The Burden of Proof in Civil Litigation

by
Chris W. Sanchirico, Columbia University

February 1996

Discussion Paper Series No. 9596-06

dp 9596-06

pages. 28

THE BURDEN OF PROOF IN CIVIL LITIGATION:

A Simple Model of Mechanism Design

CHRIS WILLIAM SANCHIRICO*

Columbia University
Economics Department and School of Law
420 West 118th Street
New York, NY 10027
(212) 854 4037
cs282@columbia.edu

ABSTRACT: The existing literature on the burden of proof has sought the rule's *raison d'etre* solely within the court's problem of decision making under uncertainty. While this search has yielded many insights, it has been less successful in providing a compelling explanation for why uncertainty in the court's final assessment should act to the detriment of one party rather than the other. By viewing the problem as one of mechanism design, this paper provides one explanation for the asymmetry. A rule resembling the burden of proof emerges from the optimal design of a system of fact-finding tribunals in the presence of: *i)* limited resources for the resolution of private disputes, and *ii)* asymmetric information--as between the parties and the court-- about the strength of cases prior to the court's having expended the resources necessary for a hearing. The paper shows that if the objective in designing a trial court system is accuracy of recovery granted, the "value" of having heard a case will depend in part on the certainty with which the court makes its final award. An optimally designed court system will then effectively filter-out "less valuable" cases by precommitting to a recovery *policy* in which plaintiffs recover nothing unless they prove their cases with a threshold degree of certainty.

1. INTRODUCTION

Few principles of law are as well settled as that which says that the plaintiff, more generally the moving party, shall have the burden of proving her claim with a "preponderance of the evidence.[1]" Yet few principles have inspired as many differing explanations and interpretations by legal and legal-economic commentators alike.

To date, most of the economic analysis of the burden of proof has attempted to make sense of the rule in the context of the theory of decision making under uncertainty--in particular, within a Bayesian framework in which the court begins with prior beliefs about the veracity of relevant factual assertions and then updates these beliefs according to Bayes' rule upon hearing the evidence placed before it. But the burden of proof is difficult to find within this framework. If we assert that it instructs the court to rule against the burdened party when the court is in "equipoise"--when its updated beliefs put exactly probability 0.5 on the truth of the assertion--then we have relegated the rule to a rare coincidence. If, on the other hand, we assert that the rule requires the court to accept the factual assertion only when its updated beliefs exceed some threshold *above* 0.5, then we save the rule from irrelevance, but beg even more ardently the question of why we should so favor one party over the other--why, for instance, we should feel less comfortable overcompensating plaintiffs than leaving legal wrongs "unrighted."

The premise of this paper is that the theory of decision making under uncertainty, by itself, does not and can not provide a satisfactory explanation of what the burden of proof is or is supposed to be; that the burden of proof must instead be understood in the context of mechanism design and asymmetric information. Under the present system of civil litigation, potential plaintiffs are vested with the power to set in motion a costly process of litigation, all of whose costs are not in the first instance their own. If this were the entire system, society would continually find itself spending more to resolve disputes than is warranted by the disputes themselves. One could internalize the full social cost to plaintiff by simply charging her for all expended resources. But this

would conflict with the immediate object of the system: to award "proper" recovery to plaintiff, which however defined, must be net of her costs. On the other hand, if the court system knew beforehand which cases would be worth the resources necessary to resolve them, it could simply refuse to take those that were not. The problem is that the court does not find out the worth of a case until after society has paid to hear it. The solution, I will argue, is to announce to all potential plaintiffs that only those cases that *turn out to have been* worth hearing will receive any recovery. Then potential plaintiffs, who have superior information about the stakes and solidity of their claims, will self-select, and on average only those cases worth hearing will make their way to court.

The link between this rule and a burden rule based on uncertainty in the court's final assessment comes through the determination of which cases are worth hearing. For a trial court, concerned primarily with the accurate application of existing law, the value of having adjudicated a case will turn in part, I will argue, on the court's confidence that its ruling was the right one. The court's announced recovery policy will then be to award nothing to plaintiff unless she proves her case with a threshold "degree of certainty" (defined within). Thus, as with the actual burden of proof, recovery under this rule will depend not only on the court's best estimate of the amount that the plaintiff should be awarded, but also on the certainty with which the court makes that estimation. Moreover, the optimal recovery policy will have the same threshold structure evident in the actual rule: roughly, if we fix the expected value of proper recovery (a random variable whose distribution is induced by the court's posterior beliefs) and progressively decrease its variance, the optimal recovery policy jumps from zero up to that expected value as the variance crosses the specified threshold; thereafter, recovery remains fixed at that expected value regardless of the degree to which the threshold exceeds variance. On the other hand, unlike the manner in which the actual burden of proof seems to be applied in practice, the optimal threshold for any given class of cases increases with the cost of litigation and/or settlement and decreases with the expected value of proper recovery.

In Section 2, I catalogue and analyze existing attempts to make sense of the burden of proof solely in the context of decision making under uncertainty. Readers familiar with this literature will want to skip first to Section 3 which contains the basic model. In Section 4, I add the possibility that parties may bargain before trial. Concluding remarks appear in Section 5.

2. ATTEMPTS TO FIND THE BURDEN OF PROOF IN THE THEORY OF DECISION MAKING UNDER UNCERTAINTY

2.1 Ties Go to the Defendant

The most common interpretation of the burden of proof portrays it as a tie-breaker rule. This is, for example, the interpretation adopted in James, Hazard and Leubsdorf (1992) and Lempert and Saltzburg (1983). In contrast to its near ubiquitous acceptance, however, the tie-breaker interpretation has the curious and unsettling property that as soon as one tries to formalize it, it disappears. As noted, the odds that the court's updated probability will land exactly on the knife edge of 0.5 are effectively nil. And since this point concerns how priors are updated on the basis of evidence, it will hold even if there are legal reasons to set the *unconditional* prior probability at 0.5.

2.2 The Raised Threshold

One way to save the burden of proof from the irrelevance of coincidence--and perhaps what those who propose the tie-breaker rule really have in mind--is to interpret the rule to mean that the court must be convinced that an assertion is true with some probability greater than a threshold $\bar{P}$, where $\bar{P} > 0.5$. But while increasing $\bar{P}$ above 0.5, may give some operational bite to the preponderance standard, it leaves unanswered the question of why we should favor the defendant in civil cases. As Posner (1973) remarks, there seems to be no *a priori* reason for favoring erroneous exoneration over erroneous liability.
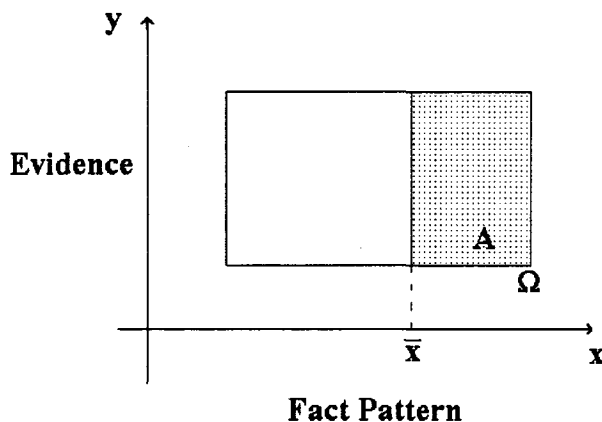
2.3 Biased Prior

Cooter and Ulen (1988) suggest that the burden of proof means that the court should bias its *prior* in favor of the defendant. The analysis of this interpretation depends on what it means to bias a prior. One possibility is that *any* shift in probability weight that causes the unconditional (i.e. prior) probability of the truth of the factual assertion to decrease is a proper bias. But biasing the prior in this manner seems somewhat arbitrary, for it implies nothing about even the direction of change in the *posterior* for any given presentation of evidence. It is, for example, possible to decrease the unconditional probability of the factual assertion while *in*creasing the posterior after all but one possible presentations of evidence.

An alternative view is that the prior should be biased in such a way that affects all conditional probabilities "uniformly." But in this case, the biased prior rule is effectively equivalent to the raised threshold rule discussed above. More precisely, for any threshold level $\overline{P}$, we can find a "biased prior" such that for all evidence, the decision made under the new prior with the threshold level set at $P = \frac{1}{2}$ is the same as the decision made under the unbiased prior with threshold level $\overline{P}$.

2.4 Confidence Levels (Cohen (1985))

Cohen views the burden of proof in terms of confidence intervals, mainstays in the tool kit of classical statistics. But though Cohen takes pains to explain the confidence interval in the context of classical statistics, it is less clear from the article how one would apply the notion to the generic problem of legal fact-finding.[2] One attempt to do so is illustrated in the accompanying graph, which identifies the set of conceivable pairs of fact patterns x and bodies of evidence y with a closed (two dimensional) interval in the

Fact Pattern

Cartesian plane. In this graph the factual assertion A-- a subset of the state space-- is true only at those states (x,y) for which x lies above a particular value[3] $\bar{x}$.

The fact finder begins with prior beliefs over the state space that induce a conditional probability measure on fact patterns x for every value of evidence presented y. Just as fixing y induces a distribution on x, fixing x induces a distribution on y, and so any function thereof. A 95% (for example) confidence interval is thus any pair of functions of y, $U(y)$ and $L(y)$ with the property that *for all x*, the probability that both $U(y) > x$ and $L(y) < x$ is 95%.

One particular type of confidence interval, and the one used by Cohen, fixes $U(y)$ at the maximal value for x, for all y. Placed in this context, Cohen's burden of proof stipulates that we accept the fact A as *legally* true if and only if the observed value l of the lower bound $L(y)$ exceeds $\bar{x}$.[4]

The first problem with applying *any* sort of confidence intervals to the legal fact-finder's problem is that none may exist. Whether we can find functions $U(y)$ and $L(y)$ with the property that the probability of $L(y) < x < U(x)$ is 95% (or any fixed percentage) **across *all* x**, is by no means clear. In contrast, the problems in statistics to which confidence intervals are applied have very special structures which guarantee the existence of such intervals.

But even if we impose additional structure on the problem sufficient to guarantee the existence of the confidence interval, a larger, more significant problem of interpretation remains. Classical statisticians will say that after learning a particular y and calculating the corresponding values $u = U(y)$ and $l = L(y)$, we know that the true fact pattern x lies in the interval (l,u) "with 95% confidence." Whatever "confidence" means here, it does *not* mean "probability" -- it is *not* correct to say that there is a 95% chance that the true fact pattern lies in the interval (l,u). In order to talk about the *probability* that x lies in any particular range, we have to view x probabilistically, an outlook which classical statistics does not admit. And even if we shift to the Bayesian viewpoint and posit prior beliefs on (x,y) updated after revelation of y, it is still not generally true that, according to our posterior belief on x, x will lie in the interval (l,u)

with probability .95. Our posterior on x, and thus our posterior belief that x lies in (l,u) depend on both the structure of the fact finder's prior and the particular y which has been observed. All we *can* say is that *ex ante* revelation of y, there was a 95% chance that the true value of x would lie between whatever values of $U(y)$ and $L(y)$ were revealed. *Ex post* revelation of the evidence, y, the confidence interval has no particular interpretation.[5]

Let us go yet another step and suppose hypothetically that the fact-finder's problem *had* a structure sufficiently specialized[6] to guarantee both that 95% confidence intervals exist and that they could be interpreted as one is tempted to interpret them: with the word "probability" substituted for the word "confidence." Indeed, let us make the even stronger assumption that it is possible to construct such "interpretable" confidence intervals with the restriction that $U(y)$ to be set constant at the maximal value taken by x. We would then know, on observing y and subsequently calculating $l = L(y)$, that x was greater than l with probability .95. Cohen's burden of proof stipulates that we find A to be true, if and only if l falls above the threshold $\overline{x}$. With our hypothetically interpretable confidence interval, this reduces to nothing more than the rule that we find A true only in the case that its posterior probability is at least .95, which is precisely the same as raising (or lowering) the threshold probability to the confidence coefficient.

2.5 Summary

Reflecting back on the four existing interpretations of the burden of proof, we see that each attempts to find the burden of proof solely within the theory of decision making under uncertainty, each at its best reduces to the second, "raised threshold" interpretation, and none provides an explanation for *why* the threshold should be so raised to favor one party over the other.

3. The Basic Model

In the model presented in this section, the burden of proof is derived from the court's optimization problem. The interpretation of the burden presented here, then, comes part and parcel with its justification. Further, the structure which emerges does not resemble a raised threshold, but instead is keyed to dispersion in the court's beliefs after hearing the evidence, a statistic which rarely appears in straight Bayesian analysis, but seems more in line with casual empiricism about the burden's true operation.

Our task is to design a *trial* court system, that is to say a system of fact-finding tribunals. We have already set "the law" according to principles such as fairness and efficiency. This law tells us how much a plaintiff should recover from a defendant[7] as a function of all the relevant factual information surrounding the case. The problem we face in designing our trial court system is that we will not know, in any particular dispute, what these facts are.[8] As a result, we want to design a system whereby we can learn more about the fact pattern before deciding how much, if anything, the plaintiff should recover.

In designing this system we must be cognizant of its opportunity cost. Resources used for the trial court system are resources that are not used for schools, national defense, or even the legislative process whereby better law might be designed. We therefore face a fundamental tradeoff in designing our trial court system--between, on the one hand, effectively rewarding no recovery to the plaintiff[9] and using the unspent resources elsewhere, and, on the other hand, spending the resources necessary to hear the case in the hope that the more informed decision we make by virtue of the hearing will be sufficiently "better," in some sense, than awarding nothing.

Though it is difficult to make precise statements about the benefit to society of awarding any particular level of recovery, it is easy to construct examples where *not* hearing the case at all seems in retrospect like the best alternative. Suppose, for example, that a $100,000 hearing determines that with probability 1, the defendant owes the plaintiff $1 in damages. In retrospect, we would have been better off not

hearing the case in the first place: we would have saved $100,000 at the cost of leaving the plaintiff $1 poorer and the defendant $1 richer than the law would like either to be.

3.1 An Objective Function for the Court System and the Induced Value of Hearing Cases

To make more interesting statements about this tradeoff, we must commit to some measure of the benefits of awarding a given level of recovery. I start with the much simpler case in which the court knows the true fact pattern and will then extend the analysis to the situation where it is unsure.

Suppose we knew for certain that the fact pattern in a given dispute was $\omega \in \Omega$, where Ω, the sample space of our uncertainty, is the set of all possible fact patterns. It seems natural in this case that we would want our *trial* court to award the *proper recovery* specified by law for that particular fact pattern. Call this amount $L(\omega)$ and view L as a random variable on Ω. Saying that we prefer to award proper recovery is the same as saying that awarding any amount greater than *or* less than $L(\omega)$ is worse than awarding $L(\omega)$ itself. In an important sense, then, we treat the two types of legal error symmetrically.

One functional form which captures this symmetry, and turns out to be relatively easy to work with, is the negative absolute difference between proper and actual recovery, $-|L(\omega) - r|$, where r stands for actual recovery awarded, our choice variable. This functional form obviously attains a maximum at $r = L(\omega)$ and is monotonically increasing to the left of this value and monotonically decreasing to the right. In the remainder of this paper, I will take this function to be the court's payoff function over certain outcomes.[10] In the case that the court is certain of the true fact pattern $\omega \in \Omega$, this function is its objective function in the (quite simple) mathematical programming problem in which it chooses optimal recovery. As usual, the *value* of the problem itself is the value of the objective function at its maximum, namely 0.

More realistically, even after the court hears the case, it is still unsure of the true fact pattern. As in Bayesian decision theory, let us say that the court began the hearing

with prior beliefs P_0 about the true fact pattern[11] and then updated these beliefs on hearing the evidence. Now it wishes to maximize expected payoffs based on these updated beliefs. If its updated beliefs over the true $\omega \in \Omega$ are denoted by P, then it will choose r to maximize $-E_P|L(\omega) - r|$.

Letting F_P denote the cumulative distribution of L under P, let us assume:

ASSUMPTION 1. F_P, *is continuous*.

Then the distribution of L has a median and this is the optimal amount of recovery:[12]

LEMMA 1. r^* *solves* $\max\limits_r - E_P|L(\omega) - r|$, *if and only if* $F_P(r^*) = \frac{1}{2}$.

Proof: A standard result.

Therefore, the *value* of the problem, $\max\limits_r - E_P|L(\omega) - r|$, is $-E_P|L(\omega) - v_P|$. On the other hand, if the court had *not* taken the case, recovery would have been in effect zero and the corresponding value would be $-E_P|L(\omega)|$. The difference between these two expressions, namely, $E_P|L(\omega)| - E_P|L(\omega) - v_P|$ is one measure of the *ex post benefit of having heard the case*.

With a little manipulation we can express this expected benefit in a more informative manner. Let $\bar{\mu}$ represent the *upper mean* of the random variable L, that is, the expected value of L conditional on its exceeding its median, vP. In the usual notation, this is $E_P\left[L \mid L \geq v_p\right]$. Define the *lower mean* $\underline{\mu}$ in a similar manner, $E_P\left[L \mid L \leq v_p\right]$. The difference $\bar{\mu} - \underline{\mu}$ between the lower and upper mean is a measure of the "dispersion" of the random variable L. If the random variable is degenerate and takes only one value, then this difference is zero. If the random variable takes only two values with equal probability, this difference reduces to the difference between these two values. This difference is as well a measure of dispersion for more complex

distributions. One can show, for example, that if the random variable is distributed either normally, uniformly or exponentially, then this difference is proportional to the standard deviation. For convenience, then, I will dub this statistic, $\overline{\mu} - \underline{\mu}$, the *dispersion* of the random variable L and denote it as δ.

LEMMA 2. $E_P|L(\omega)| - E_p|L(\omega) - v_p| = E_P|L(\omega)| - \tfrac{1}{2}\delta$.

Proof:

$$E_p|L(\omega) - v_p|$$

$$= \int_{-\infty}^{v_P}(v_P - L(\omega))dF_P + \int_{v_P}^{\infty}(L(\omega) - v_P)dF_P$$

$$= \int_{-\infty}^{v_P} v_p dF_P - \int_{v_P}^{\infty} v_p dF_P - \int_{-\infty}^{v_P} L(\omega)dF_P + \int_{v_P}^{\infty} L(\omega)dF_P$$

$$= v_p \int_{-\infty}^{v_P} dF_P - v_p \int_{v_P}^{\infty} dF_P - F_P(v_P)\frac{\int_{-\infty}^{v_P} L(\omega)dF_P}{F_P(v_P)} + F_P(v_P)\frac{\int_{v_P}^{\infty} L(\omega)dF_P}{F_P(v_P)}$$

$$= \tfrac{1}{2}v_p - \tfrac{1}{2}v_p - \tfrac{1}{2}\underline{\mu} + \tfrac{1}{2}\overline{\mu}$$

$$= \tfrac{1}{2}\left(\overline{\mu} - \underline{\mu}\right). \blacksquare$$

Thus, the value of having heard the case increases in the *ex post* expected absolute value L and decreases in the dispersion of the court's updated beliefs. Holding expected absolute value constant, then, the court prefers cases which leave it with "concentrated" beliefs to those which leave it with "diffuse" beliefs.

In determining whether the case was worth hearing, we would want to compare this measure $E_P|L(\omega)| - \tfrac{1}{2}\delta$ of expected benefit with the opportunity cost of the hearing. Of course, this measure of benefit need not be immediately comparable with our most natural measure of opportunity costs. But for simplicity let us suppose that opportunity cost, c of the hearing are stated in units that do make it comparable.[13] Then, after having heard the case, we can make the following *ex post* judgment about

whether doing so was a good idea: the case was worth hearing if and only if $E_P|L(\omega)| - \frac{1}{2}\delta$ turns out to exceed c.

3.2 Optimal Choice of Recovery Policy

If the court somehow knew what its updated beliefs P would be following the hearing, it could determine ahead of time which cases to hear and which to dismiss according to this comparison of costs and benefits. The problem, however, is that the court does not learn whether a case is worth "paying" to hear until after it has paid to hear it. The parties, however, do have much information about the value of their case even before the case is played out in court. The court's problem is, thus, characterized by an asymmetry of information.

Nevertheless, the court can get around this disadvantage, at least to some degree. The court *does* learn the value of the case before awarding recovery. It can therefore announce to all potential plaintiffs that if they bring a case to court which, in the end, was not worth hearing, according to the criterion we have laid out, then the plaintiff will receive little or no recovery. If the court sets this level below what it costs plaintiffs to bring their cases, and plaintiffs have a good idea of what the court will think after hearing all the evidence, then this policy will "filter out" plaintiffs whose cases are not worth hearing. From the court's perspective, the result will resemble the hypothetical just discussed, wherein the court knows before hand the value of each case and accepts only those which meet its criterion. The important point for our purposes is that the court's announcement to plaintiffs will bear a stark resemblance to the current burden of proof in civil cases.

Formally, suppose that the court faces a population of potential cases, each identified by the posterior belief, P regarding legal recovery L that it will inspire in the court after the hearing. Let the probability measure Q on the set of all P represent the population composition of cases.

For simplicity suppose that before bringing their cases, plaintiffs know exactly what the court's posterior beliefs, P will be. The court chooses, not just recovery in

each individual case, but a *recovery policy* $r(P)$, for implementation by the trial court, mapping posterior beliefs onto a prescribed amount of recovery for the plaintiff. Plaintiffs also know this recovery policy and believe that it will be carried out by the court system. (The credibility of the policy is an important issue and is discussed below.) We also suppose also that costs plaintiffs $\pi > 0$ to bring suit. (In its natural state, this cost lies with plaintiff; it can always be shifted via the recovery policy, $r(P)$.) Then a plaintiff who knows her case will inspire beliefs of P will file suit, if and only if $r(P) \geq \pi$.

The court, like a Stackelberg leader, knows this "reaction function" and takes it into account in setting its policy. The court's problem is therefore:

Choose $r(P)$

to Maximize $$\int_{r(P) \geq \pi} \left(-E_P |L(\omega) - r(P)| - c \right) dQ + \int_{r(P) < \pi} -E_P |L(\omega)| dQ \tag{1}$$

Assuming that plaintiff's trial costs are less than the opportunity cost of the hearing to society (which presumably contain plaintiff's costs),

ASSUMPTION 2. $0 < \pi \leq c$,

we obtain

PROPOSITION 1. *The recovery policy*

$$r(P) = \begin{cases} v_P, & \text{if } \delta \leq 2E_p |L(\omega)| - 2c \\ 0, & \text{otherwise} \end{cases}$$

solves the court's problem (1).

Under this recovery policy, the court announces that it will decide all cases in two steps. First it will test whether, after hearing all the evidence, the dispersion in its beliefs fall below a threshold level. This threshold level of dispersion will not be uniform across all cases but will depend on the expected level of proper recovery. If

this threshold is exceeded, no recovery (or at least some amount less than π) will be awarded. If, on the other hand, dispersion falls below this threshold, the plaintiff will receive recovery equal to the court's best estimate of what the law prescribes, namely the median of L.

This recovery policy bears a strong resemblance to, and thus helps make sense of, the burden of proof in civil litigation. Under this optimal rule, it is *not* enough for the plaintiff to prove that expected proper recovery is positive. The plaintiff must also do so in such manner as to inspire in the court a level of confidence in its estimation, as measured by the dispersion in the court's posterior belief. Should the court have sufficient confidence, the plaintiff is awarded the median, regardless of whether the plaintiff exceeds that level of confidence by a wide margin or no.

Moreover, this recovery policy enables the court system to do as well as it would if it knew beforehand which cases were worth hearing and which were not. The incentives which the rule creates guarantee that only those cases that the court *would have wanted to* hear retrospectively are actually heard.

Proof of Proposition 1: Define the function

$$\Phi(r,P) = \begin{cases} -E_P|L(\omega) - r| - c, & r \geq \pi \\ -E_p|L(\omega)|, & r < \pi \end{cases}.$$

The objective in problem (1) can be rewritten as:

$$\int \Phi(r(P), P) dQ.$$

Hence, it suffices to show that for all P, $r(P)$ maximizes $\Phi(r,P)$. To this end, first consider any P such that $\delta > 2E_p|L(\omega)| - 2c \Leftrightarrow E_P|L(\omega)| - \tfrac{1}{2}\delta < c$, so that $r(P) = 0$.

By Lemma 2, then, $-E_P|L(\omega)| > -E_p|L(\omega) - v_p| - c$ and so by Lemma 1, for all r,

$-E_P|L(\omega)| > -E_p|L(\omega) - r| - c$. Therefore, $\Phi(r,P)$ achieves a maximum of $-E_p|L(\omega)|$

at any r with $r < \pi$; in particular, at $r(P) = 0$.

Now suppose that $\delta \leq 2E_p|L(\omega)| - 2c \Leftrightarrow E_p|L(\omega)| - \frac{1}{2}\delta \geq c$ and $v_p \geq \pi$. Then by Lemma 2, $-E_p|L(\omega)| \leq -E_p|L(\omega) - v_p| - c$ and since $v_p \geq \pi$, $\Phi(r, P)$ must attain its maximum where $r \geq \pi$. By Lemma 1, this is at $v_p = r(P)$.

Lastly suppose that $\delta \leq 2E_p|L(\omega)| - 2c \Leftrightarrow E_p|L(\omega)| - \frac{1}{2}\delta \geq c$ and $v_p < \pi$. Then the largest value attained by $\Phi(r, P)$, *given* $r \geq \pi$, is $-E_p|L(\omega) - \pi| - c$. The largest value attained by $\Phi(r, P)$, given $r < \pi$ is trivially, $-E_p|L(\omega)|$. We see that:

$$-E_p|L(\omega)| - \left(-E_p|L(\omega) - \pi| - c\right)$$

$$= E_p|L(\omega) - \pi| - \left(E_p|L(\omega)| - c\right)$$

$$\geq E_p|L(\omega) - \pi| - \left(E_p|L(\omega)| - \pi\right)$$

$$\geq E_p|L(\omega) - \pi| - \left(E_p|L(\omega)| - |\pi|\right)$$

$$\geq 0,$$

where the last inequality is the "triangle inequality." Therefore, $\Phi(r, P)$ attains it maximum at any r with $r < \pi$, in particular, at $r(P) = 0$. ∎

The only difficulty in showing that *r(P)* is an optimal rule is showing that the court would not want to award more than the median to encourage plaintiffs with worthy cases who happen to have median recovery less than their nominal cost to bring suit. But if this nominal cost is less than the social opportunity cost of hearing the case, then there will be no such plaintiffs. By an argument based on the triangle inequality, the reduction in accuracy caused by awarding the plaintiff more than the median (in order to induce her to sue) will outweigh the benefit of having her sue. Note also that this is *an* optimal rule because, the court may award anything less than the plaintiffs nominal cost in the "otherwise" case.

It is important to note that this optimal policy is not subgame perfect. It is supported by the threat that the court will in certain circumstances take actions which are not, at the time they are made, in its own best interest. That is, after hearing a

case, whatever the case's ex post value, the best policy for the court *from that point onward* is still to award median recovery. The cost of hearing the case is already sunken and thus not relevant to the court's current decision. The recovery rule described above will only work if the court can make a credible commitment not to give in and award median recovery in cases which turn out to be, in retrospect, not worth hearing. In the typical game theoretic model of individual or firm behavior, lack of subgame perfection is a fatal flaw. However, in a model of a court system governed by rules of procedure and precedent, threats by the court *system* to follow rules which ex post seem senseless for all parties *are* plausible. Indeed, constructing models wherein the parties' beliefs about the court are not shaped by subgame perfection seems the *more* realistic alternative.

4. INCORPORATING PRE-TRIAL BARGAINING

While precise figures differ across studies, it is well accepted that a supermajority of civil suits never reach trial. How does this affect the foregoing analysis of the burden of proof? In this section I show that the optimal recovery policy with pre-trial settlement still resembles the burden of proof. Now, however, the award for those cases which meet the burden is not median legal recovery, but median recovery *plus* some function of the parties prospective trial costs. Moreover, the threshold level of dispersion is tied not to these trial costs as in the previous section, but to the cost to society of the process of litigation, including filings, discovery and settlement negotiations, up to but not including trial.

The simple litigation "game" in this model with pre-trial negotiation has several steps. First the court announces a recovery policy, $r(P)$ to all potential parties. Next, plaintiff, knowing P and $r(P)$, decides whether to file suit. Third, the case enters settlement negotiations, which imposes costs of $\sigma_\pi > 0$ and $\sigma_\delta > 0$ on plaintiff and defendant respectively. The cost to society of this settlement phase is $\sigma > 0$. In this settlement phase, defendant, who also knows P, $r(P)$, and π makes a settlement offer

s to plaintiff. Fourth, the plaintiff decides whether to accept this offer. If so, the game is over and the defendant pays s to plaintiff. If not, the plaintiff decides whether to take his case to trial. If so, the case enters the trial phase at a cost of $\pi > 0$, $\Delta > 0$ and $c > 0$ to plaintiff, defendant and society, respectively. In this phase, as before, the court hears the evidence in the case, updates its prior belief about the fact pattern from P_0 to P and then orders defendant to pay $r(P)$ to plaintiff.

If we fix a recovery policy, $r(P)$, we can solve the resulting game of perfect information between plaintiff and defendant by backwards induction. Having filed a case and rejected a settlement offer, plaintiff continues with the case if and only if $r(P) \geq \pi$. Therefore, plaintiff accepts defendant's settlement offer, if and only if

$$s \geq \max(r(P) - \pi, 0).$$

It is then optimal for defendant to offer $\max(r(P) - \pi, 0)$. For if, on the one hand, $r(P) - \pi < 0$, then the plaintiff will not bring the case to court anyway, even if $r(P) + \Delta > 0$, and so the defendant should offer $s = 0 = \max(r(P) - \pi, 0)$. If, on the other hand, $r(P) - \pi \geq 0$, then the plaintiff will bring the case to court and so the defendant stands to lose $r(P) + \Delta > 0$. The plaintiff will accept any offer above $r(P) - \pi < r(P) + \Delta$ and so the defendant should offer the smallest acceptable offer: $r(P) - \pi = \max(r(P) - \pi, 0)$.

Since plaintiff knows that this is how settlement negotiations will proceed, she files suit, if and only if $s = \max(r(P) - \pi, 0) \geq \sigma_\pi$, or equivalently, if and only if $r(P) \geq \pi + \sigma_\pi$.

In sum, only cases that meet the condition $r(P) \geq \pi + \sigma_\pi$ are filed and of these, the defendant ends up paying the plaintiff $r(P) - \pi$. The court's problem is then:

Choose $r(P)$

to Maximize $\qquad \displaystyle\int_{r(P)-\pi \geq \sigma_\pi} \left(-E_P|L(\omega) - (r(P) - \pi)| - \sigma\right)dQ + \int_{r(P)-\pi < \sigma_\pi} -E_P|L(\omega)|dQ \qquad$ (2)

We can view this problem as one of choosing $(r(P) - \pi)$ to maximize (2) and then adding π to the answer to get the optimal $r(P)$. (We need not add π for those cases where we have set recovery low enough to induce the plaintiff not the file suit.) It is clear then that the problem (2) is of the same form as the problem in the previous section with σ_π playing the role of π and σ playing the role of c. If we similarly assume

ASSUMPTION 3. $0 < \sigma_\pi \le \sigma$.

then we have already proven:

PROPOSITION 2. *The recovery policy*

$$r(P) = \begin{cases} v_P + \pi, & \text{if } \delta \le 2E_p|L(\omega)| - 2\sigma \\ 0, & \text{otherwise} \end{cases}$$

solves the court's problem (2), with pre-trial negotiation.

In this optimal policy, the court is still using its recovery policy to affect the plaintiffs incentive to file suit. But, since pre-trial negotiation lowers the cost of "adjudication" broadly defined, the universe of cases which are worth adjudicating is larger. Accordingly the threshold level of dispersion is lower--it is now keyed on settlement costs σ rather than the costs of the entire process from filing to judgment, c. On the other hand, the court is also still using its recovery policy to make awards to plaintiffs whose cases are worth adjudicating. But this less costly method of adjudication--settlement in the shadow of trial--has a biased outcome relative to what the court would award if it heard the case, namely the median, v_P. This is because plaintiff's trial costs mean that defendants need not offer full expected recovery to obtain acceptance. Therefore, the court must correct for this bias by awarding the median plus the plaintiff's trial costs.

The model is robust to the structure of pre-trial negotiations. If, for example, the plaintiff made the offer and the defendant decided whether to accept or reject, recovery,

when granted at all, would correct for the resulting bias in favor of the plaintiff. We can also allow for a more fluid negotiation scheme and merely assume an independent probability distribution over the plaintiff's share of the surplus (which encompasses the case wherein the surplus is always split in some fixed proportion). Then the optimal recovery policy would correct for the expected bias of settlement outcomes. For example, the plaintiff's share were uniformly distributed between 0 and 1, and plaintiff's and defendant's trial costs were equal, there would be no bias and no need for correction.

The model is also robust to allowing the court to condition recovery on the defendant's settlement offer, as in Rule 68. (Note again that the recovery can include fee shifting.) In this case, the defendant will offer $\min_s s : s \geq \max(r(P,s) - \pi, 0)$. If we assume that $r(P,s)$ is lower semi-continuous in s, then for every P, we can find a solution $s(P)$ to $\min_s s : s \geq \max(r(P,s) - \pi, 0)$ and at that solution

$$s(P) = \max(r(P,s(P)) - \pi, 0).$$ Defining $r(P) = r(P,s(P))$ we may proceed as above to find the optimal $r(P)$ and then set $r(P,s) = r(P,s(P))$ for all s. Hence, the court can not improve its objective by conditioning on defendant's settlement offer.

5. CONCLUSION

This paper has shown how a form of burden of proof-- in particular, a rule by which no recovery is granted to plaintiffs unless dispersion in the court's posterior falls below a threshold level--is the optimal solution to the court's basic problem of balancing its desire to 1) award its best estimate of proper recovery and 2) create the proper incentives for plaintiffs in their decision of whether to sue. In the simple model presented here, such a policy allowed the court to achieve its (*ex ante*) first best outcome.

The model might be extended in several directions. 1) A more general objective function for the court could be used--one which allowed for: a) the asymmetric

weighting of over versus under compensation of plaintiffs and b) non-linear (e.g. squared or squared-rooted) error costs. Generalizing the model in this manner would allow it to encompass situations wherein efficient incentives for precaution or activity level (in the activity generating the legal claim) dictate the asymmetric and/or nonlinear treatment of litigation error. 2) A more general model would allow for the possibility that plaintiffs themselves are unsure of what the court's posterior will be (but are still more certain than the court itself prior to the hearing). 3) A more general model would incorporate the endogenous choice of trial preparation effort.

6. REFERENCES

Cohen, N. B. (1985), "Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge," *New York University Law Review*, Vol. 60, June 1985, p. 385.

Cooter, R and T. Ulen (1988), *Law and Economics*, Illinois, Scott, Foresman and Company.

DeGroot, M. H. (1987), *Probability and Statistics*, Second Edition, Reading, Massachusetts, Addison-Wesley Publishing Company.

Fudenberg, D. and J. Tirole (1991), *Game Theory*, Cambridge, The MIT Press, Chapter 7.

Galanter, M. (1983), "Reading the Landscape of Disputes: What We Know and Don't Know (and Think We Know) about our Allegedly Contentious and Litigious Society," *UCLA Law Review*, Vol. 31, October 1983, p. 4.

Hay, B. L. and K. E. Spier, (1994) "Burdens of Proof in Civil Litigation: An Economic Perspective," mimeo, Harvard Law School.

James F., G. C. Hazard and J. Leubsdorf (1992), *Civil Procedure*, Boston, Little, Brown and Company (4th ed.), pp 337-340.

Jenkins, M. C. and J. C. Rehnquist, editors, (1986), Symposium: Probability and Inference in the Law of Evidence, *Boston University Law Review*, Vol. 66, No. 3&4.

Lempert and Salzburg (1983), *A Modern Approach to Evidence*, (2nd ed.), Chapter 9.

Polinsky, A. M. and D. Rubinfeld (1993), "Optimal Awards and Penalties when the Probability of Prevailing Varies Among Plaintiffs," *NBER Working Paper Series No. 4507.*

Posner, R. R. (1973), "An Economic Approach to Legal Procedure and Judicial Administration," Journal of Legal Studies, Vol. 2, p 399.

Posner, R. R. (1992), *Economic Analysis of Law*, Boston, Little, Brown and Company, (4th ed.).

Rubinfeld, D. L. and D. E. M. Sappington (1987), "Efficient Awards and Standards of Proof in Judicial Proceedings," <u>RAND Journal of Economics</u>, Vol. 18, No. 2, Summer 1987.

7. NOTES

---

[1] It is traditional to distinguish two types of burdens of proof in civil litigation: a burden of pleading and a burden of production or persuasion. In this paper, I consider only the latter concept, which is often referred to simply as the "burden of proof." It is also worth emphasizing that the plaintiff does not *always* have this burden, as when the defendant must prove the plaintiff's contributory negligence in tort. There seems, however, to be general agreement among legal scholars that the plaintiff *usually* bears the burden and that, whoever bears the burden, so bearing means having to prove one's case with a "preponderance of the evidence." Lastly, note that the burden (on the prosecution) in criminal cases--the "reasonable doubt standard"--is universally regarded as more difficult to bear. In this paper, I consider only civil litigation. Lastly, I ignore the fact that rebuttable presumptions may change the placement of the burden. For a general review, see James, Hazard and Leubsdorf (1992) pp. 337 *et. seq.* and Lempert and Saltzburg (1983), pp. 792 *et seq.*

[2] In his general discussion Cohen seems to indicate that the fact finder's goal is to learn about the true *probability* that the factual assertion is correct, rather than *whether* the

assertion is correct. This is a direct--perhaps too direct--analogy to the canonical

statistical problem of estimating the underlying probability of success (e.g. the

probability of heads) in a Bernoulli distribution (e.g. a possibly biased coin toss)

through repeated random sampling. To be sure, Cohen does include a hypothetical

contract dispute. But rather than illustrating how confidence intervals fit into the

typical fact finding problem, the example merely embeds a typical *statistical* problem--

estimation of the Bernoulli success probability--into a somewhat uncommon legal

setting. In Cohen's example, defendant retailer avers some probability that a product

shipped to plaintiff--though never received-- was merchantable under the U.C.C. The

mail order contract was F.O.B. and the plaintiff, not challenging this fact--does not sue

for breach. The defendant makes her case based on the random sample she obtained

when testing the shipment from her wholesaler in which plaintiff's particular product

was contained.

[3] The correspondence between this problem and the canonical statistical problem of

determining confidence intervals in estimating the mean of normal distribution with

known variance are as follows: x the fact pattern in the fact-finder's problem,

corresponds to the mean of the normal distribution; y, the evidence in the fact-finder's

problem, corresponds to the random sample generated from the normal distribution[3] y;

the probability of y conditional on any given x corresponds to the sample distribution

for a fixed mean; and lastly, the fact-finder's prior beliefs over pairs (x,y) and the

probability of x for any given y correspond to the nothing in the classical statistical

framework, but to the prior and posterior (on x), respectively, in the Bayesian statistical

framework.

[4] Cohen (1985), p. 403.

[5] See, e.g. Degroot, p. 398, *et seq.*

[6] There *are* results in statistics which give sufficient conditions on the structure of the problem for when confidence intervals, presuming they exist, *may* be interpreted in the probabilistic sense. See, e.g., Degroot, p. 398, *et seq.*

[7] I will proceed as if money damages were the only remedy.

[8] In reality trial courts do not always know what the law is either and thus face both legal and factual uncertainty. I am abstracting from legal uncertainty faced at the trial level.

[9] In reality the court could utilize some form of summary procedure. I do not consider this possibility.

[10] Using squared rather than absolute error would lead to the same results as below with expected value substituted for the median and variance for "dispersion," δ. Indeed, the essential results of the model obtain with more general objectives. The crucial, and generalizable, characteristic of the objective used here is that it is decreasing on either side of its global maximum. This creates a crude concavity in the court's value function which in turn produces what is essentially risk aversion. This risk aversion is the reason that the court prefers cases with lower dispersion.

[11] Such a prior belief is a probability measure on Ω (with sigma algebra understood). This probability measure induces a distribution for the random variable L.

[12] The number x is *median* of the random variable X if $\Pr(X \geq x) = \Pr(X \leq x) = \frac{1}{2}$. A random variable may have no median or many media. A random variable has a median

if its cumulative is continuous. If L has many media, all yield the same payoff. If L had no media, optimal recovery in this problem is the minimum value of l such that

$$\Pr(L \le l) \ge \tfrac{1}{2} .$$

[13] In what follows, I proceed as if the opportunity cost of the hearing were fixed and constant across all suits. This is merely a simplifying assumption. All that is necessary for the results is that the cost of the hearing is known beforehand to the plaintiff and revealed to the court sometime before recovery is granted. In the more general case, the court's recovery policy (see below) would vary with both P and c.

1995-1996 Discussion Paper Series
Department of Economics
Columbia University
1022 International Affairs Bldg.
420 West 118th Street
New York, N.Y., 10027

The following papers are published in the 1995-96 Columbia University Discussion Paper series which runs from early November to October 31 (Academic Year). Domestic orders for discussion papers are available for purchase at $8.00 (US) each and $140.00 (US) for the series. Foreign orders cost $10.00 (US) for individual paper and $185.00 for the series. To order discussion papers, please send your check or money order payable to Department of Economics, Columbia University to the above address. Be sure to include the series number for the paper when you place an order.

9596-01   Protectionist Response to Import Competition in Declining Industries Reconsidered
            J.P. Choi

9596-02   New Estimates on Climate Demand: Evidence from Location Choice
            M. Cragg
            M. Kahn

9596-03   Enforcement by Hearing
            C.W. Sanchirico

9596-04   PreferentialTrading Areas and Multilateralism: Strangers, Friends or Foes?
            J. Bhagwati
            Arvind Panagariya, University of Maryland

9596-05   Simplification, Progression and a Level Playing Field
            William Vickrey

9596-06   The Burden of Proof in Civil Litigation
            C. Sanchirico