kleverick, Rothschild and Winship (1984). Further, opposing litigants can have more or better information than even a sophisticated decision maker and competition among the litigants forces them to reveal relevant information. See Milgrom and Roberts (1986), Froeb and Kobayashi (1993).

Our analysis does not directly address the ability of juries to deal with complex evidence. For a discussion of this issue, see Cecil, Hans and Wiggins (1991). The problem of understanding overly complex issues is not limited to juries. See the discussion in note 4, supra. While judges are experts in the law, they are not necessarily experts with respect to evaluating statistical or factual questions. For an argument that the Supreme Court has inconsistently adopted and often misused social science research, see Bersoff and Glass (1995). For examples of "judicial missteps" in evaluating statistical evidence, see Finkelstein and Levin (1990). Evidence from private arbitration suggests that arbitrators do not posses either the
characteristics of juries or of public judges. That is, they are neither lay persons nor experts on the law, but rather are chosen based on their familiarity with the subject matter (see Landes and Posner, 1975). For a model of arbitrator behavior, see Ashenfelter and Bloom (1984).

15 Because judge trials are assumed to be more accurate, innocent defendants will choose judge trials while guilty ones will prefer juries as long as judges and juries both act naively. Sophistication on the part of juries cause all defendants to choose judge trials, while an equilibrium where judges are sophisticated and juries are not leads to innocent defendants choosing judges and guilty ones mixing between judges and juries. Gay, et. al. argue that empirical evidence on trial outcomes supports the equilibrium with naive juries and sophisticated judges.

16 In Sobel's model of dispute resolution, the parties choose between incurring a cost to report evidence or not reporting evidence to a third party decision maker. Sobel also examines how shifting the burden of proof affects the incentives of the parties to disclose costly information. However, litigants in his model cannot misrepresent their information.

17 Indeed, the perception that litigation under an adversarial system has resulted in excessive expenditures has been one of the driving forces behind recent attempts at reform, including the recent amendments to the rules governing sanctions and discovery. For a discussion of these amendments, see Kobayashi and Parker (1993), and Kobayashi, Parker, and Ribstein (1994). Other critics of the adversarial system argue that selective evidence production involves social costs, including the costly production of misleading information or the costly production of information that is not used. See Tullock (1975; 1980).

18 For evidence that experimental jurors make decisions based on non-neutral priors, see Goodman (1992). The potential for bias is not limited to individual jurors or juries. For a critique of centralized systems of producing evidence based on bias, see Parker (1994). See also McChesney (1971). Further, asymmetric stakes, and asymmetric burdens of proof and
other procedural rules often introduce bias into the litigation process. For an overview of these issues, see Cooter and Rubinfeld (1989).

19 Our results can be applied to other areas of adversarial litigation where parties expend resources to affect the probability of victory. Litigation expenditures by competing litigants have been examined by Tullock (1980), Katz (1988), and Kobayashi and Lott (1995). These models do not consider the ability of litigants to selectively use information. For models of one sided litigation expenditures, see Rubinfeld and Sappington (1987), and Miceli (1990). Our results are also similar to those found in the literature on asymmetric contests or tournaments. See, e.g., Rosen (1986), and Lazear and Rosen (1981).

20 For examples of experimental design see Grether (1992); Goodman (1992) (incorporating both types of designs); and Faigman and Baglioni (1988). See also Smith (1991).

21 See for example Camerer (1987); Franciosi et. al., (1995); Forsythe et. al. (1991); and Smith (1991) for a survey of the difference between psychology experimental literature with its focus on the individual, and the economics experimental literature, with its focus on market settings.

22 The best known empirical study of jury behavior is Kalvin and Ziesel (1966) (reporting on the results of the Chicago Jury Project). See also Cecil, Hans, and Wiggins (1991) at 745-764, Simon (1975) at 147, Guinther (1988) at 230. These studies generally find that juries are able to comprehend even complex issues, and a high rate of judge/jury agreement. For a recent empirical study of judge/jury agreement, see Claremont and Eisenberg (1992) (finding, contrary to popular belief, that judges in state cases were more favorable to plaintiffs in personal injury cases than juries, and that these results were not attributable to case selection effects). For a recent empirical study of judge/jury agreement, see Claremont and Eisenberg (1992) (finding, contrary to popular belief, that judges were more favorable to plaintiffs in medical malpractice and personal injury cases than juries, and that these results were not solely attributable to case selection effects). See also Vidmar and Rice (1993).
23 We do not explicitly consider the trial/settlement process in this article. The main results of this paper are not affected if the assumption that each litigant knows \( p \) is changed so that each litigant only possesses an unbiased estimate of \( p \) at the trial/settlement decision. Trials would be generated when parties prior estimates of the prior distribution were relatively optimistic. See, e.g., Priest and Klein (1984). Further, trials could also be generated even if the parties agree on \( p \) - e.g., if the litigants' differential valuation of precedent causes the absence of a settlement range. See, e.g., Kobayashi (1995), Kobayashi and Lott (1994), and Rubin (1977).

24 We assume that accuracy in adjudication is a relevant societal goal, so that the litigants' increased investment in information is not a social cost. For example, accurate information may improve individuals ex-ante incentives to behave properly, and can affect marginal deterrence. See, e.g., Kaplow and Shavell (1992, 1994), Kaplow (1994), and Rasmussen (1992). Uncertainty can also increase the "option value" of lawsuits, and can increase the number of non-meritorious suits. See Cornell (1990), Landes (1992), and Kobayashi and Parker (1993). In addition, more accurate (i.e., lower variance) information can more effectively offset erroneous prior beliefs held by the decision maker. For an analysis of the costs and benefits of competitively versus centrally produced information, see Froeb and Kobayashi (1993).

25 For the discrete case, in equilibrium, it must be profitable for both litigants to take the last flip (to get to the optimal stopping point), but unprofitable to flip further. These four conditions (two for each litigant) characterize the Nash Equilibrium, but do not guarantee its existence. In fact, equilibrium in pure strategies does not always exist unless evidence units become arbitrarily small. As evidence is measured in smaller and smaller units, the discrete equilibrium converges to the continuous equilibrium.

26 We choose a Beta distribution because it is the only distribution that allows an analytic solution for the posterior mean. Other prior distributions are possible, but the posterior mean could only be calculated numerically. The results that follow depend critically on the functional form of the posterior mean that derives from the use of the Beta prior.
27 For research suggesting that experimental jurors fail to appropriately combine evidence, see the references contained in note 20, supra.

28 See the discussion in note 14, supra.

29 Bias can also be imposed by varying the standard of proof. For a discussion of this issue, see, e.g., Sobel (1985), Rubinfeld and Sappington (1987), Davis (1994).

30 A distinction between the two types of evidence can be made on the basis of Hirshleifer's classic separation of information into "foreknowledge" and "discovery". The former type of information is either pre-existing or inelastically supplied, while the latter type of information will only be produced with costly effort (see Hirshleifer, 1971). For a discussion of this in the context of legal issues, see Easterbrook (1981), Kaplow and Shavell (1989), and Allen, Grady, Polsby, and Yashko (1990) (legal privileges), Kobayashi, Parker and Ribstein (1994) (mandatory disclosure) and Parker (1994), and Froeb and Kobayashi (1993) (expert testimony).

31 While the recent amendments to discovery practice under the Federal Rules of Civil Procedure attempted to reduce litigant's ability to hide information, hiding unfavorable expert information generally can be accomplished by hiring multiple experts, and selecting for trial those with favorable results. Under Rule 26(b)(4) of the Federal Rules of Civil Procedure, the ability of a party "to discover facts known or opinions held by an expert who has been retained or specially employed by another party in anticipation of litigation or preparation for trial and who is not expected to be called as a witness at trial ... ." is severely limited. See F.R.C.P. Rule 26(b)(4). See Johnston (1988) for a discussion of the practice of "expert shopping". For a discussion of the amendments to the discovery rules, see the discussion in note 12, supra.

32 When the distribution is unknown to the litigants, flipping takes on an “option” value. Using the intuition of Roberts and Weitzman (1981), if the uncertainty is large relative to the cost of flipping, litigants will always take a few initial flips because it gives them the option of flipping more if they learn the distribution is favorable. In our model, this would change the nature of
equilibrium for cases in which one of the litigants would not flip, absent the option value of flipping. It would move the estimator further away from the true value towards the prior mean. However, when the cost of flipping is low, Rothschild (1974) finds that for the Dirichelet/multinomial conjugate distributions, which are generalizations of the beta/binomial, “not invariably, but in many instances,” the distribution is "learned" as more draws are taken so that the stopping rule is similar to that used when the distribution is known in advance. For recent review of optimal stopping, see McMillan and Rothschild (1994).

33 The uniqueness of equilibrium is proved in the appendix.

34 Note that, in the cases illustrated in Figures 1 and 2, a corner solution is reached (i.e. one of the parties produces no evidence). When \( p \) is close to zero or one, there is bias, but not in cases where \( p \) is "close" to \( .5 \). According to the Priest and Klein model of litigation versus settlement, the selection of cases for litigation will be biased toward cases where \( p \) is close to \( .5 \). See the discussion in note 23, supra. Thus, under the Priest-Klein hypothesis, the set of litigated cases will be relatively free of bias. However, the amount paid in settled cases (those where \( p \) is close to zero or one) will reflect the jury's bias.

35 Note that in the case of the classical jury (Figure 3), a tax on evidence will increase welfare, as the amount of evidence and thus the costs of producing evidence will fall. However, this is not true in the cases set out in Figures 1 and 2. In these cases, a tax on evidence will increase the range over which one party produces no evidence. Although evidence production costs are lower, the accuracy of the estimator will decrease. Further, given the assumption of the model, the additional bias will be predictable ex-ante, affecting settlement, and ex-ante behavior.

36 Fee shifting, or “loser pays” rules, are somewhat difficult to address in the context of this model because of the discarded unfavorable evidence. If one of the litigants winds up paying the other’s legal costs, he would presumably learn about the other’s discarded evidence. In addition, the fee shifting used in England, and Rule 68 of the Federal Rules of Civil Procedure, apply only to marginal expenditures incurred following an offer of judgment. Since we do not
model settlement, it is difficult to evaluate these rules in the context of our model. However, putting these considerations aside, with fee shifting the party disfavored by the distribution produces no evidence because it leads to more costly evidence production for which he will likely end up paying. This leads to a bias in the jury’s estimator, in favor of the litigant favored by the distribution. For discussions of similar effects of fee shifting on the rates of filings, settlement and trial, see Shavell (1982).

37 With rising marginal costs (with the number of flips), the nature of optimal search changes for the parties. No longer are open and closed loop strategies equivalent (McMillan and Rothschild, 1994; Roberts and Weitzman, 1981). If the litigants stop and reoptimize, the optimal number of flips changes because the marginal conditions are likely to change. If for example, the plaintiff started off with an unlucky streak (no heads), he would end up settling, on average, for fewer heads than he would have produced following an open loop strategy (computed before any flips were taken). This changes the nature of equilibrium by introducing variance into the estimator of the jury. Sometimes the plaintiff will get lucky and sometimes the defendant will get lucky. On average, the “luck” will balance out, but this is difficult to prove analytically because closed loop equilibria are difficult to compute. Instead, we have examined open loop equilibria as an average benchmark for the closed loop equilibria in the rising marginal costs case. We find that principle accuracy result for the constant cost case is robust with respect to rising marginal costs for the “classical” estimator, but not for the “Bayesian” estimator. For \( a \neq 0 \) or \( b \neq 0 \), we find that while the jury’s estimator is close to the true \( p \), it is not exact.