

The Economics of Litigation

Carlos A. Abadi

April 22, 2019

All substantive areas of law have a common concern with the processes by which legal disputes get resolved. The existing corpus of economic literature on courts is modest, but understanding the litigation process has become important, as courts intrude more forcefully upon resource allocation. The current cost of trials is unprecedented, although difficult to quantify.

This article consists of four sections. Section 1 focuses on the application of economic tools to the study of courts and outlines the chronology of a legal dispute. In our framework, legal disputes are resolved at various stages of a sequential decision-making process in which parties have limited information and act in their own self-interest. Section 2 reviews the predictions obtained from modeling these decisions, and Section 3 discusses their normative significance. Section 4 concludes.

1 Chronology of a Legal Dispute

Legal scholarship has long concerned itself with how the rules and practices controlling adjudication affect the quality of court decisions. Much of the existing economic literature on courts concerns a variety of microeconomic models involving perfect competition, bargaining, principal-agent relationships, and collective choice. There are special attributes of legal disputes that must be taken into account when adapting any of these models to the study of courts. We develop such a list by briefly describing the chronology of a legal dispute.

Initially, in the first stage, there is an underlying event, such as an accident or crime, in which one person (the injurer) allegedly harms another (the victim). The frequency of harm is affected by decisions that people make concerning activities and precaution. To illustrate, the probability of a car accident increases with the amount that a person drives and decreases with the amount of precaution taken when driving. High levels of certain types of activities and little precaution in performing them increase the frequency with which one person harms another.

Curtailing the activity or taking greater precaution to lower the social cost of the harm is costly in itself. As a result, economic efficiency requires balancing the cost of harm

against the cost of avoiding it. If the parties were able to bargain together, the balance is struck by the market. This observation is the source of the best-known proposition in the economic analysis of law, the Coase theorem (Coase 1960), which states that, in the absence of impediments to exchange, legal entitlements will be allocated efficiently in the market regardless of their initial allocation by law.¹

In many situations, however, bargaining is inhibited or blocked, and the social costs of harm are externalized. For example, drivers and pedestrians do not negotiate agreements in advance to allocate accident costs. For these accidents, the balance between harm and the cost of avoiding it must be struck by law, not the market. The initial allocation of legal entitlements is therefore essential to providing efficient incentives for activity levels and precaution against external harm.

In the second stage of a dispute the party that allegedly suffered harm decides whether or not to assert a legal claim. A rationally self-interested person makes those decision by solving a sequential game that balances immediate costs (hiring a lawyer, filing a complaint) against benefits expected in the future (the proceeds from settlement or victory at trial).

The third stage occurs after a legal claim is asserted, but before trial. During this stage the parties reply to complaints, attend preliminary hearings with the judge, engage in pretrial discovery, and set trial dates. The overall objective of the court at this stage is to encourage plaintiffs (the victims) and defendants (the injurers) to bargain together and settle their disputes. The attribute of litigation bargaining – rivalry, communication, side payments, interdependency, and uncertainty – characterize bargaining games as analyzed in microeconomics. The third stage of the litigation process can be viewed, then, as a bargaining game whose cooperative solution corresponds to a settlement out of court, and whose noncooperative solution corresponds to an adversarial trial.

In settlement negotiations, as in any bargaining game, the interests of the two parties diverge with respect to division of the surplus, but converge with respect to an efficient resolution of the dispute. A legal dispute is resolved efficiently when legal entitlements are allocated to the parties who can bear them at least cost, and the transaction costs of dispute resolution are minimized.

A complicating feature of litigation bargaining is that the parties in most legal disputes are represented by lawyers, whose interests are not identical to their clients'. Designing contracts between attorneys and their clients so that incentives favor good representation is an agency (principal-agent) problem superimposed upon the basic bargaining game.

The law prods disputants to resolve their differences by private bargaining and, when negotiations fail, the courts dictate a resolution in the fourth and final stage of a legal dispute. From the perspective of settlement bargaining, the expected outcome of a trial defines the threat points of the parties. Unlike settlement bargaining, the adversarial element dominates in trials, with each party trying to win as much of the stakes as possible. Litigants, as represented by their counsel, view trials as negative-sum games.

¹Robert Cooter presents an analytical treatment of the Coase theorem.

| TABLE 1 | |
|---|--|
| MATCHING STAGES OF A LEGAL DISPUTE TO ECONOMIC MODELS | |
| Stage 1: | Harm – market models or externality models |
| Stage 2: | Assertion of legal claim – decision under uncertainty to be solved recursively |
| Stage 3: | Bargaining – strategic bargaining model with principal-agent overlay |
| Stage 4: | Trial: negative-sum game for disputants, grafted onto collective choice by impartial court |

Adjudication by the courts has two distinct outputs: dispute resolution and rule-making. From the private viewpoint, trials are a method of resolving disputes between rational self-interested plaintiffs and defendants. But, from a social viewpoint, trials are a mechanism for interpreting and creating laws to regulate and govern society. The decision-makers in appeals courts, where laws are made and interpreted, are judges whose interests differ substantially from those of plaintiffs and defendants.

In our chronology of a legal dispute we distinguish among initial harm, the assertion of a legal claim, settlement bargaining, and trial. The initial harm can be analyzed by market models or externality models of the kind economists apply to conventional economic goods. The decision to assert a claim is a decision under uncertainty to be solved recursively by computing the expected values of subsequent stages in the dispute. Microeconomic models of bargaining are applicable to settlement bargaining. The limitations of bargaining theory, however, are not as severe as the absence of an economic theory of disinterested behavior that is needed to explain how judges interpret statutes and make laws.

The match-up between stages in a legal dispute and the economic modeling of them is summarized in the table above;

A chart depicting the frequency with which disputes go from prior to subsequent stages, with "harm" at the bottom and "appeals court trial" at the top, looks like a broad-based pyramid. A typical finding is that ten disputes settle out of court for every one that is tried, although this figure varies widely by type of dispute. Generally speaking, the farther along the litigation process the dispute has gone, the better the empirical evidence. The steep slope of the "dispute pyramid" and the relative superiority of data describing the top as opposed to the bottom make the empirical study of litigation especially difficult.

2 Resolving Disputes Through the Litigation Process

What incentives do litigants face as they proceed through the litigation process? We answer this question in this section in the context of a formal model that is a hybrid of the models of suit, settlement, and trial that have been developed by William Landes, Richard Posner, Steven Shavell, and others. In Section 3 we go on to treat a number of related normative questions.

Our hybrid model of the litigation process stylizes facts to direct the reader's attention to fundamental causal relations. We assume that all accidents occur between strangers

| Table 2 | |
|-------------------------|--|
| DEFINITION OF VARIABLES | |
| c_{cp} | = cost to plaintiff of asserting a legal complaint |
| c_{sp} (c_{sd}) | = cost to plaintiff (defendant) if the case is settled |
| c_{tp} (c_{td}) | = cost to plaintiff (defendant) if the case is tried |
| D_p (D_d) | = plaintiff's (defendant's) estimate of the compensatory damage to be awarded if the plaintiff wins at trial |
| H_p | = victim's subjective value of the harm he suffers |
| L_p (L_d) | = potential plaintiff's (defendant's) subjective expected value of a legal claim |
| p_{tp} (p_{td}) | = plaintiff's (defendant's) subjective probability that a complaint that is asserted will be tried rather than settled |
| p_{vp} (p_{vd}) | = plaintiff's (defendant's) subjective probability of a plaintiff victory at trial |
| q_p (q_d) | = victim's (injurer's) subjective probability that an accident will occur and the victim will assert a claim |
| S_p (S_d) | = subjective value to plaintiff (defendant) of settling the case rather than going to trial |
| T_p (T_d) | = subjective value to plaintiff (defendant) of possible damage award by the court |
| x_p (x_d) | = plaintiff's (defendant's) precaution against harm that gives rise to legal disputes |

and, therefore, outside a market context. This rules out "Coasian" bargaining and the possibility that prices convey information to the parties. There is a single injurer, who becomes a defendant, and a single victim, who becomes a plaintiff, when a suit is filed.² Both parties can affect the probability of an accident occurring. Initially, it is assumed that each party bears its own litigation costs. This assumption will be relaxed when we analyze alternative rules for allocating litigation costs.

We will forgo chronology and discuss the four stages of the litigation process in reverse order. This allows us to emphasize the their interdependence and, in particular, the fact that a decision at each point in the process depends crucially on the parties' expectations about the future. The variables in Table 2 above will be used in the analysis.

2.1 Behavior at Trial

2.1.1 Trial Effort of Plaintiffs and Defendants

The plaintiff goes forward with the trial because he expects to win something from the defendant. The value of this transfer depends on the intrinsic merits of the case, which is determined in part by the relevant laws and in part by the particulars of the case. The pertinent laws may describe the burden of proof, the legal standard of evidence, the scope of damages (including a possible augmentation of compensatory damages reflecting punitive

²When the suit is filed by a regulatory agency, or by lawyers on behalf of a class of victims, the analysis becomes more complex because the objectives of the active plaintiff may differ from the victims'.

damages in tort cases or treble damages in antitrust cases), and the rules of procedure. The relevant facts might describe the past actions of plaintiff and defendant, and the particular circumstances that determine the application of legal rules. These actions would include the levels of precaution in a tort suit, promises in a contract suit, facts of ownership in a trespass suit, etc.³

The amount that the plaintiff expects to win is determined not by the merits of the case alone, but also by the efforts the parties devote to winning. The efforts of the parties can be measured by expenditures on the trial, denoted c_{tp} and c_{td} for plaintiff and defendant, respectively. The subjective expected trial payoff to the plaintiff is this given by the function

$$T_p(c_{tp}, c_{td})$$

The plaintiff's cost of proceeding with the trial must be set against his expected winnings. The plaintiff's expected gain from bringing suit, prior to trial and net of trial costs, is thus given by

$$T_p(c_{tp}, c_{td}) - c_{tp} \quad (1)$$

Similarly, defendant's subjective expected loss, including trial costs, is given by

$$T_d(c_{tp}, c_{td}) + c_{td} \quad (2)$$

The partial derivative $\partial T_p / \partial c_{tp}$ can be thought of as the marginal productivity of plaintiff's effort at trial. If effort is productive, then $\partial T_p / \partial c_{tp} > 0$ and $\partial T_p / \partial c_{td} < 0$, and similarly for defendant.

The expenditure of effort at trial can serve an important signaling function. The court must decide cases in which the defendant's liability in civil suits or guilt in criminal ones is uncertain, because information about the law or the facts is incomplete. The parties to the dispute usually know more than the court about crucial facts and transmitting the information to the court is costly. Thus the effort that a party puts into trial provides a signal to the court. A stronger signal increases the probability that the judge or jury will favor the facts as represented by its sender.⁴

While effort is typically productive for both parties, the relative productivity depends on the merits of the case in a complex way. For example, if the defendant is negligent, effort by the plaintiff to discover and prove the facts can be very productive. But, on the other hand, effort by the defendant to represent the facts differently could also be productive.

The variables c_{tp} and c_{td} are chosen by the litigants as part of their trial strategy. The plaintiff chooses c_{tp} to maximize his expected gain, while the defendant chooses c_{td} to minimize his expected loss. The first-order condition for the plaintiff is:

$$\frac{\partial T_p}{\partial c_{tp}} + \frac{\partial T_p}{\partial c_{td}} i_p = 1 \quad (3)$$

³The parties' perception about the law and the facts may differ from each other and may be inaccurate as well.

⁴This model is developed in Rubinfeld and David Sappington in the criminal context in which there is a single defendant.

where $i_p = dc_{td}/dc_{tp}$ is the plaintiff's conjectural variation – a measure of how the defendant's costs will change in response to the change in plaintiff's cost. Equation (3) tells us that the plaintiff will expend money at trial so that the marginal benefit from more effort is equal to the marginal cost. The decision about how much money to spend during trial thus depends on strategic considerations related to i_p . A similar condition applies to the defendant.

A number of legal variables that can influence trial effort have been explored in the literature. Consider, for example, the effect of an upward adjustment to compensatory damages on the effort of both parties. To make such an evaluation consider a game in which plaintiff and defendant take each others' effort as fixed (as in a Nash game). Then $i_p = 0$, so that equation (3) reduces to $\partial T_p/\partial c_{tp} = -1$. An upward adjustment to compensatory damages should increase the marginal productivity of plaintiff's effort ($\partial T_p/\partial c_{tp}$). It follows that c_{tp} will increase (as will c_{td} , under comparable assumptions). Thus, as the stakes increase, both parties will increase their trial expenditure.

If strategic behavior occurs, so that $i_p \neq 0$ and/or $i_d \neq 0$, the analysis is more complicated. Even when conjectural variations are not zero, we would expect a positive relationship between the upward adjustment to compensatory damages and the effort of both parties. But, without further structure, we cannot rule out the possibility that strategic behavior will lead to a contrary effect. Suppose, for example, that the game is sequential and only three levels of effort are possible – high, medium, and low. Initially, when only compensatory damages are given by the court, both parties choose a medium level of effort. Subsequently, when damages are increased, the plaintiff chooses first and opts for a high level of effort. Now, it may be in the defendant's interest to opt for a low effort level, realizing that he's unlikely to win the case whatever his choice. In this example, augmenting compensatory damages led one party to make more effort and the other party to make less.

The rule for allocating legal costs is another variable whose effect upon trial effort has been explored. In the American legal system, the parties to a dispute usually bear their own legal costs. In the UK, however, the loser in a trial bears the legal costs of the winner. Theoreticians have compared the effects of these two rules. John Hause uses a model in which the probability that plaintiff prevails at trial is a function of the legal expenditures of both parties. He concludes that a switch from the system in which both parties pay their own legal fees to one in which the loser pays the winner's costs would increase trial expenditures for those suits that go to trial. Avery Katz also takes interactions and strategic considerations into account when he models trial effort. He reaches essentially the same conclusion as Hause, arguing that higher stakes raise the marginal value of additional expenditures to both parties. At the same time, the possibility that the losing side will pay part of the winner's litigation expenses lowers the expected marginal cost of litigating the case.⁵

⁵Other trial effort papers include Ronald Braeutigarn, Bruce Owen, and John Panzer.

2.1.2 The Outcome of the Trial

The outcome at trial (a win for plaintiff or defendant) is the result of a complex interaction between the efforts that both parties put into the trial and the facts and law of the case. If both parties are only interested in winning the stakes in this trial, rather than being interested in the law or reputation on future disputes, then the levels of effort chosen and trial outcomes will depend on the relative productivities of both parties.

In many cases, however, parties are likely to engage in similar litigation in the future – so that a repeated game framework becomes more appropriate. When one or both parties is concerned about the future, the probability that the plaintiff will win may increase or decrease. Typically, the probability of winning will increase for the party with a future interest in victory. To see why, consider a Nash game in which the parties initially have a 50 percent chance of winning. Now suppose that the defendant acquires a future interest in victory, so that the cost of a trial judgment increases by a multiple m , where $m > 1$. It follows from equation (3) that the defendant’s expenditures on trial, which were formerly determined by $\partial T_p / \partial c_{td} = -1$, are now given by $\partial T_p / \partial c_{td} = -1/m$. The defendant’s trial effort will consequently grow, and the probability of defendant’s victory will increase to a level above 50 percent.⁶

Jeffrey Perloff and Rubinfeld have suggested that defendants typically have more at stake than plaintiffs because defendants are likely to be involved in future litigation of the same type.⁷ In this situation, the loss to the defendant is greater than the plaintiff’s gain. The defendant will, consequently, choose to spend more on trial than the plaintiff and will, therefore, have a greater than 50 percent chance of winning. Using an antitrust dataset, Perloff and Rubinfeld find support for this view, because approximately 70 percent of all antitrust cases in their data set are won by defendants. This percentage is substantially higher than the rate of defendant victories in the cases studied by Priest and Benjamin Klein.

Settlement bargaining is a filter and the small percentage of cases that pass through it and go on to trial are not a random selection of all suits. Consequently, the frequency with which plaintiffs win at trial depends on the nature of the selection process. Hypotheses about the selection process and the frequency of plaintiff victory have been advanced and studied empirically. These hypotheses all build on the view that cases fail to settle as a consequence of a mistaken prediction about the outcome of a trial made by one of the parties. If, for example, the predictions of defendants and plaintiffs are normally distributed around the true mean, each party is equally likely to make a mistake. Priest and Klein use such an argument, supported by data, to conclude that cases go to trial in which defendants

⁶See Rubinfeld. This is not an equilibrium argument, because the plaintiff will respond to the defendant’s increase in effort, but the direction of change should be the same in equilibrium as in the first round of responses, provided that the reaction functions have the expected shape.

⁷Econometric evidence also confirms that juries will award greater damages when there are corporate defendants, all else equal (James Hammitt, Steve Carroll, and Daniel Relles). Jury awards are discussed more generally by Mark Peterson, Peterson and George Priest, and Mark Shanley and Peterson.

and plaintiffs each have a 50 percent probability of winning (see also Priest).

Donald Wittman replies that when the parties disagree about expected trial award the 50 percent rule of Priest and Klein can be seriously biased. Disagreement concerns in part the meaning of "winning", which is ambiguous in the context of trials. The plaintiff "wins" a civil suit, in one sense of the word, if a court awards damages or provides injunctive relief. Many civil suits, however, concern not the fact of defendant's liability but its extent. From this perspective, the plaintiff "wins" at trial only if the damage award is larger than the defendant's settlement offer.

A further ambiguity arises when one of the parties to a dispute has a future interest in the trials's outcome. An interest in reputation or precedent by one of the parties makes the stakes asymmetrical. Even if the 50 percent rule were true when the parties are symmetrically situated, it will not be when there are asymmetries. To illustrate, a defendant who wants to cultivate a reputation for tough bargaining will contest cases that he has little chance of winning. Conversely, a defendant who wants to avoid the publicity of a trial will settle cases that he has a high probability of winning.

2.2 Settlement Versus Trial

The economic issues surrounding whether suits are settled or brought to trial have a long history in the law and economics literature. The early literature, including work by Landes, John Gould, Posner, and William Baxter, treated the private incentives of the parties, while Shavell went further by distinguishing private from social incentives. Most of the more recent literature on the economics of settlements has moved toward a game-theoretic framework in which there are information asymmetries and a variety of sequences by which settlement offers are made by one or both parties. In this subsection we treat the parties' incentives, and then briefly survey the theoretical results concerning the effect of changes in policy instruments when the parties behave strategically.

2.2.1 The Incentives of Plaintiffs and Defendants

In some legal disputes there is scope for settlement, whereas in others trials may be inevitable. To distinguish between them, consider a civil dispute in which the parties have no future interest, so the bottom line is how much defendant pays plaintiff. The parties have expectations about the size of the transfer that would result from a trial and its cost. Plaintiff's expected gain from going to trial, net of trial costs, is given in equation (1) above, while defendants's expected loss, including trial costs, is given in equation (2) above. These expected gains and losses represent the subjective threat values of the parties. Any change that strengthens one player's threat value should increase his gains from the bargain. For example, Hugh Gravelle shows that plaintiffs with smaller risk aversion will receive larger settlements in a model in which courts have imperfect information.

The sum of the subjective threat values equals the players' assessment of the games's noncooperative value:

$$\begin{aligned} \text{Noncooperative value} &= (T_p - c_{tp}) - (T_d + c_{td}) \\ &= (T_p - T_d) - (c_{tp} + c_{td}) \end{aligned}$$

If a trial can be avoided, the parties must still bear the transaction costs associated with settlement, which are denoted c_{sp} and c_{sd} for plaintiff and defendant, respectively. In a settlement, the net transfer necessarily equals zero. The cooperative value of the game thus equals the actual net transfer (zero) less the transactions costs incurred:

$$\text{Cooperative value} = -(c_{sp} + c_{sd})$$

The difference between the cooperative and noncooperative values of the game equals the surplus;

$$\text{Surplus} = (c_{tp} + c_{td}) - (c_{sp} + c_{sd}) + [T_d - T_p] \quad (4)$$

The surplus from cooperation equals the sum of the term in braces, representing the difference in the costs of trial and settlement, and the term in brackets, representing the difference in subjective expectations about damages awarded at trial.

Transactions costs are less when a case is settled than tried:

$$(c_{tp} + c_{td}) - (c_{sp} + c_{sd}) > 0$$

Indeed, trial costs are so much greater than settlement costs that many authors choose the simplifying assumption that settlement costs are nil, that is, $c_{sp} = c_{sd} = 0$. In this case, the surplus reduces to the gap of expectations of the parties concerning the value of trial:

$$\text{Surplus} = (c_{td} + c_{sp}) + (T_d - T_p) = (T_d + c_{td} - (T_p - c_{tp}))$$

For a risk-neutral plaintiff, the subjective value of the possible damage award at trial, denoted T_p , equals the money value of expected damages, D_p , times the subjective probability of their award, denoted p_{vp} , that is, $T_p = p_{vp}D_p$. Similarly, for a risk-neutral defendant, $T_d = p_{vd}D_d$. When plaintiff and defendant have the same expectation about trial ($p_{vp} = p_{vd}$ and $D_p = D_d$), they concur about its expected value, so that $T_p = T_d$. If the parties are relatively pessimistic about the prospects at trial ($p_{vp} < p_{vd}$ and $D_p < D_d$), plaintiff will expect to win less than defendant expects to lose, so that $T_p > T_d$.

If the surplus is negative, the disputants prefer a trial to any possible settlement, so trial is inevitable. If the surplus is positive, however, there is scope for settlement out of court. The frequency of settlements presumably increases with the magnitude of the surplus. There is more scope for settlement when litigation is costly (c_{tp} and c_{td} are large), negotiations are inexpensive (c_{tp} and c_{sd} are small), and the disputants are pessimistic about trial outcomes ($p_{vp} < p_{vd}$, $D_p < D_d$). As a result, any policy that increases litigation

costs. lowers settlement costs. or makes disputants pessimistic about their trial prospects, will increase settlements.

Now consider the effect of risk aversion upon litigants with the same information about possible outcomes of trials. A trial represents a gamble, so the subjective value to risk-averse disputants will diverge from its expected value. For example, when the parties are both risk-averse and they have the same expectations about trial, their subjective values of trial diverge:

$$T_p < p_{vp}D_p = p_{vd}D_d < T_d$$

Risk aversion thus increases the surplus as given in equation (4), which presumably increases the probability of a settlement. Notice that risk aversion increases the surplus in the same way as pessimism – by increasing the difference between the subjective values of plaintiff’s trial gains and defendant’s trial losses.

2.2.2 The Effects of Legal Rules

Most models have assumed that settlement occurs automatically whenever the surplus in equation (4) is positive. This assumption has the effect of ruling out strategic behavior. Its main justification is pragmatic – predictions can be derived readily from nonstrategic bargaining models, whereas strategic models are often intractable. Given the fact that the term in braces $(c_{tp} + c_{td}) - (c_{sp} + c_{sd})$ is positive, and assuming nonstrategic bargaining, the trial/settlement split falls into two zones determined by the sign of the surplus, with one intermediate point:

$$\begin{aligned} [T_d - T_p] > -(c_{tp} + c_{td}) - (c_{sp} + c_{sd}) &\Rightarrow \text{settlement} \\ [T_d - T_p] > -(c_{tp} + c_{td}) - (c_{sp} + c_{sd}) &\Rightarrow \text{tipping point} \\ [T_d - T_p] > -(c_{tp} + c_{td}) - (c_{sp} + c_{sd}) &\Rightarrow \text{trial} \end{aligned}$$

These relationships help to generate a prediction about the effect of treble damages and punitive damages upon the frequency of trials. Consider how augmenting damages affects a case at the tipping point between settlement and trial. The fact that the term in braces is positive implies that $[T_d - T_p] < 0$ at the tipping point. Augmenting damages increases the absolute value of this negative magnitude, which tips the case into the trial zone, so there are more trials and fewer settlements. Augmenting damages in a nonstrategic bargaining model thus strengthens the tendency of optimism to cause trials.

This conclusion must be modified once account is taken of the resulting change in trial effort. Augmenting damages increases the stakes of the trial, which typically elicits more effort at trial by the parties, as explained above. With more effort, the term in braces $(c_{tp} + c_{td}) - (c_{sp} + c_{sd})$ increases in value. The resulting increase in the surplus from cooperation presumably makes settlement more likely.

Risk aversion also affects the comparison. Augmenting damages, by increasing the stakes at trial, makes trial more risky, which makes trial less attractive to risk-averse

disputants. Risk-averse disputants at the tipping point under a regime of compensatory damages may be nudged into settlement by a change to a regime of augmented damages because trial has become too risky.

In sum, augmenting damages increases the stakes at trial, which has the opposing effects upon the ability to settle out of court in a nonstrategic model. On one hand, more weight is given to the parties' optimism, which tends to increase the frequency of trials. On the other hand, trials become more costly and more risky, which tends to decrease their frequency.

Similarly, changing the rule for distributing trial costs has the opposing effects upon the ability to settle out of court. Under the American rule, the parties know with certainty that they will pay their own costs, and under the UK rule the loser pays all the plaintiff is expected to bear trial costs, $c_{tp} + c_{td}$, only if he loses, which occurs with probability $(1 - p_{vp})$. Assuming risk neutrality, we can modify equation (4) to contrast the two rules:

$$US = (c_{tp} + c_{td}) - c_{sp} + c_{sd} + [p_{vd}D_d - p_{vp}D_p] \quad (5)$$

$$UK = (1 - p_{vp} + p_{vd})(c_{tp} + c_{td}) - (c_{sp} + c_{sd}) + [p_{vd}D_d - p_{vp}D_p] \quad (6)$$

Equation (6) reduces to (5) when the parties have the same subjective beliefs about the probability of plaintiff's victory, $p_{cp} = p_{vd}$, but not otherwise.

Consider the effect of the change in rules on a case at the tipping point between settlement and trial under the American rule. The surplus is zero at the tipping point, so (5) = 0 by assumption. The term in braces in equation (5) is positive by assumption. Hence, $p_{vp}D_p > p_{vd}D_d$, which indicates that the parties are optimistic about trial. Assume that this optimism extends to the expectations about the probability of plaintiff's victory, so that $p_{vp} > p_{vd}p_{vp} > p_{vd}$ implies (5) > (6). This fact and the fact that (5) = 0 imply that (6) < 0. Thus, a switch to the UK rule causes the surplus at the tipping point to turn negative, resulting in more trials.

This conclusion must be modified when trial effort and risk aversion are considered. The switch in cost distribution rules from US to UK increases the stakes at trial by including trial costs in the gamble. The effect of higher stakes upon trial effort and risk aversion has already been discussed – the effect of a switch in the distribution rule for trial costs parallels the effects of augmenting damages.⁸ Theory tells us, therefore, that a switch from the rule of each pays his own (American) to the rule of loser pays all (British), tends to increase the frequency of trials by giving more weight to the parties' optimism, and to decrease it by making them more costly and more risky.

The direction the overall effect upon the frequency of trials from changing the cost distribution can, therefore, not be determined by theory alone. The common belief among lawyers that fewer suits will occur when the loser pays more of the legal costs enjoys

⁸Katz compares trial efforts of the parties under the British and American rules. See also Braeutigam, Owen, and Panzar, Shavell, and Posner

some support from empirical economics.⁹ This belief has motivated a modification of the American rule to more closely resemble the British rule. Geoffrey Miller considers one prominent example. Suppose plaintiff rejects the defendants' final offer to settle for a specified sum of money and that, after a trial, plaintiff is awarded less than the final settlement offer. Under these circumstances the plaintiff can be said to have lost in court relative to the settlement offer. Rule 68 of the code governing procedure in Federal courts, which is similar to procedural rules in a variety of states, specifies that a plaintiff who loses in court relative to the defendant's settlement offer must pay some of the winner's court costs, including such items as the cost of depositions and filings, and excluding attorneys' fees.¹⁰ Whatever effect this rule has on the frequency of trials, it strengthens the bargaining position of defendants.

2.2.3 Strategic Aspects of Settlement Behavior

The nonstrategic bargaining model in the preceding section assumes that disputes will always settle out of court when the cooperative surplus, as perceived by the players, is positive, whereas trials will occur when it is perceived as negative. There is, however, another case of trials – the distribution problem itself. The problem of dividing the surplus created by settlement is a source of instability that can lead to bargaining breakdowns.

Attempts by theorists to model the distribution problem in bargaining games in general have produced not a consensus among economists, but a variety of predictive and normative theories that rival each other. The unsatisfactory state of bargaining theory is reflected in strategic models of the litigation process. One approach to settlement bargaining generates definitive predictions by making restrictive assumptions about the scope of bargaining, the timing of offers, and the ability of the parties to transmit information. Thus, in Janusz Ordover and Ariel Rubinstein and Ivan P'ng, the settlement amount is fixed and not open to bargaining. In P'ng the defendant knows whether he is negligent and uses this information to decide whether to make a settlement offer, whereas plaintiff responds without knowing for certain whether the defendant was negligent. In Lucian Bebchuk, the settlement amount is endogenous, but the plaintiff knows the actual harm and the defendant knows only the probability distribution of possible harms. In Bebchuk's model, the relatively uninformed plaintiff makes the first and only settlement demand, which the defendant must either accept or reject in favor of a trial. The response of defendant to plaintiff's offer conveys some information about the defendant, but uncertainty persists, so cases go to trial.

The models discussed so far do not face the distribution problem squarely. When bargaining over distribution of the cooperative surplus, the players are uncertain about the extent to which other parties will concede. A rational player will gauge his demands such

⁹Donald Coursey and Linda Stanley found a higher settlement rate when their experimental subjects decided disputes under Rule 68 than under the British Rule

¹⁰The Supreme Court extended Rule 68 to cover attorneys' fees in cases where the statute under which the action is brought allows recovery of attorneys' fees. See *Merrick v. Chesney*, 473 U.S. 1, 105 Ct. 3012

that the gain from settling on slightly more favorable terms is offset by the increased risk of negotiations breaking down. Thus, the optimal strategy in settlement bargaining balances a larger share of the stakes against a higher probability of trial. When these considerations are balanced at the margin, expected utility is maximized relative to the distribution of an adversary's possible strategies. A bargaining equilibrium can thus be characterized as a situation in which everyone maximizes expected utility given complete knowledge about the distribution of strategies followed by others.¹¹ This equilibrium concept has the advantage of permitting strategic behavior to cause trials.

While the optimal strategy, as characterized above, is best relative to the distribution of the other party's strategy, it is not necessarily best against the actual strategy that will be chosen. A party may overestimate a particular opponent's willingness to make concessions, which can cause a breakdown in settlement negotiations and a trial. To illustrate, suppose that the parties must choose between a hard strategy (make no concessions) and a soft strategy (concede). Each party knows the frequency with which these strategies are chosen by others, but no one finds out his particular adversary's strategy in a specific dispute until after it is resolved. Trials occur under these circumstances when both parties commit to hard strategies, and settlements occur otherwise. To illustrate, suppose that in equilibrium 30 percent of plaintiffs and 30 percent of defendants are pursuing hard strategies. Then 9 percent of disputes end in trial and 91 percent settle out of court, and no one is surprised by these proportions.

This equilibrium concept presupposes some means by which the parties generate their expectations about the probability that other players will concede. A full account of the genesis of concessionary expectations would go beyond the legal process into psychology and sociology. For the purposes of economic analysis, however, it is usually sufficient to assume that disputants have expectations prior to beginning a legal dispute, and then to predict how the legal process modifies them. An earlier example of this approach by Cooter, Stephen Marks, and Mnookin sought for conditions under which the predictions of the nonstrategic models could be extended to strategic bargaining. The specification of the information structure of the game is not adequate in this early work, but has been corrected in subsequent work. In William Samuelson both parties make settlement offers simultaneously, so uncertainty persists and bargaining can fail. Stephen Salant assumes that plaintiffs come in two types, slightly injured and badly injured, and defendants cannot tell them apart in pretrial bargaining.

Several papers have applied the concept of sequential equilibria (David Kreps and Robert Wilson) to settlement bargaining, notably Urs Schweizer, Jennifer Reinganum and Louis Wilde, and Barry Nalebuff for civil suits, and Reinganum for criminal cases. At each node or state of the game, each party chooses the strategy that is optimal for the remainder of the game, given uncertainties about other players and their future actions. Parties update their beliefs in light of information provided at each stage of the game.

¹¹For an analysis of this Bayesian-Nash equilibrium, see John Harsanyi

To generate definite predictions, these approaches must exploit facts about the litigation process that prescribe sequences of moves and generate asymmetric information. Thus, plaintiff must make the first move to assert a legal claim. In settlement bargaining for civil disputes, the defendant often has more information concerning the existence of liability (e.g., whether negligence can be proved), and the plaintiff has more information about the extent of liability (e.g., how severe was the injury).

If enough structure is imposed to generate sequential moves with asymmetric information, some predictions can be derived that may contradict the nonstrategic models. To illustrate, an important topic in strategic bargaining is the information transmitted by the exchange of offers. Signaling in settlement bargaining was studied by Nalebuff, who relied upon information asymmetries to generate predictions about equilibria in a sequential subgame. In the first step, plaintiff makes a single demand; next the defendant either rejects the demand or settles the case; finally, if the demand is rejected, the plaintiff decides whether to proceed to trial. Plaintiff's demand in the first stage conveys information to defendant about the probability that plaintiff is prepared to proceed to trial. Defendant's rejection of plaintiff's demand in the second stage conveys information to plaintiff about the strength of defendant's case. In equilibrium, plaintiffs know the distribution over the strength of the case of defendants who settle, and defendants know the proportion of cases that plaintiffs litigate.

A comparison of Bebchuk, Nalebuff, and our hybrid model illustrate that different specifications of the game affect important predictions about the litigation process. Consider the effect of an increase in plaintiff's trial costs on the terms of settlement. An increase in plaintiff's trial costs weaken plaintiff's threat position, which leads to lower settlement offers in Bebchuk's model and our hybrid model. While agreeing with this general argument, Nalebuff points to an alternative possibility. He argues that when trial costs increase, plaintiff will not be prepared to go to trial unless he expects to win a larger judgment. To make the threat of going to trial credible, he will demand a larger settlement. A full specification of the information structure in settlement bargaining, including the signal contained in the offers they make, may thus lead to predictions that contradict the nonstrategic model.

2.2.4 Empirical Studies of Trial/Settlement Split

Courts have been studied by sociologists and other social scientists from both a longitudinal and a cross-sectional point of view. However, it is relatively recently that economists have begun the task of specifying and estimating structural models of the behavior of the parties during the dispute resolution process. The greatest attention has focused on the settlement decision. A satisfactory model of settlement must take account of uncertainty in settlement bargaining, which results in specification errors, and the possibility that bargaining breaks down due to strategic behavior.

A structural model for empirical research on the trial/settlement split can be developed

from our hybrid model. Suppose we posit that the plaintiff's expected gain from trial consists of the systematic component T_p and a randomly-distributed error. Similarly, a defendant's expected loss from trial consists of a systematic component T_d and a randomly-distributed error. Trial costs may be random as well. Combining all these terms, the cooperative surplus from settlement becomes

$$\text{Surplus} = (T_d - T_p) + (c_{td} + c_{tp}) + \epsilon \quad (7)$$

where G is the systematic component and ϵ is a random disturbance term. In this framework, a dispute may fail to settle even though the systematic component of the cooperative surplus is positive, provided that the error is large and negative.

A reduced-form model can be obtained from equation (7) in which the probability of trial, denoted p_t , is determined by evaluating the probability distribution function of the systematic component of the settlement surplus:

$$p_t = p_t(G) \quad (8)$$

A question investigated empirically by using equation (8) is whether augmenting compensatory damages will result in more or fewer trials. Recall our previous discussion in which we concluded that, when damages are augmented, the tendency of optimism to cause trial is strengthened, the tendency of risk aversion to discourage trials is strengthened, and more costly effort is elicited to win trials that occur, which further discourages trials. Perloff and Rubinfeld found evidence suggesting that in antitrust cases, where reputational effects are important, and where the parties tend to be pessimistic ($T_d - T_p > 0$), treble damages lead to a decrease in the proportion of cases resolved by trial and an increase in the number of settlements.

Other studies, however, have suggested a contrary result, including Danzon and Lee Lillard . They applied a model of the settlement process to medical malpractice claims. Their model consists of two trial equations that explain the probability of plaintiff winning ($p_{vp} = p_{vd} = p$) and the amount of verdict ($T_p = T_d = T$), and two settlement equations that explain the minimum demand of the plaintiff ($T_p - c_{tp}$) and the maximum offer ($T_d + C_{td}$) of the defendant. Both the minimum demand and maximum offer depend positively on the perceived probability of winning and the perceived verdict. As in our hybrid model, the authors assume that cases will settle when the minimum asking price of the plaintiff is greater than zero , but less than the defendant's maximum offer. When the minimum asking price is greater than the maximum offer, the case will go to trial.

Danzon and Lillard assume that an increase in the stakes involved in the case (brought about when damages are augmented) will increase random errors proportionally. The costs of litigation, however, increase less than proportionally. So G in equation (7) is negative in more cases, and more cases will be litigated. ¹²

¹²Posner comes to the same conclusion as Danzon and Lillard but makes the assumptions that the costs of litigation are fixed, and that the parties disagree only about the probability that the plaintiff will win at trial.

Among the interesting results of this study are the following: (i) the higher the award at trial, the greater the probability that the case will go to trial; (ii) the higher the plaintiff's probability of winning at trial (as perceived equally by both parties), the lower the probability that the case will go to trial; and (iii) plaintiffs win only 28 percent of the cases that go to trial.

2.3 Assertion of a Legal Claim

A dispute is initiated when a party with a complaint asserts it, either formally by filing the required legal document, or informally by private communications between the parties. Some legal disputes are settled privately and never come to the court's attention. In other cases, such as tortious injuries to minors, the resolution of the dispute is not legally binding until approved by the court. In the best-documented disputes, however, an official complaint is filed by the plaintiff against the defendant.

The decision to assert a legal claim is difficult to investigate empirically because cases that do not come to the attention of judicial authorities never enter official records. Danzon and Lillard partly avoided this problem by studying insurance records. In two data sets of medical malpractice, they found that 50 percent of cases were resolved before a suit was filed, 40 percent were settled before a verdict, and 10 percent were tried to a verdict.

Asserting a complaint, whether informal or official, uses plaintiff's time and/or money. The expected benefit of asserting a legal claim consists of the possibility of settlement or a favorable court judgment. Shavell and Posner, among others, have assumed that rational decision makers assert a complaint because the cost of doing so is less than the expected benefit. Let c_{cp} denote the cost to plaintiff of asserting a legal claim, and let L_p denote the plaintiff's subjective expected benefit.

The expected benefit of asserting a legal claim can be determined explicitly from preceding sections of this article. The plaintiff's subjective value of a possible court judgment, conditional upon a trial occurring, has been written as $T_P(c_{tp}, c_{td})$. Let p_{tp} denote the plaintiff's subjective probability that a complaint will eventually lead to a trial. The plaintiff's (unconditional) subjective value of a court judgment that could result from asserting a complaint, net of litigation costs, is thus $p_{tp}[T_P(c_{tp}, c_{td}) - c_{tp}]$. Similarly, let S_p denote the subjective expected value of settlement for plaintiff, conditional upon a settlement being reached, which occurs with probability $1 - p_{tp}$. The plaintiff's subjective expected value of the legal claim, L_p , is thus

$$L_p = p_{tp}[T_P(c_{tp}, c_{td}) - c_{tp}] + (1 - p_{tp})(S_p - c_{sp}) \quad (9)$$

Equation (9) implies that claims are more valuable to a victim when litigation costs and bargaining are inexpensive (low c_{tp} and c_{sp}), and plaintiff is optimistic about his prospects at trial or settlement (high T_p and S_p). Further, the plaintiff knows that a settlement will occur only if it makes him better off than going to trial. As a result, L_p must be decreasing in P_{tp} .

The decision rule for the rational plaintiff balances the subjective value of a legal claim against the cost of asserting it:

$$c_{cp} < L_p \rightarrow \text{Assert a legal claim}$$

$$c_{cp} = L_p \rightarrow \text{Tipping point}$$

$$c_{cp} > L_p \rightarrow \text{Do not assert a legal claim}$$

The literature is divided, however, on the appropriate measure of L_p for the rational plaintiff. Suppose both parties have complete information about trial costs and outcomes and certain other conditions in dispute are met. Then suits will be brought only when the plaintiff's expected benefit from trial net of trial costs is positive. Settlement costs, probabilities, and amounts are irrelevant under these conditions. This can most easily be seen in a model of repeated litigation in which both parties know the plaintiff's expected net benefits from trial to be negative. Then it will be in defendant's interest not to agree to settle such a case and, consequently, plaintiff will not choose to bring the case in the first place. Thus, when both parties have complete information, the settlement probability is zero, and L_p is equal to the plaintiff's expected net benefit from trial.

Bebchuck has shown, however, that under a different assumption involving asymmetric information, the victim's decision to sue may depend on the likelihood and/or the magnitude of a settlement. In this framework the more general definition of L_p applies.

To analyze the relationship between legal costs and legal disputes, consider a person at the tipping point of asserting a legal claim, where $c_{cp} = L_p$. A change in the law that increases trial costs will immediately lower the value of l_p . The equation will thus tip in the direction $c_{cp} > L_p$, where the claim will not be asserted. More generally, laws that increase the costs of resolving disputes are likely to decrease the frequency with which legal claims are asserted and increase the cost of settling those that are asserted.

These conclusions have applied to the explanation of nuisance suits. A nuisance suit can be defined as a suit that both parties recognize as having no merit, in which case the expected damage award is nil: $T_p = T_d = 0$. thus the plaintiff's benefit from asserting a nuisance complaint from equation (9) reduces to:

$$L_p = (1 - p_{tp})(S_p - c_{sp}) - p_{tp}c_{tp} \tag{10}$$

The value of equation (10) obviously cannot be positive unless S_p is positive. It is irrational to file a nuisance suit unless the expected value of a possible settlement is positive. In general, the decision rule allows for the possibility of asserting claims whose expected trial value is nil only if their settlement value is positive.¹³

Why would a defendant pay damages to a plaintiff to settle a suit without merit? The answer offered in several models turns upon asymmetric costs. The central role of cost

¹³See also Bradford Cornell, who uses an option pricing model to suggest that victims will file some suits for which the net present value is negative.

asymmetries can be illustrated by applying the Nash bargaining solution to our hybrid model. The Nash bargaining solution gives each player his threat value plus half of the surplus from cooperation. Assuming risk neutrality, the general solution for plaintiff (for all suits) can be written:

$$\underbrace{p_{vp}D - c_{tp}}_{ThreatValue} + \underbrace{(1/2)[(c_{tp} + c_{td}) - (c_{sp} + c_{sd}) + (p_{vd}D_d - p_{vp}D_p)]}_{HalfofSurplus} \quad (11)$$

Consider the effect of asymmetric trial costs on equation (11). Assume that defendant's trial costs are greater than plaintiff's, $c_{td} > c_{tp}$, and assume that the players are symmetric with respect to settlement costs and information about trial, so that $c_s = c_{sp} = c_{sd}$, $P_v = p_{vp} = p_{vd}$, and $D = D_p = D_d$. Under these assumptions, equation (11) reduces to

$$p_v D + (1/2)(c_{td} - c_{tp}) - c_s/2 \quad (12)$$

The Nash bargaining solution under these assumptions requires that plaintiff's payoff, net of all costs, equal expression (12). The plaintiff pays his own settlement costs, so (12) will be satisfied if the defendant pays to the plaintiff a settlement amount, S , equal to the expected value of the trial judgment plus half the difference in trial costs:

$$S = p_v D + (1/2)(c_{td} - c_{tp}) \quad (13)$$

An important conclusion follows from equation (13): Assuming strict symmetry in information and costs (including $c_{td} = c_{tp}$), the Nash solution to settlement bargaining requires the defendant to pay the plaintiff the expected judgment from trial. Furthermore, assuming asymmetry in costs, the Nash solution to settlement bargaining requires the party who saves relatively more from avoiding trial to share these gains with the party who saves relatively less.

A precise prediction about the nuisance suits follows from equation (13). For nuisance suits, $P_v D = 0$ by definition, so $S = (1/2)(c_{td} - c_{tp})$. Thus, the bargaining solution between risk-neutral players requires that defendant refuse to settle nuisance suits in which trial costs are symmetric ($S = 0$ when $c_{td} > c_{tp}$).

The preceding model explains nuisance suits by asymmetries in the costs of defendant and plaintiff. An alternative explanation rests upon asymmetries in the timing of costs. For example, David Rosenberg and Shavell propose a sequential game in which plaintiff files a suit at a negligible cost. Following this, the defendant must either settle or incur litigation costs. Only after the defendant's action must the plaintiff either withdraw or incur costly litigation. So long as the defendant must expend effort on litigation prior to the plaintiff, the defendant might find a small settlement cheaper than litigation.

Bebchuk uses a slightly different model that focuses on the settlement process itself. He shows that nuisance suits can lead to a settlement when the defendant cannot be sure whether the plaintiff will go to trial or withdraw if there is no settlement. Finally,

Thomas develops a sophisticated model of strategic bargaining and shows that asymmetric information, not asymmetric costs, can lead to settlement of nuisance suits for a positive sum of money.

The condition under which the victim will assert a claim is also sensitive to fee arrangements that the victim makes with his lawyer. Under a contingent fee arrangement, the incentive to assert a claim is different from what it would be under an hourly fee arrangement. Under the former, the lawyer bears some of the client's risk in exchange for a portion of the proceeds if the victim receives an award at settlement or trial. Under the latter, the fee paid to the lawyer is independent of the victim's recovery. Danzon analyzes the effect of these two fee arrangements on the assertion of claims. She shows, for example, that a risk-preferring contingent fee attorney will accept a case that a risk-neutral hourly fee attorney would not take. However, it is also true that some claims that would be filed by a risk-neutral client using an hourly fee attorney would not be taken by a contingent fee lawyer.¹⁴

2.4 Precaution Against Harm

Our analysis has proceeded through three stages in a legal dispute in reverse chronological order, beginning with trial, followed by settlement bargaining, and then turning to assertion of legal claims. The fourth and final stage to consider is the harm that one person does to another. Harm can take many forms, such as tortious injury, breach of contract, and trespass upon property. Injurers and victims usually have access to forms of precaution that reduce the probability and severity of harm.

By far the greatest focus of economists who study common law rules has been on incentives created for precaution by injurers and victims. Coverage ranges from the article by John Brown (1973) in which alternative liability rules (e.g. strict liability, negligence, and comparative negligence) are compared in a model in which accidents are treated as externalities, to the article by Coase (1960) in which similar rules are analyzed in a framework in which injurers and victims bargain over the level of precaution that both parties take. The level of precaution is determined by these models by a profit-maximizing or utility-maximizing calculus in which the cost of precaution is traded off against its benefits, often in the form of reduced liability.

Only in a few instances, however, has the analysis taken explicit account of the relationship that is the subject of this section – incentives for precaution created by the litigation process itself (see, for example, Jerry Green and P'ng). Because litigation is expensive, expenditures on precaution will be made to reduce the probability and extent of litigation.

In extending our hybrid model to cover the relationship between litigation costs and incentives for precaution against harm, we proceed on the assumption that harm is an externality that cannot be cured in the market. In an externality model, harm done by the injurer affects the victim, and the victim's assertion of a legal claim affects the injurer.

¹⁴Other contingent fee studies include Herbert Kritzer et al. and Danzon.

However, there is no bargaining between the parties to allocate the costs of harm before it occurs. As a consequence, levels of precaution by the parties are determined by the legal assignment of liability.

In an externality model the injurer trades off the cost of additional precaution against the resulting reduction in legal claims. To formalize this optimization problem, the injurer's subjective probability that the victim will assert a legal claim, q_d , is assumed to be a decreasing, concave function of the injurer's precaution, denoted x_d , and other variables not made explicit: $q_d = q_d(x_d)$, where $dq_d/dx_d < 0$, $d^2q_d/dx_d^2 > 0$.

Let L_d denote the subjective expected cost to defendant of plaintiff's assertion of a legal claim. Analogous to equation (9), this expected cost is given by:

$$L_d = p_{td}[T_d(c_{tp}, c_{td}) + c_{td}] + (1 - p_{td})(S_d + c_{sd}) \quad (14)$$

From equation (14) it follows that more precaution by the injurer typically decreases the expected cost of legal claims.¹⁵ This relationship is assumed to be concave: $L_d = L_d(x_d)$, where $dL_d/dx_d < 0$, $d^2L_d/dx_d^2 > 0$. The injurer thus chooses precaution to minimize the sum of the costs of precaution and legal claims:

$$\min[x_d + q_d(x_d)l_d(x_d)] \quad (15)$$

Turning from injuries to victims, let x_p denote the victim's expenditure on precaution against harm caused by injurer, and let q_p and L_p indicate the probability and value, respectively, of the victim's subjective expectation concerning the assertion of a legal claim. To keep the analysis simple, assume that the victim asserts a claim if an accident occurs, but not otherwise.¹⁶ Thus q_p can be interpreted as victim's subjective probability of an accident, which is assumed to be a decreasing, concave function of victim's precaution (and other implicit variables): $q_p = q_p(x_p)$, where $dq_p/dx_p < 0$, $d^2q_p/dx_p^2 > 0$.

The potential victim who suffers harm equal to H_p receives a legal claim whose value is denoted L_p . The loss H_p is a concave, nonincreasing function of the victim's precaution (and other variables): $H_p = H_p(x_p)$, where $dH_p/dx_p \leq 0$, $d^2H_p/dx_p^2 \geq 0$. The expected value of the legal claim L_p is also a function of x_p : $L_p = L_p(x_p)$. The sign of the derivative

¹⁵Under every rule of law known to us, additional precaution by the injurer (weakly) decreases injurer's liability. To illustrate, under a negligence rule, more precaution by injurer increases the probability that the court will find the injurer not liable because he satisfied the legal standard of care. Similarly, under a rule of strict liability, more precaution by injurer reduces the magnitude of the expected damage award. Additional precaution by the injurer thus tends to reduce his expected cost of trial, T_d . Furthermore, lowering his expected cost of trial, T_d , strengthens his threat position in settlement bargaining, so expected settlement, S_d , tends to fall as well. More precautions may even reduce the probability, p_t , that disputes will end up in trial. (More precaution tends to reduce the severity of accidents, which reduces the stakes, and our previous analysis reached the tentative conclusion that larger stakes cause more trials.) Thus L_d is a nonincreasing function of x . (Strategic effects of larger x_d on plaintiff's choice of variables are not discussed here.)

¹⁶Thus, we are assuming that $c_{cp} < L_p$ for all accidents under consideration - the cost of asserting a claim is less than plaintiff's expected benefit L_p .

is not generally determinate, as can be seen from equation (9).¹⁷ The difference in value between the harm H_p and the legal claim L_p measures the victim's net loss. The potential victim thus chooses x_p to minimize the sum of his precaution costs and his net loss from harm:

$$\min[x_p + q_p(x_p)[H_p(x_p) - L_p(x_p)]](16)$$

A change in law that shifts the function $L_d(x_d)$ up, or increases its marginal value dL_d/dx_p , will, according with equation (15), induce more precaution from injurers. Conversely, a change in law that shifts $L_p(x_p)$ up, or increases its marginal value, dL_p/dx_p , will, according to equation (16), induce less precaution from victims.

As an example, consider the effect of augmenting compensatory damages. This will simultaneously shift $L_d(x_d)$ and $L_p(x_p)$ up. Assuming independence, this will lead to more precaution by injurers and less by victims. This illustrates the contrast between distribution and efficiency in courts. In general, therefore, compensation rules that effectuate transfers increase incentives for precaution by one party, but they reduce incentives for the other party.

Alternatively, consider a change in the law that increases the defendant's subjective probability that a given case will go to trial, that is, $s_d + c_{sd} < T_d + c_{td}$. It follows from equation (14) that L_d is an increasing function of the settlement probability. Consequently, as A. Mitchell Polinsky and Rubinfeld suggest, the change in the law will shift $L_d(x_d)$ up, thereby generating more injurer precaution, and greater deterrence.

Finally, suppose another change in law increases the plaintiff's costs of litigation. The value of legal claims, L_p , to plaintiffs, will decrease, thereby causing potential victims to take more precaution. In addition, higher costs of trials will cause plaintiffs to assert fewer claims. The cost of legal claims, L_d , to defendants can either increase or decrease as a consequence, depending upon whether the effect of fewer claims or more costly claims dominates. In general, legal rules that increase the cost of resolving disputes increase incentives for precaution by victims and may either increase or decrease incentives for precaution by injurers.

3 Normative Issues

Legal policy has traditionally been evaluated by standards of fairness, whereas the normative standard in most economic models is efficiency. Although efficiency is more controversial as a goal for law as opposed to markets, claims about efficiency have had significant impact on legal scholarship, teaching, and, possibly, on courts (see Jerome Culp). In this

¹⁷The plaintiff's subjective expected value of asserting a legal claim, L_p , equals the expected benefit, given by equation (9), less the cost of asserting the claim. The sign of the derivative of L_p with respect to x_p is indeterminate because more precaution by the victim generally reduces the magnitude of harm and increases the proportion of the costs of harm borne by the injurer. The magnitude of these effects, which go in opposite directions, depends upon the particulars of law and fact.

part of the article we will discuss some normative concerns that can be treated within an efficiency framework. The first issue deals with the fourth stage of litigation, but the remaining issues combine several stages. The last section goes further by examining the behavior of judges as lawmakers.

3.1 Trial Effort of Litigants - A Normative Analysis

Courts and other lawmakers have several policy instruments to affect trial effort, including (i) the legal standard of care,¹⁸ (ii) the magnitude of damages (Rubinfeld and Sappington), (iii) the burden of evidence production (Joel Sobel), (iv) the standard of proof (Rubinfeld), and v) court costs (Rubinfeld and Sappington). Each of these policy instruments directly affects the expenditure of parties at trial, and indirectly influences decisions at each stage of a legal dispute – trial, settlement, assertion of claims, and precaution. This section considers the direct effect of policy instruments on effort at trial and the outcomes of trials.

First consider the effect of trial effort. Define a court decision as correct if it would be reached by applying the law under conditions of full information.¹⁹ Instead of having full information, however, courts must make their decisions based on information provided largely by the disputants themselves. Effort by plaintiffs and defendants, and the rules governing evidence and procedure, determine a probability distribution of errors of Type I (finding violations where conformity occurred), and Type II (finding conformity where violations occurred).

It can be argued that more effort by both parties will disclose more information to the court, so its decision will come closer to the full-information decision. More information can thus reduce errors of both types. Furthermore, cases with large stakes induce more effort by both parties. This view leads to the conclusion that bigger cases are more likely to be decided correctly (see Posner).

One policy tool that typically induces greater trial effort by both parties is augmenting compensatory damages. The question of whether the increase in litigation effort from increasing damages has social value comparable to its cost has been investigated in several studies of treble damages in antitrust law (William Breit and Kenneth Elzinga; Steven Salop and Lawrence White). Polinsky and Rubinfeld show how the optimal damage level changes when costly litigation is taken into account.

A different perspective is provided by considering incentives that cause more effort to one party and less by the other. If effort is productive, then more effort by the party that deserves to win increases the accuracy of court decisions, whereas more effort by the party

¹⁸To our knowledge, the complex relationship between the standard of care and trial effort by plaintiffs and defendants has not been formally studied.

¹⁹Notice that a "correct" decision gives everyone their due under law. This is Plato's first definition of justice in the *Republic*. Hobbes takes the view in *Leviathan* that there is no other concept of justice (no justice in nature).

that deserves to lose has the opposite effect. In the context of criminal trials, Rubinfeld and Sappington argue that more effort by innocent parties decreases the probability of Type I errors (convicting the innocent), and more effort by guilty parties increases the probability of Type II errors (not convicting the guilty).

Some legal rules, such as "loser pays all", may increase the accuracy of court decisions by providing incentives for more effort by parties that deserve to win than parties that deserve to lose. This point was already discussed in connection with nuisance suits, where the British and American rules for allocating litigation costs were discussed. Thus, Marilyn Simon shows, assuming risk neutrality (but not otherwise), that a change from the American to the British rule reduces the probability of court error.

Instead of reducing both types of errors, some policy variables decrease one type while increasing the other. Thus, a shift in burden of proof from defendant to plaintiff might reduce Type I errors while increasing Type II errors. Identifying the "best" point in the frontier between the two types of errors involves perplexing normative issues.

Insight into these normative issues can be obtained from a game-theoretic framework. Suppose, as Sobel suggests, that both parties are bargaining with private information, but cannot publicly misrepresent that information. The judge has a prior distribution about the claims of the parties, but does not know their accuracy with certainty. In one equilibrium, each party has a positive probability of winning even if he does not provide evidence. In another equilibrium, a party wins only by presenting substantial evidence. In general the rules of the game and, in particular, which party has the burden of production of the evidence, will determine the equilibrium outcome.

Sobel shows, for example, with respect to this second equilibrium, that the overall cost of obtaining evidence is lowest if the burden of production is placed on the party that has the lowest cost of proving his claim. If the objective is to maximize the social value of the trial process, it is often better to place the burden of production on the party with the higher cost of providing evidence. If that party has a relatively weak claim, it will not present a case. But, if it has a relatively strong claim, the presentation will be worthwhile, despite the higher cost of evidence production.²⁰

A central normative issue in discussions of legal procedure is balancing the cost of additional information against the benefit of reducing cost errors. There is reason to wonder whether disputants value cumbersome procedural rules designed to produce accuracy as highly as do courts. Private systems of dispute resolution in which the parties choose their own rules, such as Visa's system of arbitration among member banks, typically employ much cheaper procedures than those adopted by public courts. Random inaccuracies are not too serious when the stakes are small relative to the disputants' wealth.

Unlike random inaccuracies in trial outcomes, which are unavoidable when information is costly, systematic inaccuracies have the appearance of bias. An alleged source of bias

²⁰Sobel's results are sensitive to the nature of the game and the assumptions concerning asymmetry of information. Ordover and Rubinstein, and Samuelson describe some game-theoretic perspectives that could lead to different conclusions.

is defendant's identity. Econometric evidence has confirmed the belief among lawyers that juries will award greater damages when defendants are corporations rather than individuals, all else being equal (see Hammitt, Carroll, and Relles; and more generally Peterson; Peterson and Priest; Shanley and Peterson). Perhaps courts impose rules that are too cumbersome in an effort to reduce random inaccuracy, whereas bias is a serious concern.

3.2 Private Versus Social Incentives to Bring Suit

Many disputes involve claims for damages, which can be resolved by transfers of income. In general, the rules for making these transfers affect the incentives of the parties subject to current and future disputes. It is not surprising, therefore, that the private and social value of suits may diverge.

Our hybrid model can be used to trace this divergence with respect to incentives to assert legal claims. Recall that plaintiff decides whether to assert a legal claim by balancing the subjective expected benefit from trial with the cost of filing the claim, which yields the tipping point $C_{cp} = L_p$, where

$$L_p = p_{tp}[T_p(C_{tp}, c_{td} - c_{tp})] + (1 - p_{tp})(S_p - c_{sp}) \quad (17)$$

Note that the defendant's litigation costs c_{td} , which are triggered by the assertion of a legal claim, are not borne by the plaintiff and do not figure directly in his decision to assert a legal claim. Shavell suggested that the private costs of asserting a legal claim are less than the social costs under the American system because the plaintiff does not bear the defendant's litigation costs. Thus a plaintiff who runs the gamble of asserting a legal claim externalizes part of the cost of finding out whether it is worthless or valuable. Shavell also noted, however, a consideration pointing in the opposite direction. When trial costs are substantial, the private net benefit from trial may be negative even though the social gain from deterring injurers is large.

Peter Menell countered Shavell's argument that the private costs of suit are less than the social costs by pointing out that when victims do not pay injurers' costs of resolving disputes, injurers may respond by taking additional precaution. The additional precaution may or may not be socially efficient. Louis Kaplow refined this argument by distinguishing between the effect of precaution on the extent of harm and its probability. Arguments about the divergence of private and social incentives to sue were subsequently synthesized by Susan Rose-Ackerman and Mark Geitsfeld.

3.3 Deterrence with a Costly, Uncertain, Litigation Process

Pioneering work on incentives for precaution, such as Brown (1973), compared the efficiency of alternative rules such as strict liability versus negligence for allocating the cost of harm.

These studies assumed that all harm is pecuniary,²¹ disputes can be resolved without cost, and courts apply clear legal standards without error. Our hybrid model will be extended to modify the conclusions when dispute resolution is costly and courts apply obscure standards or make errors.

A full extension would compare incentives for precaution by injurers and victims under alternative liability rules. For this article, however, the discussion will be restricted to the incentives for precaution under the rule of strict liability and then under a negligence rule. A basic conclusion of the early studies is that, assuming costless dispute resolution, the injurer's incentives for precaution under strict liability are efficient when the defendant must fully compensate the plaintiff. Full compensation is achieved when the victim is indifferent between avoiding the harm or suffering it and receiving compensation.

When the result is extended to the context of costly litigation, full compensation must include the cost of resolving the dispute, not just the harm that gave rise to it. Suppose disputes are resolved by trials. To achieve full compensation in our hybrid model under this assumption, a victim who is certain to win at trial must be compensated, not just for the harm H_p caused by the accident, but for his trial costs c_{tp} plus his costs of asserting a claim c_{cp} . If, in addition, the court sometimes makes errors, so that the injured plaintiff wins at trial with probability p_{vp} , full compensation requires setting the damage award D_p so that

$$D_p = \frac{H_p + c_{cp} + c_{tp}}{p_{vp}} \quad (18)$$

A rule requiring full compensation of victims by injurers causes the latter to internalize costs, which induces efficient precaution by them. American law, however, typically requires the injurer to compensate the victim for the harm that gave rise to the legal claim, but not for the victim's cost of resolving the dispute. The injurer who expects accidents to result in trials will, consequently, choose a level of precaution knowing that he must pay for his precaution, the expected harm, and *his* litigation costs.

The resulting externality will distort the injurer's precaution. It might appear that this element of externalized cost will always cost injurers to take too little precaution relative to the socially efficient level. In fact, this will be true if litigation is relatively costly and precaution is relatively expensive. (In this case the efficient level of precaution will be higher with costly litigation than without, because additional precaution reduces the expense of litigation substantially.)

However, Polinsky and Rubinfeld have shown that this level of externalized cost may result in a greater than efficient level of injurer precaution in cases when litigation is relatively inexpensive and precaution is quite costly. This surprising result occurs when additional injurer precaution substantially reduces the number of suits that victims bring, and thereby reduces the injurer's liability and his litigation cost. (Recall that victims only

²¹Nonpecuniary injuries, such as pain and suffering in tortious accidents, can affect the total utility without affecting the marginal utility of money. In this situation, costly compensation is inefficient.

bring suit when their expected benefit exceeds their litigation cost. Additional precaution by the injurer can tip many victims from the region in which they bring suit to the region in which they do not.)

A further qualification of the efficient standard is required when the outcome of litigation is uncertain and the rule of strict liability is replaced by a negligence rule. Early models showed that if courts set the legal standard of care equal to the efficient level of precaution, and if they apply this standard without error, injurers will exactly conform to the legal standard, as required for efficient precaution. Suppose, however, that courts make errors in applying a negligence rule, as a consequence of which some negligent defendants escape liability and some non-negligent defendants are found liable. It may be advantageous for injurers to depart from the legal standard under these circumstances. Whether they exceed it or fall short is indeterminate in principle, although it seems likely that injurers will want to exceed the legal standard to allow for a margin of error by the courts within which liability is avoided.²² Efficiency can be achieved by appropriate adjustment in the legal standard to offset the departure of injurers from it.

Errors by the court need not be symmetric, and damage rules other than compensatory damages are possible. P'ng has pursued this line of thought by focusing on the deterrent effects of Type I and Type II errors. He points out (in the context of a rule of strict liability) that Type I errors, in which penalties are mistakenly assessed against non-violators, lower the relative cost of violating the law, rather than conforming to it. To insure that the appropriate incentives are created, P'ng proposes a positive adjustment to compensatory damages so that violators must pay more, along with a subsidy for those who engage in the activity that runs the risk of being found in violation of the law by mistake.

3.4 Lawmaking By Courts

Most analyses relating to the courts have focused on the behavior of the parties prior to and during the litigation process. An important area of study that has received less attention is the role of courts in the lawmaking process. This section briefly summarizes the state of the lawmaking literature and speculates about some fruitful avenues of research.

Some economically-oriented scholars of the common law accept the positive and normative efficiency thesis, according to which judge-made law tends toward efficiency and reinforcing this tendency is good public policy (Posner). The positive thesis is testable, at least in principle, but there has been little quantitative research on how the common law changes.²³ The normative thesis, while not a statistically-demonstrable conclusion, is

²² Calfee and Craswell show that a reduced standard may be preferred, because increased precaution by the injurer reduces the probability that a suit will be successful and that the victim will recover damages.

²³ In a recent paper, Priest tried to test whether changes in doctrine by judges, which increase uncertainty, cause an increase in the scope of disagreement among litigants. Priest's data apparently shows that doctrinal change and increased disagreement occur in the same year, but not which occurs first. This fact is consistent with his hypothesis or with the rival hypothesis that changes in doctrine resolve uncertainties that cause litigants to disagree (Priest; Cooter)

rather a conviction that some people reach by reading many cases.

How might lawmaking by courts lead to efficient outcomes? Two different explanations correspond to two different conceptions of the common law. One conception, which regards litigation as a market, views the common law process as driven by competition among rationally self-interested actors. The other conception, which regards judging as an exercise in public reason, views the common law process as driven by the theories of law embraced by judges. We review the hypothesis that the law is market-driven here, but omit a discussion that it is idea-driven because of the limited economic work that has been done.

Several ingenious attempts have been made to explain how competition among litigants, like market competition among businesses, can produce efficiency without anyone consciously aiming for it.²⁴ One such mechanism is selective litigation. Suppose that inefficient laws are more likely to be litigated than efficient laws. If inefficient laws are repeatedly challenged in court, they may be overturned, whereas if efficient laws are less frequently challenged, they are more likely to persist unchallenged. Selective litigation could work like a strainer that catches inefficient laws while allowing efficient laws to slip past. The product, being repeatedly sieved, becomes more efficient with the passage of time. Two assumptions are enough to cause the law to evolve towards efficiency, at least weakly: (i) a rule's efficiency is negatively correlated to the probability that the litigants will test it in court, and (ii) efficiency is not negatively correlated to the probability of a rule surviving such a test before a judge.²⁵ For the process to operate, judges need not favor efficiency, but they must not disfavor it.

Does litigation tend to select inefficient laws? Theory suggests a weak "Yes". The more someone values a contested legal entitlement, the more that party will be prepared to spend on litigation to obtain it. Larger litigation expenditures increase the frequency of court challenges and improve their quality, which, in turn, increases the probability of winning. Thus the value that a person places upon a legal entitlement should correlate with the probability of winning it through litigation. By transferring legal entitlements from parties who value them less to parties who value them more, the common law tends toward efficiency.

This process can be redescribed as a contrast between distribution and efficiency. The allocation of legal entitlements affects both the quantity of wealth and its distribution. When legal entitlements are allocated inefficiently, the plaintiff who overturns the misallocation stands to gain from both the increase in wealth and from its redistribution. In contrast, when legal entitlements are already allocated efficiently, the plaintiff who overturns the allocation stands to gain from the redistribution of wealth and to lose from the decrease in its quantity. Because the value of overturning inefficient laws exceeds the value of overturning efficient laws, the frequency and quality of challenges to inefficient laws should be higher than that for efficient laws.

²⁴This possibility was first raised by Paul Rubin. See also Priest; and John Goodman. For a discussion of dynamic efficiency, see Landes and Posner; and Lawrence Blume and Rubinfeld.

²⁵A precise statement of the conditions for such evolution is found in Cooter and Lewis Kornhauser.

Selective litigation is similar to the "invisible hand" in markets but, unfortunately, the grip of the invisible hand in courts is far weaker than on markets. A law is, by its nature, general in the scope of its application, so challenging a law affects everyone who is, or will be, subject of it. Most plaintiffs appropriate no more than a fraction of the value that new precedent creates and redistributes.

The effects of a new, more efficient, precedent spill far beyond the litigants in the case where it is set. Litigants, however, may have little regard for the social costs that an inefficient rule imposes on others. The bias toward efficiency may be overwhelmed by the inclination of plaintiffs to challenge laws when they can capture a large share of the precedent's value. Plaintiffs may thus bring suit when they expect the redistributive gains of a successful challenge to be large, regardless of the law's efficiency or inefficiency. The problem with viewing litigation as a market is that redistributive gains are frequently more important than inefficiencies in channeling litigation.²⁶

An exception to this pessimistic conclusion concerns laws that are vague. Bargaining games are hard to settle when the parties do not know each others' threat points (Elizabeth Hoffman, Matthew Spitzer). An implication is that vague laws cause litigation. Laws whose inefficiency derives from their vagueness will tend to be litigated until the courts achieve a clear allocation of the underlying entitlements.

Our view is that, so far as common law tends toward efficiency, it must be driven by the ideas of judges, not by competitive pressures in the market for litigation. There is some evidence that the judiciary is giving a larger role to economic reasoning (Frank Easterbrook), but there is also evidence that the judiciary tends to expand its own powers, much as a bureaucracy engorges itself, without regard to benefits and costs. In addition, the fact that important legal cases are decided by majority vote of panels of judges raises the possibility that courts are afflicted by the same voting paradoxes as legislatures (Easterbrook; Kornhauser and Lawrence Sager; Spitzer).

The ideal of an independent judiciary implies creating circumstances under which judges decide cases that do not affect their private interests. The salary and tenure of federal judges are independent of their performance, and their performance is apparently unrelated to promotion to a higher court (Richard Higgins and Paul Rubin). These facts raise an issue about whether disinterestedness provides the best incentive structure, or whether competition among adjudicators might improve the efficiency of dispute resolution.

Economists have compared the incentives of judges and arbitrators (Landes and Posner; Robert Cooter; Christopher Bruce). Arbitrators maximize their own incomes by deciding disputes so as to maximize the demand for their services. If an arbitrator's decisions were not on the Pareto frontier, a rival arbitrator could lure away the former's customers by offering decisions that both parties prefer. There is then, a strong incentive for arbitrators to achieve Pareto efficiency with respect to the disputants. However, the parties to a

²⁶This problem is not solved by class action suits where the plaintiff represents a whole class of people whose legal rights will be extinguished by resolution of the dispute.

dispute who hire a private adjudicator do not internalize all the benefits of changing the law. Better rules will benefit future cases to which current disputants are not a party. Thus the incentives of arbitrators for creating new laws may be deficient.

Besides making common law, judges interpret statutes. Interpreting statutes involves supplying operational definitions for statutory language and applying these definitions to decide cases. Economists tend to conceive of legislation as the product of bargaining among the representatives of various interests. This view suggests that statutes should be interpreted according to the understandings and purposes of the underlying bargain, much like the interpretation of business contracts (Easterbrook). The purpose of legislative bargains, like business bargains, is to maximize the surplus from exchange. If this view is persuasive, then efficiency considerations should enter directly into the interpretation of statutes.

Unlike the collective choice literature, which is replete with impossibility theorems, the efficiency theory sounds an optimistic note: Courts are efficient. This thesis, when combined with the impossibility theorems, implies that courts are better than elected officials at shaping efficient laws. This proposition, if true, has important implications for judicial review: Instead of deferring to elected officials, courts should vigorously review legislation and regulations (William Riker and Barry Weingast).

4 Conclusion

The economic models of legal disputes and their resolution by courts described in this article represent a substantial improvement along some dimensions over traditional legal scholarship. Explaining the process of dispute resolution as an equilibrium in the interaction of self-interested decision makers draws upon a well-developed behavioral model that permits a comparison of the efficiency of alternative legal rules. Indeed, the greatest strength of this literature is its careful working out of the inexorable logic of self-interest. The models consequently provide a point of reference that legal theory needs for an understanding of courts and for deliberation over proposed changes in rules.

There are, however, significant obstacles and resistances that leave scope for development and, possibly, breakthroughs in the future. First, the literature suffers from unsatisfactory state of bargaining theory of legal disputes and courts. Indeed, the insights needed to improve strategic bargaining theory may be inspired partly by law's institutional detail. Second, as long as disinterested decision making remains a mystery to economics, the motives of judges cannot be endogenous in economic models. A better model of judicial decision making will force economists into the mainstream of jurisprudential debate about the motivations of judges. Third, the law and economics literature have yet to pursue adequately the modern economic theory of organizations, and to apply that theory to the operation of courts and other institutions whose governance rules are primarily legal. Finally, and most important, empirical research has lagged woefully behind theoretical ad-

vances. Improved data collection and additional econometric studies are needed to improve the empirical grounding of the economic analysis of law.

We would be remiss if we failed to mention the gain to economics from the interaction with law. The courts, like the stock market, respond quickly to shocks in ways that economists cannot seem to anticipate. In fact, the legal institutions that have evolved to deal with the externalities created by injurers are more varied and subtle than the traditional taxing institutions that are the focal point of many economists. For example, a decade of effort by economists to develop theories of tort law succeeded on its own scholarly terms, but economists all too often provided efficiency proofs for institutions that most lawyers now view as inefficient. The proposals for reform that the "tort crisis" has put on the agenda of legislatures and courts raise issues of institutional design that economists have just begun to consider. Economists can learn from lawyers how to make our policy science more deft, flexible, and responsive to a living institution.