Chapter 4

LITIGATION

KATHRYN E. SPIER*

School of Law, Harvard University, and National Bureau of Economic Research

Contents

* The final draft of this manuscript was submitted in 2004. Because of the inevitable publication delays, this chapter cannot cover all of the important recent developments in the field. The author thanks A. Mitchell Polinsky and Steven Shavell for invaluable help writing this chapter and the Searle Fund at Northwestern University for financial support. Jim Dana, Andrew Daughety, Claude Fluet, Xinyu Hua, Bentley MacLeod, and Jennifer Reinganum. Kristina Steffenson McElheran contributed outstanding editorial assistance.

Handbook of Law and Economics, Volume 1
Edited by A. Mitchell Polinsky and Steven Shavell

DOI: 10.1016/S1574-0730(07)01004-3

Abstract

The purpose of this chapter is to survey the academic literature on the economics of litigation and to synthesize its main themes. The chapter begins by introducing the basic economic framework for studying litigation and out-of-court settlement. One set of issues addressed is positive (or descriptive) in nature. Under what conditions will someone decide to file suit? What determines how much is spent on a lawsuit? When do cases settle out of court? Important normative issues are also addressed. Are the litigation decisions made by private parties in the interest of society as a whole? Next, the chapter surveys some of the more active areas in the litigation literature. Topics include rules of evidence, loser-pays rules, appeals, contingent fees for attorneys, alternative dispute resolution, class actions, and plea bargaining.

Keywords

litigation, trials, courts, settlement, bargaining

JEL classification: C72, D82, K00, K41

1. Introduction

The topic of this chapter—Litigation—is one of the liveliest research areas in the field of Law and Economics. Why do some lawsuits go to trial while many others are resolved out of court? Is out-of-court settlement in the interest of society? How confident should a judge or jury be before finding for one party or the other? Should the losing party be required to reimburse the winning party's legal expenses? Should policy makers restrict the contracts between lawyers and their clients or are these contracts best left to the free market? To what extent should the ruling in a lawsuit be constrained by prior rulings in similar cases? The economic issues surrounding these and other important questions are surprisingly subtle and the techniques used to examine them have grown increasingly refined over the years. The purpose of this chapter is to survey the academic literature on the economics of litigation and to synthesize its main themes.[1]

In the United States, 2.3 million non-criminal cases were filed in federal courts and 20.1 million were filed in state courts in the year 2000.[2] These numbers, while staggering, underestimate the true "reach" of litigation because they exclude the countless cases that are never filed, yet are resolved in the shadow of litigation through private dispute resolution mechanisms. In all, legal services account for 1 percent of the United States' labor force and 1.3 percent of its gross domestic product.[3] The sheer magnitude of these numbers speaks to the importance of the study of litigation from both a positive and a normative perspective. Developing a better understanding of the economic forces behind litigation decisions is critical for the participants in litigation—plaintiffs, defendants, lawyers, and judges—and policy makers, alike.

The premise of this chapter is that the main purpose of the court system is to facilitate value-creating activities and deter value-destroying activities through the enforcement of contracts and laws.[4] Mortgage contracts, for example, create economic value by allowing would-be homeowners to borrow against their future earnings. If the court system refused to uphold these contracts (preventing the bank from foreclosing in the event of non-payment, for example) then borrowers would default on their loans more often and, in anticipation, banks would choose to lend less money. Similarly, insurance

[1] Previous surveys include Cooter and Rubinfeld (1989), Hay and Spier (1998), and Daughety (2000).

[2] Ostrom, Kauder, and LaFountain (2001, p. 13). The state court figure also does not include traffic cases and juvenile cases.

[3] Statistical Abstract of the United States (2001, tables 596 and 641, pp. 384 and 418).

[4] The court system has other purposes, of course, some of which are also economic in nature. For instance, the courts can also provide a mechanism for *creating rules*. In particular, judges are often put in the position of "filling gaps" when private contracts are incomplete—in effect, creating a rule where one did not previously exist. In common law regimes, judges have the power to change or modify inappropriate or obsolete laws, decisions that may bind on future decisions through precedent. Moreover, litigation is a mechanism for *disseminating information*—it communicates to would-be offenders, for example, the consequences of their actions. Observing litigation may also teach us right from wrong and inform us of the risks associated with various activities.

policies (e.g., car insurance, homeowners insurance, life insurance, and health insurance) create economic value because risk-averse individuals would rather pay for their expected losses upfront through installments or insurance premiums. The value that the consumers derive from avoiding risk would be forgone if the insurance contracts were not upheld by the court system. Criminal penalties create value by deterring would-be criminals from stealing or destroying the valuable property of others. If these penalties were unenforceable by the court system, then deterrence would be compromised and more crimes would be committed.[5]

In an ideal world, the court system would be accurate, unbiased, and free. The enforcement of rules would take place immediately and no transactions costs would be incurred. But the world is far from ideal. Mistakes are made at trial: acquitted defendants may in fact be guilty of the charges and convicted defendants may be innocent. Judges and juries may also bring their personal or political biases into the courtroom: a sympathetic jury may award astronomical damages to a severely injured child or be overly harsh towards a corporate executive with deep pockets. Errors, biases, and expected litigation costs also distort the economic activities that take place in the shadow of litigation. A physician, for example, may prescribe unnecessary diagnostic tests for patients and may avoid certain practice areas (such as obstetrics) because of these added litigation risks.

The chapter is organized as follows. Section 2 introduces the basic economic framework for studying litigation and out-of-court settlement. Some of the issues are positive (or descriptive) in nature. For instance, under what conditions will someone decide to file suit? What determines how much is spent on a lawsuit? When do cases settle out of court? This section also addresses important normative issues. In particular, it explores whether the litigation decisions made by private parties are in the interest of society more broadly. Section 3 then uses the basic framework to survey some of the more active areas in the litigation literature. Topics include the rules of evidence, loser-pays rules, appeals, contingent fees for attorneys, alternative dispute resolution, class actions, and plea bargaining. Concluding remarks follow.

2. Basic framework

This section presents the basic economic framework used in the study of litigation. Although future sections of this chapter consider more sophisticated models, this section assumes that there is a single case involving two *litigants*: one *plaintiff* and one *defendant*. The plaintiff is the injured party who seeks compensation for her injuries. The defendant is the party who is potentially responsible for the plaintiff's injuries. We assume that litigation is costly for both the plaintiff and the defendant. Indeed, the average

[5] Jail time can also create value by incapacitating the criminal, thereby preventing future crimes.

hourly billing rate of a law firm partner in the United States in 2001 was $246.[6] Because of the private expenses associated with using the court system to resolve disputes, plaintiffs and defendants (and arguably the lawyers who represent them) have economic incentives to weigh their options carefully and make prudent decisions in litigation.

Section 2.1 abstracts from settlement decisions and considers the plaintiff's decision to pursue litigation. Section 2.2 considers the incentives of the litigants to resolve their case out of court through private settlement.

2.1. The decision to litigate

2.1.1. Bringing suit

A plaintiff's decision to litigate hinges on her private benefits from pursuing the case as well as her private costs of pursuit. [7] She will rationally choose to bring suit when the expected gross return from litigation, call it x, exceeds the cost of bringing the case to trial, c_p. The gross return, x, could reflect the expected judgment at the end of a long and costly trial or a settlement that takes place at some time prior to the trial. More generally, it could reflect issues that are somewhat beyond the scope of the current dispute, such as the impact that a court decision will have on future cases or the plaintiff's concern for her business reputation. The plaintiff's litigation cost, c_p, would typically include money paid to private attorneys, but it could also include the plaintiff's personal cost of effort, time, and any other opportunity costs associated with the plaintiff's involvement in the lawsuit. As with other economic decisions, the plaintiff will choose to pursue litigation when the strategy has positive expected value, $x > c_p$, and she will choose to not pursue litigation if the case has a negative expected return, $x < c_p$.[8]

2.1.2. Private litigation spending

In practice, the plaintiff—often with the help of her attorney—must decide how much time and effort to invest in the lawsuit. In other words, the plaintiff's private litigation costs are endogenous rather than exogenous. The plaintiff's investment choice will reflect both the underlying facts of the case and the beliefs that the plaintiff holds about the future of the case—including those concerning the investments and responses of the defendant.

[6] The 2001 Survey of Law Firm Economics (pp. 11–39).

[7] In this subsection I will not distinguish between cases that are settled and those that go all the way to final judgment. Settlement will be taken up in the next subsection, and appeals are considered in section 3.

[8] In this section, it is assumed that each litigant bears his or her own litigation costs, regardless of the outcome at trial. This known as the American Rule. Under the English Rule, on the other hand, the loser in litigation is required to reimburse the winner for certain expenses. The economic effects of these rules are discussed in section 3.

Formally, suppose that the plaintiff's expected recovery from litigation depends on her own spending as well as that of the defendant. The litigation investments made by the plaintiff and defendant, c_p and c_d, respectively, affect the plaintiff's future recovery at trial, $x(c_p, c_d)$. This function is increasing in the first argument and decreasing in the second.[9] The equilibrium investments can be viewed as the outcome of a non-cooperative game (which could be simultaneous or sequential) where the plaintiff maximizes her expected litigation return,

$$x(c_p, c_d) - c_p,$$

and the defendant minimizes his expected total payments,

$$x(c_p, c_d) + c_d.$$

As described earlier, the plaintiff will bring suit when her anticipated return from litigation, $x(c_p, c_d) - c_p$, is positive.

The structure of this game is similar to those of other types of contests, including patent races, tournaments in internal labor markets, and a variety of sports. The dynamic strategies employed by the participants in these contests hinge on the anticipated reactions of the rival—each player would like his opponent to "back off" and invest less in the contest (see Dixit, 1987; Katz, 1988).[10] To this end, the plaintiff might derive a strategic benefit from aggressive spending, for example, when the defendant's best-response function is decreasing in the plaintiff's investment (i.e., is downward-sloping). Conversely, the plaintiff would benefit from a commitment to lower spending levels if the defendant's best-response curve slopes upward. Either situation can arise in the general framework.[11] Along these lines, a credible commitment by the defendant to build a defensive war chest could lower the plaintiff's expected return from litigation, leading the plaintiff to drop the claim altogether.

2.1.3. Normative implications

The private decisions of the plaintiff and defendant to invest time and money in a lawsuit are not generally aligned with the interests of society as a whole. This section highlights why the plaintiff's private incentive to use the legal system diverges from the social

[9] This general structure captures the specific case where private litigation expenditures influence the plaintiff's probability of winning but the damages conditional on winning are fixed.

[10] See Bulow et al. (1985) for a general discussion of strategic commitment in competitive environments.

[11] Interestingly, the mathematical condition for the defendant's best-response curve to slope down—that the cross partial derivative of the recovery function is positive—implies that the plaintiff's best-response curve is upward sloping. Although the plaintiff may want to commit to aggressive spending, the defendant might actually refrain from tough spending strategies since aggressive spending would only elicit an aggressive response from the plaintiff. For some functional specifications, the best-response curves are backward-bending, i.e., increasing in some range and decreasing in others. This, of course, raises challenges for both theoretical inquiry and empirical estimation.

incentive.[12] In some circumstances, the plaintiff will not litigate often enough (or spend too little on the case), while in others she will litigate too often (or spend too much on the case).

To better illustrate both situations, first consider a simple moral hazard example where the potential defendant can take precautions (which require effort and/or monetary investments) to reduce the probability of an accident. If the defendant fails to take precautions, then the accident occurs with probability p_0 and the plaintiff sustains damages x. By investing e, the defendant reduces the probability of harm to $p_1 < p_0$. The defendant's decision to take precautions will hinge, of course, on what he expects the plaintiff to do following the accident. If the plaintiff's litigation costs are high relative to the magnitude of harm, $c_p > x$, then the plaintiff will choose not to litigate following an accident. Anticipating no future legal action, the defendant would have no incentive to take precautions *ex ante*. Conversely, if the plaintiff's litigation costs are relatively low, $c_p < x$, then the plaintiff will certainly litigate in the event of an accident. In this case, the defendant will rationally choose take precautions if his total expected cost associated with taking precautions, $e + p_1(x + c_d)$, is smaller than his total expected cost associated with being careless, $p_0(x + c_d)$.[13]

Insufficient litigation activity It is easy to construct examples where the plaintiff has an insufficient incentive to bring suit. Take, for example, the case where $p_1 = 0$ and $p_0 = 1$. Accidents are completely avoided if the defendant takes precautions and occur with certainty, otherwise. Suppose further that $e < x$, so that it is socially efficient for the defendant to take precautions. If the plaintiff could commit ahead of time to aggressively pursue litigation in the event of an accident, then, in anticipation, the defendant would rationally choose to take precautions to avoid the accident to begin with. There would be no accidents and no litigation costs spent—the first-best outcome! But the plaintiff's commitment to litigate following an accident may not be credible: when the plaintiff's cost of pursuing a case is high relative to the expected judgment, i.e., $c_p > x$, the plaintiff will choose not to pursue the claim. In anticipation, the defendant has no incentive to take precautions and deterrence is compromised.

This problem may be especially severe if the defendant has deep pockets and a reputation for tenacious defense spending. In the shadow of a powerful defendant, an injured plaintiff may rationally decide *ex post* to scale back or drop the suit altogether. This problem may be especially severe in situations where the defendant's failure to take precautions affects many victims rather than just one. Oil spills (e.g., the Exxon Valdez incident) and poorly-designed consumer products (e.g., the Ford Pinto) are two examples where the harms were broadly dispersed among victims while the responsibility for harm was concentrated. In situations like these, an individual victim's decision to sue

[12] See Shavell (1982b, 1997) for more thorough—and excellent—discussions of these important issues.

[13] The defendant's total expected cost includes the cost of precautions, the future expected liability, and the future expected litigation costs.

is a public good: the indirect effect on deterrence helps the population of victims more broadly. Due to the incentive of victims to free ride on the litigation efforts of a few (who must incur the litigation costs) there is likely to be insufficient litigation, from a societal perspective.

These ideas are also important in criminal law enforcement. When a victim of a crime presses charges, the benefits of a successful prosecution—which may include the incapacitation of dangerous individuals in addition to deterrence—will accrue to society more broadly. Furthermore, a criminal conviction does not entitle the victim to any monetary compensation (although the victim can sometimes successfully pursue a civil action.) The victim's costs of pressing charges may be significant, including the time and effort associated with the collection of evidence and appearing as a witness as well as the psychological costs associated with recounting bad experiences. There may also be indirect costs associated with embarrassment or loss of reputation or social standing as the victim of a crime (e.g., the stigma associated with rape that makes survivors reluctant to testify). In these situations, the socially desired level of litigation exceeds what would otherwise be observed in equilibrium.

Excessive litigation activity Using the simple model outlined at the beginning of this section, it is not difficult to construct examples where the plaintiff's private incentive to bring suit is *socially excessive*. Take the stark example where the potential defendant's precautions are totally ineffective: accidents occur with the same likelihood whether or not the defendant took care, $p_1 = p_0$. Note that the defendant would never take precautions in this extreme case. (Indeed, it would be socially wasteful if he did.) The plaintiff will sue the defendant following an accident when litigation has positive expected value, $c_p < x$, and the defendant will be forced to bear the litigation costs as well. This is socially wasteful. There is nothing the defendant could have done to avoid the accident, and consequently the litigation costs are a pure deadweight loss for society.

This type of divergence between private incentives and social objectives is very important in practice. Take, for example, the case of a divorcing couple battling in court over the financial assets acquired during their marriage and perhaps the custody of their children. One can imagine situations where every dollar spent by the wife is exactly cancelled out by a dollar spent by the husband. Both parties may end up spending a lot of money, time, and effort just to stand still! While divorce lawyers could profit handsomely in these types of cases, the litigants themselves and their children sometimes end up poor (not too mention bitter and emotionally drained).

This type of divergence can also arise in situations where precautions taken by the potential defendant reduce the expected social harm. Let's return to the more general case where $p_1 < p_0 < 1$. Suppose the plaintiff has a private incentive to litigate following an accident and that the defendant, in anticipation, takes costly precautions to reduce the likelihood of an accident (and, consequently, his future liability). The total social cost in this scenario is $e + p_1(x + c_p + c_d)$, the sum of the defendant's cost of precautions, the expected damages suffered by the plaintiff, and the expected litigation costs of both parties. If the plaintiff refrained from pursuing litigation (and the defendant rationally

refrained from taking precautions) then the total social cost would simply be $p_0 x$. This latter case may be preferable. In particular, if $(p_0 - p_1)x < p_1(c_p + c_d) + e$ then the level of private litigation activity is too high from the point of view of society as a whole. Intuitively, the society-wide benefit of litigation—the lower accident rate that results when the defendant takes more care—is outweighed by the total costs of litigation.

Discussion If the direction of the divergence between private and social incentives is clear, then a variety of public policies may be introduced to bring them into greater alignment. Damage multipliers, where a plaintiff can receive several times her actual damages, or punitive damages can give individual plaintiffs an additional incentive to litigate claims that would otherwise have negative expected value.[14] The ability of individual plaintiffs to consolidate their claims—through a joinder, perhaps—and the ability of lawyers to represent the interests of dispersed plaintiffs through class representation mitigates the externalities that otherwise would constitute an obstacle to recovery and deterrence. Conversely, other policies can be used to reduce the level of litigation. Strict liability, where a defendant is liable for a plaintiff's damages regardless of his level of care, can lead to high levels of litigation activity. The negligence rule, where the defendant is liable only if he took inadequate care, can give the defendant the appropriate incentives for care while reducing the total volume of litigation and its corresponding social costs. These brief examples are given for illustrative purposes only—the subtleties of these research areas and many others will be discussed in greater detail in section 3.

2.2. *Out-of-court settlement*

The preceding discussion focused on the decision of a plaintiff to bring a lawsuit while abstracting from the private settlement of litigation. We will now focus on the private settlement of legal claims. In practice, the vast majority of cases that are filed ultimately settle before trial and countless others are settled before a case is filed at all. Less than 4 percent of civil cases that are filed in the U.S. State Courts go to trial.[15] In the U.S. Federal Courts, only about 2 percent of civil cases go to trial.[16] The fact that most private litigants choose to "opt out" of formal litigation channels should not be too surprising. The pursuit of litigation is expensive, time-consuming, and distracting. In short, trials are a decidedly inefficient way for private parties to resolve their disputes.

As above, suppose that x is the expected judgment at trial and let c_p and c_d be the plaintiff and defendant's trial costs, respectively. If the case went to trial, the plaintiff would receive a net payoff of $x - c_p$, the expected judgment minus his litigation costs,

[14] Polinsky and Rubinfeld (1988b) argue that compensatory damages, where the plaintiff's award is exactly equal to her damages, is not generally optimal. The optimal damages would either be adjusted upward or downward, depending upon the circumstances.

[15] Ostrom, Kauder, and LaFountain (2001, p. 29).

[16] Judicial Business of the United States Courts (2001, p. 154, table C-4).

and the defendant's payoff would be $-x - c_d$. Although x represents a simple transfer from the defendant to the plaintiff, the total cost of litigation, $c_p + c_d$, is a deadweight loss—"money down the drain," so to speak. The plaintiff and the defendant can typically avoid this loss through a private agreement to end the dispute before the litigation costs are incurred. Specifically, a binding settlement contract that specifies a transfer from the defendant to the plaintiff, $S \in (x - c_p, x + c_d)$, leaves both the plaintiff and the defendant better off than they would be from going to trial.

The insight that out-of-court settlement Pareto dominates trials (at least for the litigants themselves) raises a number of interesting questions. For what amount will the case settle? Will the defendant agree to settle if the plaintiff has a negative expected value claim, $x - c_p < 0$? When will the case settle—shortly after filing or on the courthouse steps? Why do some cases fail to settle? Is private settlement in the interest of society, more broadly? We will see that the answers depend on many factors, including the timing of offers and counteroffers, the information and beliefs of the two litigants, and the way that the particular lawsuit fits into the broader economic, legal, and strategic environment.

2.2.1. Settlement with symmetric information

The simplest economic framework considers settlement under symmetric information. Suppose that the two litigants have exactly the same beliefs about what will happen if the case goes to trial—i.e., they are symmetrically informed about the stakes of the case, the litigation costs, and all other relevant parameters. We will begin by assuming that $x - c_p > 0$, so the plaintiff clearly has a credible threat to take the case all the way to trial. Later, we will relax this assumption and assume that $x - c_p < 0$. Even though, in this latter instance, the case has negative expected value if pursued all the way to trial, the plaintiff may nevertheless succeed in extracting a settlement in equilibrium.

"Lumpy" litigation costs Consider a simple extensive form game where there are $T - 1$ rounds of bargaining before a costly trial in round T. The litigants alternate in making offers. Suppose that the plaintiff is designated to make the last settlement offer before the trial, call it S_{T-1}. If the case fails to settle then x is transferred from the defendant to the plaintiff and the litigation costs, c_p and c_d, are incurred. The costs of litigation are lumpy in the sense that they are all incurred at trial instead of gradually over time. Finally, suppose that the two litigants discount time at the same rate and let δ be their common discount factor.

This game is solved by backwards induction. In period $T - 1$ the defendant will accept any offer that is better than going to trial, where he would pay $x + c_d$ in total. The plaintiff's best offer in period $T - 1$ is $S_{T-1} = \delta(x + c_d)$, minus a penny perhaps.[17]

[17] For notational simplicity, I will assume that the defendant will accept an offer when he is indifferent between accepting and rejecting. The hypothetical penny would make him strictly prefer acceptance, but it would muddy the algebra.

Working backwards, we consider period $T-2$, where the defendant has the right to make a settlement offer. In this period, both litigants realize that if the case fails to settle in period $T-2$ then it will certainly settle for $S_{T-1} = \delta(x + c_d)$ in the next round. The very least that the plaintiff is willing to accept is therefore δS_{T-1}, and so the defendant would offer no more than $S_{T-2} = \delta^2(x + c_d)$ and the plaintiff would accept. This logic suggests that the case will settle in the first round for $S_1 = \delta^{T-1}(x + c_d)$.

The allocation of the bargaining surplus is sensitive to the timing of the settlement offers. Suppose instead that the defendant is the one to make the last offer. In this case, the defendant would offer $S_{T-1} = \delta(x - c_p)$ in the last round and the plaintiff would accept. Working backwards to the first offer we would find $S_1 = \delta^{T-1}(x - c_p)$. Comparing these two cases—one where the defendant makes the final offer and one where the plaintiff makes the final offer—delivers an important insight: the party who makes the last offer succeeds in extracting all of the bargaining surplus! In this alternating offer game with all litigation costs being incurred in a "lump" at trial, being last is much more important than being first.

The bargaining surplus would, of course, be more evenly shared under different assumptions about the timing of offers and counteroffers. Suppose instead that the two litigants flip a coin in each bargaining round to determine who will make the offer. In the last round (if it were reached) the parties would settle, on average, for $S_{T-1} = \delta[x + (c_d/2) - (c_p/2)]$. Working backwards, one can see that the case could settle in the first round for $S_1 = \delta^{T-1}[x + (c_d/2) - (c_p/2)]$, regardless of who makes the offer. Note that if the plaintiff's and defendant's litigation costs are of roughly the same magnitude, then the settlement amount would accurately reflect the discounted expected judgment at trial.

It is interesting to note in this example that, although it is efficient for the case to settle before trial, *it is no more efficient for the case to settle in round 1 than in round* $T-1$. In the random-offer framework of the previous paragraph, if the case settles in round 1 the plaintiff's payoff is $S_1 = \delta^{T-1}[x + (c_d/2) - (c_p/2)]$. If settlement is delayed so that the case settles on the courthouse steps instead, then the plaintiff's payoff (discounted back to round 1) is the same: $\delta^{T-1}S_{T-1} = \delta^{T-1}[x + (c_d/2) - (c_p/2)]$. To put it somewhat differently, the plaintiff is indifferent between settling early and settling late and a similar logic shows that the defendant is indifferent as well. The reason for this is simple: there is no inefficiency associated with delay when the costs of litigation are "lumpy" and are all borne at trial.

Finally, the simple framework presented here differs in several important respects from the related (and certainly more famous) framework of bilateral trade. Suppose that a buyer's valuation for a good or service, B, exceeds the seller's cost, C. Following trade, the buyer enjoys surplus $B - P$ and the seller enjoys profit $P - C$. It is clearly in the buyer's and seller's mutual interest for trade to take place sooner rather than later because discounting causes the total value that must be shared between them to diminish. If they trade immediately, their joint surplus is $B - C$. If they were to wait and trade in period $T-1$ (say) then their joint surplus would be $\delta^{T-1}(B-C)$. In bilateral trade, discounting causes the total "pie" to shrink. This is not the case in the settlement

framework. When all of the costs of litigation are incurred at trial, the bargaining surplus in each bargaining round is zero: the plaintiff receives exactly what the defendant pays.

Divisible litigation costs The previous section assumed that all of the costs of litigation were incurred at trial. In reality, the costs of litigation are incurred by the litigants in a variety of ways while preparing for the trial. In addition to the direct costs of trial preparation, there may be costs of distraction (as litigants focus on lawsuits rather than their jobs and families) and risk management issues associated with open claims.[18] One can easily extend the earlier framework by assuming that the costs of litigation are divided among the T rounds—pretrial bargaining rounds as well as the trial round. In contrast to the scenario with lumpy litigation costs, there is a unique subgame perfect equilibrium where the case settles in round 1 (Bebchuk, 1996).[19] When litigation costs are incurred over time, then delay is inefficient and there are strong economic incentives to settle early.

Negative expected value (NEV) claims The previous sections assumed that the plaintiff had a credible threat to take the case to trial if it did not settle. We will now suppose that the plaintiff has a "negative expected value claim," one that would be unprofitable for the plaintiff if pursued all the way to trial. In particular, we assume $x - c_p < 0$. The defendant will certainly refuse to pay a cent if the plaintiff cannot credibly commit to litigation. Can the plaintiff succeed in extracting a settlement offer from the defendant under these circumstances?

The key to credibility for the plaintiff hinges on some of the factors identified above: (i) the divisibility of the litigation costs and (ii) the timing of offers. Suppose that the costs are "lumpy" as above—all of the litigation costs are borne at trial—and that the plaintiff has the option of dropping the case at any point in time. Subgame perfection implies that if settlement negotiations fail, then the plaintiff will surely drop the case on the courthouse steps. Backwards induction implies that the defendant would never make or accept an offer to settle.

It would be a mistake to conclude that negative expected value claims cannot succeed, however. Bebchuk (1996) shows that when the litigation costs are divisible and spread over the bargaining phase, then the set of circumstances under which a plaintiff can succeed in extracting a settlement is broader. To see why, suppose for simplicity that the litigation costs, c_p and c_d, are equally divided among the T periods and that there is no discounting of time ($\delta = 1$). Although the case may begin with negative expected value, $x - c_p < 0$, it can be "transformed" into a positive expected value case when T, the number of periods, is sufficiently large. To see why, imagine that the case actually reaches the courthouse steps. Most of the litigation costs are sunk at that point in time.

[18] A corporate defendant, for example, may need to keep additional resources on hand in anticipation of the possibility of a large adverse judgment or other surprises.

[19] With discounting, this expression would be modified to reflect the fact that the costs are not all borne at the same time.

When standing on the courthouse steps, the plaintiff has a credible threat to proceed to trial when $x - c_p/T > 0$. (The plaintiff's ultimate threat to litigate is stronger when T, which captures the divisibility of the costs, is larger.) If they were to "flip a coin" at that point, the case would settle on average for $S_{T-1} = x + (c_d/2T) - (c_p/2T)$. Working backwards[20] (and assuming that the plaintiff's litigation costs are not too large relative to the defendant's), we would find that the case could settle for approximately $S_1 = x + (c_d/2) - (c_p/2)$ in round 1. The plaintiff's threat to litigate is thus credible at each stage of the game!

2.2.2. Settlement with asymmetric information

In the simplest settlement games with symmetric information—including the examples discussed above—cases either settle out of court for a positive amount or are dropped without future costs being incurred. In short, settlement with symmetric information is privately efficient. In this section we will see that bargaining may be privately inefficient if the litigants are asymmetrically informed. These inefficiencies result in costly trials as well as costly delays in negotiations.

Asymmetric information has a variety of sources in litigation and manifests itself in a variety of ways. The plaintiff, for example, may have first-hand knowledge of the level of damages she has suffered; the defendant may have first-hand knowledge about his degree of involvement in (or liability for) the accident. Both litigants may know better the credibility of their own witnesses and the quality and work ethics of their lawyers.[21] It is important to note that some of this information will become commonly known over time—the parties may learn a great deal through pretrial settlement proceedings, for example. Other information may not come to light at all, but can nevertheless affect the trial outcome. The salient point here is that the revelation of information—through pretrial discovery activities and through formal legal proceedings—is both privately and socially costly. All else equal, litigants have private incentives to settle their dispute before the costs are sunk.

For the rest of this section we will assume that the defendant has private information about x, the expected judgment at trial. A similar analysis would follow if it were the plaintiff who had the private information, instead.[22] Formally, he observes the parameter x where x is drawn from a probability density function $f(x)$ on $[\underline{x}, \bar{x}]$ with cumulative

[20] The second to last offer would be $S_{T-2} = x + (c_d/T) - (c_p/T)$ on average, and so on.

[21] The sources of asymmetric information mentioned in this paragraph all affect the expected judgment at trial. The parties may also be privately informed about their own level of risk aversion or their discount rates, as in Farmer and Pecorino (1994). Private information along these lines will certainly affect the settlement offers that the information holder is willing to make or receive, but is not directly payoff relevant to the rival. Private information could also be endogenously determined by the litigants' effort choices, as in Hay (1995).

[22] Schweizer (1989) and Daughety and Reinganum (1994) consider extensive form games with two-sided asymmetric information. Spier (1994b), discussed in more detail later, explores two-sided asymmetric information settings with mechanism-design techniques.

density $F(x)$. For reasons that will become apparent later, we will also assume that this distribution has a *monotone hazard rate*, so $[1 - F(x)]/f(x)$ is everywhere decreasing in x.

Screening models Starting with P'ng (1983) and Bebchuk (1984), many papers have considered a somewhat special framework where the uninformed player—the plaintiff in our example—makes a single take-it-or-leave-it settlement offer, S, before a costly trial. The settlement offer "screens" the defendants into two groups: those who accept and those who reject. When faced with a settlement offer, S, the defendant will certainly choose to accept the offer if it is lower than what he would expect to pay at trial, $S < \delta(x + c_d)$. The settlement offer corresponds with a cutoff, $\hat{x} = \delta^{-1}S - c_d$, where defendants with types above the cutoff accept the plaintiff's offer to settle and those below the cutoff go to trial instead. Note that the sample of cases that go to trial is not a random sample—a defendant who is more confident about his prospects at trial (a defendant with low x) is more likely to reject the settlement offer and go to trial. Thus, the cases that go to trial would, on average, have lower judgments than those that settle out of court.

This case-selection result would, of course, be reversed if the plaintiff had private information instead and the defendant could make a take-it-or-leave-it offer before trial. The defendant's settlement offer would correspond to a cutoff where plaintiffs with low x's accept the offer and those with high x's reject the offer and go to trail. In this instance, cases that go to trial would, on average, have higher judgments than those that settle. This observation should make it clear to the reader that the empirically testable implications of these models depend very strongly on the source of the information asymmetry.

For now, we return to the case of private information on the part of the defendant. If given the power to make a take-it-or-leave-it offer, the plaintiff would choose her settlement offer to maximize her expected profit. Since each settlement offer corresponds to a unique cutoff, we may write the plaintiff's optimization problem as a function of the cutoff, $\hat{x}$:

$$\underset{\hat{x}}{Max} \int_{\underline{x}}^{\hat{x}} \delta(x - c_p) f(x) dx + [1 - F(\hat{x})]\delta(\hat{x} + c_d).$$

The first term represents the payments associated with the confident defendant types who reject the settlement offer and go to trial. Note that we are assuming that the plaintiff has to bear her own litigation costs in these cases. Fee-shifting rules, including the English rule, are discussed in section 3. The second term reflects payments, $S = \delta(\hat{x} + c_d)$, from the defendant types above the cutoff, $\hat{x}$. An interior solution, if it exists, is characterized by the following first-order condition:[23]

$$1 - F(\hat{x}) - (c_p + c_d) f(\hat{x}) = 0.$$

[23] The monotone hazard rate assumption guarantees that this solution is unique.

This condition may be understood intuitively: when the plaintiff raises her offer slightly from $S = \delta(\hat{x} + c_d)$ to $S = \delta(\hat{x} + c_d + \Delta)$, there are both benefits and costs. The benefit is that those defendants with types above $\hat{x} + \Delta$ will pay $\hat{x} + \Delta$ more than before, a benefit which is approximately $\Delta\delta[1 - F(\hat{x})]$, the discounted additional payment, $\Delta\delta$, multiplied by the probability of acceptance $1 - F(\hat{x})$. The cost is that defendants with types between $\hat{x}$ and $\hat{x} + \Delta$ go to trial instead of settling. This cost to the plaintiff is approximately $\Delta\delta(c_p + c_d)f(\hat{x})$. This includes both the plaintiff's cost of litigation (borne directly by the plaintiff) and the defendant's cost of litigation (borne indirectly by the plaintiff through the foregone settlement offer).

The first order condition implies that at least some cases will settle—the plaintiff will certainly make a settlement offer that is accepted by the most liable defendant. This can be seen by setting $\hat{x} = \bar{x}$. The left hand side of the first-order condition is negative when evaluated at this value, implying $\hat{x} < \bar{x}$. Furthermore, if $(c_p + c_d) > [1 - F(\underline{x})]/f(\underline{x})$ then the plaintiff will make a "low ball" offer, $S = \delta(\underline{x} + c_d)$, that all defendant types will accept. Intuitively, all cases settle if the litigation costs outweigh the degree of asymmetric information. An interior solution exists when the costs of litigation are not too high, i.e., when $(c_p + c_d) < [1 - F(\underline{x})]/f(\underline{x})$.

This screening example assumed that the plaintiff had a credible commitment to pursue the case all the way to trial. This is not necessarily true. The defendant types who reject the offer are the ones who believe that they have strong cases (i.e., they have low x's). The plaintiff, understanding this, may therefore have an incentive to drop the case following the rejection of her settlement offer! This credibility issue may be seen most readily by considering the case where $\underline{x} = 0$ and the litigation costs are large. The preceding analysis tells us that the plaintiff should offer to settle for $S = \delta(\underline{x} + c_d) = \delta c_d$. But if the defendant rejects the offer, the plaintiff may rationally believe that the defendant is a low type and will choose to drop the claim. If so, the defendant has an incentive to reject the settlement offer. Nalebuff (1987) extends this framework to incorporate a credibility constraint and shows that when the constraint is binding, the equilibrium settlement offer is higher than before. (A higher offer implies that the average return at trial from the bottom of the truncated distribution is higher as well, making the plaintiff's commitment to proceed credible.)

The screening example above also assumed that there was a single offer before trial. Spier (1992a) extends Bebchuk's framework to consider a sequence of settlement offers before trial. When litigation costs are "lumpy," a striking pattern emerges: the plaintiff waits until the very last moment to offer $S_{T-1} = \delta(\hat{x} + c_d)$, where $\hat{x}$ is defined above. That is, all settlement occurs on the courthouse steps! This result is important for several reasons. First, we often observe these 11^{th} hour settlements in practice, demonstrating the practical relevance of the asymmetric information framework. Second, this result implies that the common way of modeling these bargaining games—that of take-it-or-leave-it offers—is less restrictive than it may first appear. The finitely-repeated screening model where all of the costs are borne at trial is equivalent to the simple model with a single offer.

The reason why the plaintiff refrains from early offers is not hard to see. Suppose that, for the sake of argument, the plaintiff could commit to a take-it-or-leave-it offer in round 1, taking the defendant all the way to trial if he rejects the round 1 offer. The best offer that she could possibly make is $S_1 = \delta^{T-1}(\hat{x} + c_d)$, implementing the very same cutoff as above. Note that the plaintiff is indifferent between receiving $S_1 = \delta^{T-1}(\hat{x}+c_d)$ in round 1 and the present discounted value of receiving $S_{T-1} = \delta(\hat{x}+c_d)$ on the courthouse steps. As discussed earlier, when the plaintiff and defendant discount time at the very same rate, then the passage of time alone does not impose costs on the two litigants. It is not credible, however, for the plaintiff to make a take-it-or-leave-it offer in round 1. Following a rejection, the plaintiff would update her beliefs about the remaining distribution and make a lower settlement offer. In anticipation, the defendant would wait to receive a better offer. The plaintiff can thus use delay to her advantage. By postponing settlement talks until the last moment, the plaintiff optimally extracts rents from the defendant.

Spier (1992a) also shows that when the costs of litigation are divisible over time, the plaintiff's optimal strategy does involve some settlement in each round. In particular, there is more settlement in the first rounds than in the middle and, if the costs borne at trial are disproportionately large, one ought to observe a pronounced deadline effect where many cases settle on the courthouse steps (generating a "U-shaped" pattern of settlement, overall). Importantly, this result is still obtained in the limit as the number of rounds, T, approaches infinity. The reason that this result differs from the familiar Coase Conjecture is that the passage of time before settlement does not screen among defendant types in the ordinary sense. Since they discount time at the same rate, the different defendant types have the same preferences between a settlement offer in round 1 and a settlement offer in round T. The types differ, however, in their preference for going to trial, generating delay in equilibrium. See also the empirical work of Fournier and Zuehlke (1996).

Signaling models We will now suppose that the informed defendant, rather than the uninformed plaintiff, makes a take-it-or-leave-it offer on the courthouse steps.[24] The informed defendant's settlement offer potentially signals his private information and the uninformed plaintiff must form Bayesian inferences when deciding how to respond to the offer.

Reinganum and Wilde (1986) characterize an elegant fully-separating equilibrium of this game where the defendant's offer perfectly reveals his type and the plaintiff mixes between accepting and rejecting the defendant's offer. The defendant's equilibrium settlement offer,

$$S(x) = \delta(x - c_p),$$

[24] Daughety and Reinganum (1993) present a model where both the timing of offers and the information structure are endogenous.

gives the plaintiff exactly the same payoff that she would get at trial.[25] The plaintiff subsequently randomizes between accepting and rejecting the offer where the probability that the plaintiff accepts $S(x)$ is given by

$$\pi(x) = e^{-(\bar{x}-x)/(c_p+c_d)}.$$[26]

Note that the probability of acceptance is increasing in the defendant's expected liability, x. This property is implied by the defendant's incentive compatibility constraint. Intuitively, the defendant must be rewarded in equilibrium for making higher settlement offers with a higher rate of acceptance by the plaintiff (who is indifferent between accepting and rejecting).

Mechanism design The screening and signaling models we have just seen are similar in some ways and very different in others. Importantly, the two models have similar implications for the selection of cases for trial. When the defendant has the private information then, on average, the cases that settle out of court have higher expected liability than the cases that go to trial. While the screening model generates this feature very starkly through the cutoff, $\hat{x}$, the signaling model generates it through the plaintiff's mixed strategy. (If the plaintiff had the private information instead, then this result would, of course, be reversed: cases that settle would have a lower expected liability than those that go to trial.) One difference is that, in the signaling model, every type of case (with the exception of type $\bar{x}$) has a positive probability of proceeding to trial in equilibrium.

An alternative approach to studying pretrial bargaining games is to consider the entire class of bargaining games using the Revelation Principle (Myerson, 1979). This principle tells us that any equilibrium of any extensive-form game may be represented as an outcome of a direct-revelation mechanism in which the defendant announces his type, $\tilde{x}$, and the mechanism designer subsequently maps this announcement into an outcome of the game. In the simplest settlement context, the mechanism would be a settlement offer, $S(\tilde{x})$, and a corresponding probability of settlement, $\pi(\tilde{x})$. Incentive compatibility for the defendant requires that he prefer to announce truthfully, and participation constraints imply that the players weakly prefer to play the direct revelation game than

[25] We will maintain the assumption that the plaintiff has a credible threat to take the case to trial regardless of his beliefs, i.e., $\underline{x} - c_p \geq 0$.

[26] Here is the derivation. If the defendant of type x were to mimic another type $\tilde{x}$ his expected payments would be $\pi(\tilde{x})\delta(\bar{x}-c_p)+[1-\pi(\tilde{x})]\delta(x+c_d)$. Incentive compatibility for a defendant of type x requires that he prefers to "tell the truth" and offer $S(x)$ rather than pretend to be someone who he is not by offering $S(\tilde{x})$. In other words, the derivative of this expression must be zero when $x = \tilde{x}$, or $\pi(x) - \pi'(x)(c_p + c_d) = 0$. The general solution of this differential equation is $\pi(x) = \alpha e^{x/(c_p+c_d)}$ where α is a positive constant. The boundary condition is $\pi(\bar{x}) = 1$. Suppose this boundary condition did not hold. The defendant of type $\bar{x}$ could raise his offer slightly above $\delta(\bar{x} - c_p)$ and the plaintiff would accept with certainty. (This offer dominates going to trial for any beliefs that the plaintiff might have about the defendant's type.) The D1 refinement of Cho and Kreps (1987) may be used to get rid of pooling equilibria.

go to trial. Both the screening and the signaling model discussed can be recast in this conceptual framework.

The mechanism-design approach provides us with proof that some cases will *necessarily* go to trial when the litigation costs are not too large. Formally, there does not exist a direct-revelation mechanism where all cases settle out of court. This may be easily established with a "proof by contradiction." Suppose that $\pi(\tilde{x}) = 1$ for all announced types—every case settles out of court. This implies that all defendant types must settle for exactly the same amount, call it S^*. If not, then the defendants, regardless of their information, would imitate the type with the lowest assigned settlement amount. The defendant's participation constraint requires that $S^* \leq x + c_d$ for all values of x, including the very lowest. (The defendant weakly prefers playing the mechanism to going to trial.) The participation constraint for the plaintiff requires that the plaintiff do better in settlement than she would do on average at trial, $S^* \geq E(x) - c_p$. Taken together, the value S^* can exist only if $E(x) - \underline{x} \leq c_p + c_d$, an assumption clearly violated when $c_p + c_d$ are small.

This result may be contrasted with that of Myerson and Satterthwaite's (1983) famous results for bilateral trade. They show in the bilateral-trade framework that if the buyer's valuation were private but the seller's type were known then a mechanism that set the price equal to the seller's cost would realize the gains from trade. In the litigation context, however, the plaintiff and defendant are unable to resolve their dispute even with one-sided incomplete information. It is also interesting to note that the breakdown occurs despite common knowledge that gains from trade exist: the litigants both know that they will jointly save $c_p + c_d$ by settling. In Myerson and Satterthwaite (1983), breakdowns only arise when the supports for the buyer's valuation and the seller's cost do not overlap, corresponding to the situation in which trade is not always efficient.

Although the literature has not delivered many positive or normative results about the entire class of pretrial bargaining games, several scholars have studied the mechanisms that achieve the Pareto frontier. For this so-called "optimal" mechanism, it can be shown that the selection effects described above hold here as well (so more liable defendants are more likely to settle). These mechanisms have also been investigated in more applied research. In particular, Spier (1994b) and Neeman and Klement (2005) use these mechanisms to consider the shifting of legal fees between winners and losers and pleadings rules, respectively. Although these mechanisms can provide a useful upper bound on private and social welfare, the approach has certain drawbacks. Unlike the bilateral trade mechanism of Myerson and Satterthwaite (1983) discussed above, the pretrial bargaining mechanisms are not necessarily implemented by standard extensive form bargaining games. In practice, these mechanisms would require players to commit to abide by the rules of the mechanism and for courts to enforce them.

2.2.3. Alternative frameworks

Mutual optimism Before the popularity and widespread adoption of techniques from information economics, law and economics scholars took a non-Bayesian approach to

settlement breakdowns. Starting with the work of Landes (1971), Posner (1973), and Gould (1973), many scholars have taken the position that litigants may have different (and possibly inconsistent) priors about the outcome at trial. The plaintiff, for example, may believe the expected judgment at trial to be x_p while the defendant may believe it to be x_d. These divergent beliefs may arise when the two litigants receive different signals of the "true" expected damages, x, and may be influenced by their different backgrounds and experiences. Examining the bargaining zone, $[x_p - c_p, x_d + c_d]$, shows that the case will fail to settle when the plaintiff is much more optimistic than the defendant: $x_p - x_d > c_p + c_d$.

The optimism framework has two important advantages for applied work in litigation: its tractability and (arguably) its realism in many litigation settings. Scholars have used this framework to explore diverse topics such as the selection of cases for trial (Priest and Klein, 1984), fee-shifting (Shavell, 1982a), conflicts between lawyers and clients (Miller, 1987), and bifurcation of trials (Landes, 1993). It has also served as a foundation for empirical work on settlement (see Waldfogel, 1998). There is also interesting experimental and anecdotal evidence that litigants and their lawyers do tend to exhibit self-serving biases (Loewenstein et al., 1993). As a group, plaintiffs may have a tendency to overestimate the expected judgment at trial, $x_p > x$, while defendants as a group may tend to underestimate them, $x_p < x$. Indeed, these self-serving biases may serve as an advantage in bargaining—they allow the optimistic litigants to grab a greater share of the bargaining surplus—and can arise in evolutionary settings (Bar-Gill, 2006).

However, the optimism framework described above has disadvantages as well. In many ways, the optimistic litigants are "too stubborn": they stick with their inconsistent prior beliefs "come Hell or high water." In reality, many litigants—especially those with skilled lawyers—update their beliefs over time as new information emerges. They learn about the underlying merits of the case and are aware of strategies that their opponents employ. A careful understanding of this learning process is critical for both positive and normative analyses. It gives us a better understanding of the private litigation strategies we observe in reality and it is helpful in evaluating the effects of litigation reform. See Aumann (1976) for early work on common knowledge, Yildiz (2003, 2004) for recent theoretical work on learning and delay without common priors, and Watanabe (2005) for an empirical analysis.

Settlement externalities Cases may also fail to settle where there are externalities among existing claims.[27] To take a simple example, suppose that there are two plaintiffs and one defendant and that the payoff for each plaintiff at trial depends on how many cases ultimately go to trial. In this setting, settlement by one plaintiff changes the bargaining position and ultimate recovery of the second. Formally, suppose that each

[27] These types of externalities will be discussed in several topics in section 3, including joint and several liability and settlement with limited liability constraints.

plaintiff's expected payoff is x_2 if both plaintiffs go to trial and x_1 if only one plaintiff goes to trial. Depending upon the setting, x_2 may be either larger than or smaller than x_1.

First, suppose that $x_2 < x_1$. This may arise, for example, when the defendant enjoys economies of scale in case preparation. We can imagine settings where the defendant has a credible commitment to spend more on his defense when facing two plaintiffs than he has when fighting only one. Note that settlement by one plaintiff confers a positive externality on the second plaintiff, raising the payoff from x_2 to x_1. It is implausible that both plaintiffs will agree to settle out of court for x_2: if one plaintiff expects the other to settle out of court, he would rather reject the offer of x_2 and go to trial where he receives $x_1 > x_2$. Consequently, one or both plaintiffs will be able to command a settlement premium above and beyond x_2. It is not hard to see that, if the defendant's costs of litigation are not too large, then the defendant would rather forego settlement altogether than pay the plaintiffs a premium to compensate them for the positive externalities of settlement.[28]

Now suppose instead that $x_2 > x_1$. This could arise if the plaintiffs have a large joint fixed cost of pursuing their cases such as legal representation. Here, the plaintiffs are better off going to trial together rather than separately. Indeed, if one plaintiff were to settle then the second might decide to drop the case rather than pay a lawyer to pursue it independently. In contrast to the case where $x_1 > x_2$, here settlement by one plaintiff imposes a negative externality on the other. We would expect both cases to settle in this scenario. If the defendant offered each plaintiff $S = x_2$ (plus a penny) on the courthouse steps then both plaintiffs would be thrilled to accept. Indeed, the defendant may be able to induce them to settle for much less than that.[29]

Other reasons for bargaining failures There are many other reasons why settlement negotiations may fail. First, imagine that the object of litigation is indivisible—the custody of a child, for example—and that the litigants are liquidity-constrained. If the litigants value the object above and beyond the scope of monetary payment, then there is no scope for settlement. (See Shavell, 1993, and Mnookin and Kornhauser, 1979.) Second, one of the litigants (or their lawyer) may derive independent value from having his or her day in court (perhaps they enjoy the publicity!) or derive non-pecuniary pleasure from imposing costs on the opponent. Third, suppose that the parties have very

[28] The careful reader will notice that the defendant would try to "tie" the offers together: "My offer to settle for $S = x_2$ is good only if both of you accept. If only one accepts, the deal is off and I will take both of you to court." These types of offers are in fact observed in some class action settlements where a requirement is that a certain percentage of plaintiffs remain in the class.

[29] One way is to make settlement offers in sequence. There may be better strategies, however. If he offered between $S \in [x_1, x_2]$ then there are two equilibria: one where both plaintiffs accept and another where both reject. Although the latter is Pareto superior, the former is the risk-dominant outcome when S is not too small. See Spier (2002).

asymmetric stakes in the case. A corporate defendant, for example, may derive particular economic value from establishing a judgment in an early case that could chill the filing of future cases.

2.2.4. Normative implications

At first blush, there are strong normative arguments in favor of settlement. Take a single lawsuit—a personal injury case, perhaps—that would otherwise go to trial. As we have seen, it is certainly in the litigants' interest to resolve their dispute out of court. Through a private settlement, the parties can avoid their private litigation costs $c_p + c_d$ and (if they are risk averse) the risk premium associated with trials. It is also important to note that social costs are avoided as well—there are, of course, large fixed costs of maintaining the court system and legal infrastructure and significant marginal costs associated with any given trial (the judge's and jury's time, for example). *All else equal*, private settlement serves society's interest.

What makes this topic more interesting—and sometimes exceptionally challenging—is that *all else is not equal*. The rest of this section will discuss some of the real economic effects of settlement.

Primary incentives Suppose, for the sake of argument, that the potential defendant can take precautions at an *ex ante* stage to reduce the probability of an accident. Suppose further that all injured victims—the future plaintiffs—have positive expected value claims and that there is symmetric information during bargaining. These assumptions, while clearly unrealistic, allow us to isolate some of the basic effects of settlement on primary incentives.

Following an accident, the defendant is better off if he has the option to settle his claim. This is his revealed preference—if he were made worse off by settling he could simply refuse to settle and go to trial instead. Since the defendant anticipates settling on relatively advantageous terms, he has less incentive to take precautions to avoid the lawsuit to begin with. Simply put, settlement dilutes the defendant's incentives for care. See Polinsky and Rubinfeld (1988a). Perhaps surprisingly, the fact that the defendant takes less care is not necessarily a bad thing from a social welfare perspective. To see why, we will consider a specific example.

Suppose that damages are compensatory in the sense that x reflects the true harm that the plaintiff has suffered in the accident. If settlement were prohibited, then the socially optimal level of the defendant's precautions would reflect all of the private and public litigation costs associated with accidents in addition to the plaintiff's injuries. With compensatory damages, the defendant clearly underinvests relative to the social optimum—the defendant doesn't take into account all of the costs that the accident imposes on others, namely the plaintiff's litigation costs and society's costs of running the court system.

The desirability of settlement in this setting hinges on the defendant's bargaining power. Let's take the extreme situation where the plaintiff has all of the bargaining

power: the defendant pays in settlement only slightly below what he would pay in total if the case went to trial, $x + c_d$. In anticipation of settlement, the defendant takes only slightly less care than before. In this case, settlement significantly reduces the litigation costs—a first-order improvement for social welfare—but incentives are not compromised. It follows that settlement enhances social welfare so long as the defendant does not have "too much" bargaining power. If, on the other hand, the defendant has significant power in negotiations then the effect on incentives could potentially outweigh the benefit of cost savings. The defendant, anticipating settling for $S = x - c_p$, would invest significantly below the first-best level. If $x - c_p$ is close to zero, for example, then the defendant has little incentive to take precautions at all.

Settlement can further dilute the defendant's incentives when asymmetric information is present. Recall that the screening model of settlement with asymmetric information featured some pooling of types. Consider, for example, a model where the uninformed plaintiff makes a final offer to a defendant who is privately informed about his liability for the accident. In equilibrium, defendants whose types (corresponding to expected liability) are above a threshold accept that offer. The pooling of defendant types may be bad for incentive reasons: the defendant has little marginal incentive to reduce his liability if he anticipates being in the pool. Spier (1997) presents a simple example along these lines where there is "too much" settlement in equilibrium. Incentives would be improved—and social welfare would be higher—if the plaintiff could commit to being tougher in settlement negotiations, making higher settlement offers to the defendant.[30]

Bringing suit The preceding discussion of primary incentives focused on the defendant's precaution decisions. We will now turn to the plaintiff's decision to bring suit. Following an accident, the plaintiff is made better off through settlement than she would be going to trial. Again, this is revealed preference—if the plaintiff were made worse off by settling she could simply refuse to settle and go to trial instead. Since the plaintiff expects to settle on relatively advantageous terms, she has a greater incentive to bring suit to begin with.[31] We will see that the plaintiff's increased incentive to pursue litigation may be either good or bad from a public policy perspective.

First, consider the effect of settlement on total litigation costs, taking the defendant's precautions as fixed. The ability to settle out of court surely reduces the private and social costs associated with *a given case*. (The cost will not be driven to zero, however; we have seen that asymmetric information is a robust obstacle to settlement and litigation costs are incurred in equilibrium.) At the same time, the ability to settle raises the *overall volume of cases* that are pursued. Indeed, the additional litigation costs associated

[30] In Spier's equilibrium, the defendant randomizes between taking due care and being negligent and the plaintiff randomizes between high and low settlement offers.

[31] This may be exacerbated when an uninformed defendant is making a final offer to a plaintiff who privately observes his damages. The screening equilibrium features a cutoff where plaintiffs with damages below the cutoff accept the offer. The pooling of plaintiff types may be bad for incentive reasons as well: plaintiffs who essentially have no injuries at all may succeed in extracting positive offers of settlement.

with the increase in cases may swamp the reduction in litigation costs associated with existing cases. In sum, settlement can lead the total litigation costs to either fall or rise.

Next, recall that the defendant's primary incentives are, of course, endogenous. Thus, a change in the plaintiff's incentives to file suit can generate additional indirect effects by affecting the defendant's incentives for care. Section 2.1.3 considered the social desirability of litigation, assuming that all cases that were filed went all the way to trial. We saw that there could be either too much litigation activity or too little litigation activity, depending on the nature of the feedback effects on the defendant's incentives. Bringing settlement into that analysis does not simplify matters. To illustrate, suppose that there were too many cases before allowing settlement because the costs of litigation were low relative to the benefit of improved incentives. Allowing settlement raises the number of cases further—a bad thing if we hold all else equal, even though the litigation costs associated with a given case fall.

Discussion The preceding discussion of the normative implications of settlement models showed that although there are a number of robust *positive implications* of settlement models, the *normative implications* depend very much on context. The well-established topics of nuisance suits, the shifting of legal fees, and accuracy—all topics that will be discussed in detail in section 3—raise normative issues along the lines discussed above. There is still more work to be done on the desirability of settlement from a *forward looking perspective*. Litigation is, by nature or design, a public good. Judicial decisions from early cases can influence the future in a variety of ways. In common law regimes, the opinions of judges—not the private settlement contracts—determine how the law itself evolves. Litigation can create social value by promoting the efficient evolution of laws to govern *future* economic activity. Trials may also create information that has independent economic value. Individuals who have suffered injuries may benefit directly from the groundwork laid by earlier claims.[32] The public also learns about the hazards of products and risky activities through the litigation activities of others.[33] A consumer, for example, may refrain from purchasing a risky product after observing the harms that similar products have caused to others.

3. Topics

3.1. Accuracy

It is generally thought that accuracy is valuable when imposing sanctions on offenders (whether they are criminals, tort offenders, or violators of private contracts). Intuitively,

[32] See also the discussion of secret settlements and the publicity effect in section 3.

[33] Hua and Spier (2005) consider a model where future actors use this information to fine-tune their accident avoidance behaviors. In general, the settling parties do not fully internalize the benefit that litigation has on these future actors.

if the legal rules are designed appropriately then the anticipation of accurate adjudication should create better incentives. Furthermore, accuracy should also help encourage innocent activities. Without it, individuals would tend to avoid engaging in value-creating activities that could be mistaken for violations and subsequently sanctioned. But accuracy does not come for free: there are significant private and public costs associated with designing more accurate legal rules. It is therefore important to identify circumstances where the additional costs of creating accurate outcomes are outweighed by the social benefits of accuracy.[34]

Suppose that an injurer (the potential defendant) must decide whether or not to engage in an activity. If the injurer takes no precautions, then the benefit to the injurer from engaging in the activity is b and the harm suffered by the victim (the potential plaintiff) is h. The injurer can avoid harming the victim if he invests e in precautions. The first-best outcome has the injurer taking precautions when his cost of taking them is smaller than the harm to the victim, $e < h$, and engaging in the activity when his private benefit is sufficiently high, $b \geq \min\{e, h\}$. This outcome could be obtained easily in a perfect world where litigation is accurate and free.[35] A damage rule that specifies compensatory damages where the victim is "made whole" would lead the injurer to make the correct cost-benefit tradeoff.

Suppose instead that liability is determined with error. While it is obvious whether or not the injurer engaged in the activity, the precaution taken by the injurer (and the associated harm borne by the victim) is not directly observable in a court of law. If the injurer took precautions, there is still a probability θ_1 that he will be held liable for damages, h. This is a "type 1 error"—the probability that an innocent person will be convicted. If the injurer failed to take precautions there is a probability θ_2 that he will get away with it. This is a "type 2 error."

Legal error distorts the injurer's decisions in this example. First, the injurer will take precautions if his expected cost from taking them, $e + \theta_1 h$, is smaller than his expected liability if he fails to take them, $(1-\theta_2)h$. With a compensatory damage rule, the injurer will take precautions when $e < h(1 - \theta_1 - \theta_2)$. His decision is distorted because the type 1 error increases his cost of taking precautions while the type 2 error reduces his liability if he fails to take them. Second, the injurer will engage in the activity when $b \geq \min\{e + \theta_1 h, (1 - \theta_2)h\}$. Thus, the type 1 and type 2 errors clearly distort the level of economic activity as well.[36]

[34] See Kaplow and Shavell (1994) for an analysis of this tradeoff when liability is measured with error. See Kaplow (1998) for an excellent discussion of the many issues relating to accuracy more broadly.

[35] Polinsky and Shavell (1989) present a model where litigation is both inaccurate and costly. The litigation costs introduce an issue that is not addressed here: the decision of the victim to bring suit. They consider public policies such as imposing fines on losing plaintiffs to discourage their bringing suit.

[36] It is important to note that these distortions cannot be alleviated through a simple damage multiplier. A multiplier $\alpha = 1/[1 - \theta_1 - \theta_2]$ leads to the correct precaution incentives. However, this multiplier inflates the cost to the injurer of engaging in the risky activity to begin with, leading to a chill on economic activity. Additional instruments may include subsidies to those found innocent. See P'ng (1986).

To illustrate the problems associated with acquitting the guilty, suppose that $\theta_2 > 0$ and $\theta_1 = 0$ (so there is no chance that an innocent person will be convicted). This gives rise to two distortions: the injurer takes too few precautions and he engages in the activity too often. On the other hand, to illustrate the problems with convicting the innocent, suppose that $\theta_1 > 0$ and $\theta_2 = 0$ (so there is no chance that a guilty person will be acquitted). As before, the injurer takes too few precautions. In contrast, however, he will now engage in the activity too little rather than too much. The type 2 error—the chance that an innocent person will be convicted—discourages the relevant economic activity.

The previous example took the victim's harm level, h, as fixed. In reality, the injuries that arise due to the negligence of injurers are stochastic in nature. First, the injurer's failure to take precautions will not always cause an accident. Second, the victim's damages conditional upon an accident occurring are variable as well. Kaplow and Shavell (1996) argue that the *ex post* accurate verification of the victim's harm level may, or may not, be socially valuable in this setting. Accuracy is valuable if the injurer knew (or should have known) the victim's damages at the time when he chose his precaution level: the anticipation of an accurate award *ex post* gets the injurer to make the correct tradeoff *ex ante*. Accuracy is not valuable, however, if the victim's damages are purely stochastic and could not have been known by the injurer *ex ante*. There is no loss from setting the damage award equal to the expected or average harm in this case.[37] Indeed, when litigation is costly it is socially wasteful to devote resources to accurate outcomes when there is no corresponding benefit.

This section has highlighted two benefits of accuracy: deterrence and the encouragement of innocent economic endeavors. There are other benefits of accuracy as well that are important, but which fall outside this simple taxonomy. Accuracy may be valuable for risk-sharing reasons. If our victim above is risk averse, an accurate system that makes the victim whole is socially valuable because it reduces the risk premium that the victim would otherwise bear. Accuracy is also socially valuable when it creates better information, and therefore better incentives, for future actors. In a repeated litigation environment, the information created by earlier trials may help actors fine-tune their actions in the future.[38] Finally, the anticipation of precise investigations in the future will get injurers to do their homework ahead of time—they have a strong incentive to learn about the injuries that their risky activities cause.[39]

[37] Spier (1994c) argues that accuracy is valuable when precautions affect the magnitude as well as the probability of an accident.
[38] Hua and Spier (2005).
[39] Kaplow and Shavell (1992).

3.2. Evidence

3.2.1. The burden of proof

Black's Law Dictionary defines the burden of proof as a "legal device that operates in the absence of other proof to require that certain inferences be drawn from the available evidence ..." It is useful to think about this burden as having two parts. First, there is the so-called "burden of production" where a party must present sufficient evidence in favor or his claim or risk automatically losing without a full trial. Second, there is the burden of persuasion (or standard of proof) which provides the judge or jury with guidelines for how strong the evidence must be in order to find for the plaintiff (or prosecutor in a criminal proceeding). This burden is typically described in qualitative terms: "beyond a reasonable doubt," "clear and convincing evidence," "preponderance of the evidence," etc. A favorable decision for the plaintiff under the preponderance of the evidence standard, for example, is the point where the judge or jury believes that it is more likely that the plaintiff is in the right. In Bayesian terms, this corresponds to a posterior tipping point of 50%. "Beyond a reasonable doubt" is more subjective—what is "reasonable" is in the eye of the beholder.

The topic of evidence can be studied with the frameworks and techniques of information economics. Consider the following basic moral hazard example. An agent chooses a level of effort, either low effort (e_L) or high effort (e_H), which is not directly observed by the principal. The agent's choice generates a stochastic signal, s, which is distributed according to density function $f(s, e)$. The signal is informative in the sense that higher signals strengthen the decision maker's posterior belief that the effort taken was e_L rather than e_H.[40] Formally, this corresponds to the monotone likelihood ratio condition.[41] In our legal example, the signal s may be interpreted as the "evidence" suggesting the defendant's guilt. Correspondingly, the burden of persuasion (or standard of proof) corresponds to a cutoff, s^*, where evidence below the cutoff leads to acquittal and evidence above the cutoff leads to conviction.

This framework yields a stark implication for the design of both evidentiary standards (s^*) and the level of sanctions (D). Suppose that the cost of high effort is 1 and the cost of low effort is 0. The defendant will choose high effort so long as his cost of choosing high effort, 1, is smaller than the reduction in the expected sanction associated with the higher effort level,

$$1 \leq [F(s^*, e_L) - F(s^*, e_H)]D.$$

[40] Readers familiar with principal-agent models will note a subtle difference, here. In this example, a higher signal is associated with greater evidence of guilt, and hence *lower* effort on the part of the defendant. This is contrary to most standard models where higher signals are generally associated with *higher* effort. This distinction is highlighted to minimize confusion.

[41] Formally, the monotone likelihood ratio condition holds that $f(s, e_L)/f(s, e_H)$ is increasing in s for $e_L < e_H$.

Any pair of policy instruments, s^* and D, that satisfy this expression will provide the defendant with adequate incentives for care. But there are additional concerns associated with wrongful convictions. The monotone likelihood ratio property (MLRP) implies that by simultaneously raising the standard of proof, s^*, and the sanctions, D, perfect deterrence may be achieved at zero social cost. This is, of course, a variant of Becker's (1968) famous enforcement result—that penalties should be maximal while very little money should be spent apprehending criminals.

There are several reasons why we do not observe these extreme schemes in practice. First, civil sanctions are limited by a defendant's wealth and criminal sanctions are limited by a defendant's remaining lifetime. A significant increased probability of conviction is required to maintain incentives.[42] Second, juries may be unwilling to convict defendants when the sanctions are very high. Indeed, the subjective nature of criminal standards of proof (e.g., "beyond a reasonable doubt") gives the decision maker the discretion to define for himself what is reasonable. In Andreoni (1991), when sanctions are increased the jury convicts less often to avoid the higher cost of convicting the innocent (the cost of a type I error goes up since the innocent defendant is imprisoned longer). Since juries consequently convict less often, higher penalties may encourage rather than discourage crime. Third, the simple example above assumed that the evidence was exogenously generated. In practice, evidence is both costly to gather and subject to manipulation and misrepresentation.[43]

Rubinfeld and Sappington (1987) present a framework where the defendant chooses a level of litigation effort to influence the signal received by the court. They assume that defendants who are in fact innocent have "more productive" effort than their guilty counterparts. Innocent defendants consequently spend more money on their defense, endogenously leading to a better distribution of evidence at trial. The optimal social policy in this setting, which includes both the evidentiary standard (s^*) and the level of sanctions (D), will be chosen to balance the litigation costs and *ex ante* deterrence concerns. The defendant—whether innocent or guilty—will spend more money when the sanctions are higher and it is shown that the optimal policy has less than maximal sanctions.[44]

Sanchirico (1997) presents a model where plaintiffs, as well as defendants, make investments in their cases. Increasing the standard of proof for conviction has a good side benefit: plaintiffs who know that the defendant is innocent will choose to not file

[42] See Demougin and Fluet (2006) for an excellent discussion of optimal rules of evidence under wealth constraints. They argue that the penalty scheme that creates maximal incentives resembles the standard of preponderance of the evidence. Formally, rules that penalize an injurer when the evidence is "more likely" under negligence than due care (and conversely does not penalize the injurer when the evidence is "more likely" under due care) make the incentive compatibility constraint easier to satisfy.

[43] Other reasons include risk aversion, marginal deterrence, and heterogeneous injurers. See Bebchuk and Kaplow (1992) and the references therein.

[44] Their results are quite sensitive to the signaling technology of their model. See also related work by Sanchirico (2001).

suit, while plaintiffs who know that the defendant is guilty will litigate. In other words, the "self selection" may enhance social welfare. Bernardo et al. (2000) put additional structure on the litigation technology and find that, for evidentiary standards in an intermediate range, making a rule more pro-defendant can lead to more shirking and more litigation than before.

Hay and Spier (1997) assume that the body of evidence is fixed and not subject to manipulation. However, the two litigants may have different costs of acquiring or presenting the information. The burden of production may be viewed as assigning to one party the task of presenting the evidence to the court (and relieving his opponent to some extent of that task). Optimally used, this burden may minimize the expenditures devoted to gathering, presenting, and processing information in litigation. In practice, it is typical for plaintiffs to have the initial burden of producing enough evidence of the defendant's involvement in the case and indication of wrongdoing to justify proceeding. This makes sense in settings where there are innocent explanations for the plaintiff's injuries. Counterexamples exist, however. The taxpayer bears the burden of proof in income tax deficiency actions brought in the tax court by the IRS, for instance. This makes sense because the taxpayer would typically have greater access to evidence concerning his financial affairs than would the IRS.[45]

3.2.2. Disclosure and discovery

The basic framework presented in section 2 argued that asymmetric information between litigants about the likely outcome of the lawsuit could lead settlement talks to break down. The inefficiencies associated with bargaining impasses raise the question: "Why doesn't all of the information in the litigant's possession come out before trial?" Indeed, litigants will often voluntarily share information with each other before trial. An injured plaintiff, for example, may submit evidence of injury (x-rays, doctor's reports, etc.) to the defendant at the onset of filing to prove that she has a legitimate claim. In addition, in the United States and elsewhere there are laws that require litigants to disclose information when specifically requested to do so by the other side. An "interrogatory" is a set of questions that one side submits to the other side (and that must be answered) before trial. In a "deposition," lawyers may interview the other side's witnesses under oath.[46]

At an intuitive level, both voluntary and involuntary disclosure of evidence can serve important social objectives. First, the sharing of information before trial puts the parties on a more level playing field in the courtroom. This may well help to improve the accuracy of court decisions by making the game less one of rhetoric and more a fair contest based on the facts. Second, disclosure and discovery help to align the beliefs of

[45] See for example, Portillo v. Commissioner of Internal Revenue, 932 F2d 1128, 1133 (5th Cir. 1991).

[46] See Cooter and Rubinfeld (1994) for a good discussion of these institutions and a simple non-Bayesian model of disclosure where discovery can eliminate the "false optimism" of the litigants.

the parties about what will happen at trial, thereby facilitating private settlement. Critics of legal discovery point to its abuses, however, such as the ability of litigants to impose unfair costs on the other side. We will consider each of these issues in turn.

Trial outcomes Intuitively, both sides in a lawsuit have incentives to hide information that harms their case and to present evidence that helps their case. Discovery involves a set of formal rules and procedures that compel each side to share evidence that may, in the absence of discovery, never make it into the courtroom. It is thought that, in such a world, discovery may level the playing field by giving both sides access to the same information and thus improve the accuracy of legal decisions. While these ideas have some intuitive appeal, the formal models of disclosure of evidence can be subtle and often contradict this intuition.

Suppose, as in section 2, that a defendant privately observes his expected liability, represented by a parameter x drawn from a probability density function $f(x)$ on $[\underline{x}, \bar{x}]$. The defendant can costlessly and credibly reveal this information at trial. The court, a Bayesian player, does not receive any independent signals regarding the realization of x, but does know the distribution from which x is drawn. The defendant with the strongest case, $x = \underline{x}$, has an obvious incentive to disclose his innocence to the court and thereby secure an advantageous judgment. Indeed, since a Bayesian judge or jury would make an adverse inference if the defendant remained silent, the defendant with a slightly worse case, $x = \underline{x} + \Delta$, has an incentive to disclose this as well. This reasoning suggests that all information comes out at trial and an "accurate" outcome is obtained. This tendency for unraveling is familiar from the classic economic analyses of product quality and warranties (Grossman, 1981).

Grossman's unraveling result—that accuracy at trial is not compromised by the private incentives to withhold evidence—is sensitive to the underlying assumptions of the model.[47] First, not all defendants are able to reveal their private information—in many cases, hard evidence of innocence does not exist. Those defendants who do have hard evidence of guilt (they are in possession of a "smoking gun," perhaps) have an incentive to pool with the general population of defendant types who are simply unable to signal their cases' quality. See Shavell (1989a). Similarly, if the defendant has costs of disclosing evidence at trial then the unraveling will be incomplete in a world of voluntary disclosure.

With legal discovery, which mandates that the defendant must submit to interviews and answer questions before trial, it is more likely that the plaintiff will gain access to the information that the defendant would otherwise withhold. For example, the plaintiff's discovery activities may succeed in finding the "smoking gun," leading to the conviction of a defendant who otherwise would get off the hook. See Hay (1994) for further discussions and examples. In a complementary piece, Cooter and Rubinfeld (1994)

[47] See also Shin (1998), Shin (1994), and Sobel (1985).

argue that discovery may also improve accuracy as a consequence of "eliminating surprises" in the courtroom. Surprises at trial would typically lead to more spontaneous and less thought-out courtroom activities.[48]

Lewis and Poitevin (1997) present a model of costly disclosure to regulatory tribunals where disclosure creates a signal that is imperfectly correlated with the true state of affairs. With voluntary disclosure, they show that a litigant's decision to disclose information is itself a signal of strength: in equilibrium, strong litigants disclose and weak litigants do not (accepting the fact that non-disclosure will identify them as weak). Paradoxically, mandatory disclosure can reduce the accuracy of the ultimate court decision. When disclosure is voluntary, the weak types identify themselves by not sinking the costs of disclosure. When disclosure is mandatory, the court loses an important signal and some of the weak defendants may be exonerated.[49]

Settlement behavior To see the roles that disclosure and discovery play in private settlement negotiations, we return to the simple example from the section 2: A defendant privately observes his expected liability at trial, represented by a parameter x drawn from a probability density function $f(x)$ on $[\underline{x}, \bar{x}]$. Recall that if the uninformed plaintiff can make a single take-it-or-leave-it offer before trial (Bebchuk, 1984), the equilibrium will be characterized by a cutoff $\hat{x}$ and a settlement offer $S = \hat{x} + c_d$ where defendants whose types are above $\hat{x}$ will accept the offer to settle and those below will reject the offer and go to trial.

The game changes considerably when the defendant can credibly and voluntarily disclose his private information before the plaintiff makes the settlement offer. At first blush, one might think that the typical "unraveling" results would hold. A defendant who expects to pay the lowest damages at trial (type $\underline{x}$) is happy to disclose the information to the plaintiff and secures a low offer of settlement ($S = \underline{x} + c_d$). Continuing with this logic, it appears that defendants with slightly weaker cases will disclose as well. Perhaps surprisingly, complete unraveling does not arise in this setting. Shavell (1989b) shows that there is an equilibrium where defendants with types below $\hat{x}$ reveal their private information while those with types above $\hat{x}$ keep it hidden. The plaintiff, believing that the silent defendants come from the truncated distribution above $\hat{x}$ offers to settle with these types for $S = \hat{x} + c_d$. By remaining silent, a defendant with type

[48] See Cooter and Rubinfeld (1994, p. 446).

[49] Jost (1995) assumes that the defendant's disclosure is not credible in itself, but may be verified through discovery by the other side (which is analogous to costly auditing). He also takes the penalty structure, including penalties for misrepresentation, as exogenous. He shows that the defendant does not truthfully reveal his information in equilibrium. If he did, then the plaintiff would take everything he says at face value and not bother to spend time and money in discovery. And if the plaintiff does not audit, then the defendant would surely lie and pretend to have a strong case. Jost argues that it is better for a central authority, which has commitment power and a vested interest in long-run deterrence, to check the validity of the defendant's claims.

$x > \hat{x}$ earns information rents equal to $x - \hat{x}$.[50] Although unraveling is incomplete here, all cases ultimately settle out of court.

In contrast to the case of voluntary disclosure, all of the defendant's information would come to light if the plaintiff could somehow force the defendant to reveal his private information (through formal discovery channels, perhaps). By forcing the silent defendants to disclose their types, the plaintiff can tailor the settlement offer appropriately, offering $S = x + c_d$ instead of $S = \hat{x} + c_d$ to a defendant of type $x > \hat{x}$. To put it another way, the plaintiff can use discovery to "grab" the defendant's information rents of $x - \hat{x}$.

Shavell (1989b) shows that discovery plays a more important role when a fraction of the defendants are simply unable to disclose their private information before trial. As before, defendants who can prove that they have strong cases will voluntarily do so and receive low settlement offers. There is a group of defendants who remain silent, however: Some have strong cases but cannot credibly prove it to the plaintiff. Others have weak cases and decide to remain silent for strategic reasons. Mandatory disclosure plays an important role here because it "weeds out" the latter types—the weak defendants with credible information—from those defendants who are simply unable to reveal their information. Following discovery, the "silent" defendants have stronger cases on average, and so the plaintiff makes an even better offer than before. In this way, mandatory disclosure increases the rate of settlement.

Mnookin and Wilson (1998) present a model where discovery is both imperfect and expensive: each side can sink costs to get a more accurate signal of the opponent's type. These discovery efforts increase the probability of settlement by reducing the degree of asymmetric information. In their model, discovery is a public good: both the plaintiff and the defendant benefit (in expectation) from the discovery activities. The party engaging in the discovery benefits more, however, since one effect of discovery is to reduce the information rents captured by the other side.

There has been some empirical support for the idea that discovery facilitates settlement. Farber and White (1991) present an empirical analysis of 252 medical malpractice cases. Upon filing, it is likely that the plaintiffs were not well informed about the likelihood that the hospital or physician was negligent. Within this sample, 37% were dropped before trial, 58% were settled before trial, and 5% went to trial. Farber and White argue that the fact that so many cases were dropped following discovery indicates that the plaintiffs learned, through formal discovery channels, that the defendants

[50] The unraveling would be complete if the litigation costs were zero. See Hay (1994). If the defendant, rather than the plaintiff, made the settlement offer (as in Reinganum and Wilde's (1986) signaling model), then (with appropriate refinements) there could be complete unraveling even with positive litigation costs. The defendant with the strongest case would reveal it and offer $S = \underline{x} - c_p$ which would be accepted with probability one. In the absence of disclosure, this same offer would be accepted with a probability smaller than one. Sobel (1989) presents a model with two sided incomplete information where each litigant may be one of two types. He compares the set of equilibria that arise with mandatory discovery and no disclosure and discusses the voluntary disclosure case informally.

were not negligent. Similarly, the settlement of most of the remaining cases is consistent with the greater alignment of information following discovery.

The costs of discovery A private discovery request is, of course, costly and the costs of discovery are not typically internalized by the party who requests the information. Indeed, one side can force the other side to spend many months screening documents, sorting through private materials, and submitting to depositions. The potential for abuse here is obvious: discovery may be used as a strategic weapon. While the requesting party surely does not internalize all of the costs of discovery, he also does not internalize all of the benefits. Discovery, insofar as it increases the accuracy at trial, has the potential to create social benefits. The social benefits, which include increased deterrence and incentives for care, are enjoyed by society more broadly. It is therefore difficult to draw general conclusions about the desirability of discovery.[51]

Shepherd (1999) studies the time that litigants spend seeking discovery using a survey of attorneys in 369 federal civil cases.[52] He shows that defendants increased their discovery efforts, "tit-for-tat," in response to heightened discovery requests by the plaintiff. Interestingly, this "counterpunch" strategy was not observed for plaintiffs, who did not increase their requests in response to increased pressure from defendants.[53]

3.2.3. Admissibility of settlement negotiations at trial

Should the litigants' private settlement activities, including the offers that they make, their discovery requests, and the extent of their litigation expenditures, be admissible as evidence at trial? At first glance, one might assume that the answer is yes—after all, pretrial settlement activities can reveal valuable information. For example, in the section 2 we saw that litigants who possess more valuable information are more likely to forego settlement and litigate instead. Their failure to settle is an informative signal about the stakes of the claim. If the court has imprecise information to begin with, it can learn more about the truth—and consequently rule more accurately—when it can fully observe the settlement activities of the parties. It is perhaps a puzzle, then, that Rule 408 of the Federal Rule of Evidence in the United States prohibits the use of this information at trial.

Daughety and Reinganum (1995) provide an interesting economic rationale for the inadmissibility of settlement offers at trial: admissibility increases the rate of litigation.

[51] Schrag (1999) argues that placing limits on discovery reduces the costs of litigation and stimulates earlier settlement. For example, the 1983 revisions of rules 30, 31, and 33 of the Federal Rules of Civil Procedure placed limits on the number of depositions and interrogatories that each side could request of the others. In his model, increased discovery efforts do not unearth evidence *per se*, but instead influence the expected judgment directly.

[52] The results are similar when he also includes the time that litigants spend responding to discovery requests.

[53] These results suggest that the defendant's reaction curve slopes upward, while the plaintiff's slopes down. Shepherd (1999) also shows that hourly-fee attorneys made more discovery requests than their contingent-fee counterparts.

They extend the signaling model of Reinganum and Wilde (1986), where an informed plaintiff makes a take-it-or-leave-it offer, to include a Bayesian court. The court observes the settlement offer, updates its beliefs about the true state of the world and awards damages accordingly. As in Reinganum and Wilde (1986), there is a separating equilibrium where the plaintiff's offer reflects exactly the defendant's expected payments at trial. Incentive compatibility requires that the defendant mix between accepting and rejecting the offer, and that the probability of acceptance is decreasing in the size of the offer. When compared with the case of inadmissible settlement offers, the plaintiff has an incentive to "exaggerate" her offer and pretend to be of a higher type—by doing so, she influences the court and secures a higher damage award. Consequently, the defendant must reject with a higher probability than when settlements are inadmissible.[54] This argument implies that the litigation costs are lower when settlement offers are inadmissible at trial.

3.3. Sequential litigation

3.3.1. Appeals

An important characteristic of most legal systems is the right of a litigant who is dissatisfied with a lower court's decision to seek reconsideration by a higher court. This is true with civil, criminal, and administrative procedures in the United States as well as legal systems in other countries.

Shavell (1995) considers a stylized model where appeals are an efficient means of correcting the errors made at the lower court level. It does this by harnessing the private information of the litigants themselves. It is assumed that the litigants know whether a lower court ruling was in error or not and may launch a costly appeal. Shavell assumes that an incorrect decision is more likely to be overturned by the higher court than a correct decision. Litigants will tend to self-select in this environment: a litigant is more likely to sink the cost of appealing an earlier ruling if the probability of reversal—and hence the expected return from an appeal—is higher. This tendency to self-select is not perfect, however. When the cost of appeal is too low, then a litigant will appeal whether or not the lower court had rendered a correct decision. (Conversely, when the cost of appeal is sufficiently high, then no cases will be appealed.) By choosing an appropriate subsidy or tax, however, a social planner can align the litigants' appeal decisions with those of society more broadly.

An interesting implication of Shavell's analysis is that increasing the accuracy of lower court decisions is not a perfect substitute for the appeals process. "Increasing trial court accuracy reduces the frequency with which the appeals process is needed

[54] Kim and Ryu (2000) consider a model where the uninformed defendant makes the final settlement offer. They show that if the plaintiff's acceptance/rejection decision is admissible then the litigation rate will rise as well. Intuitively, the plaintiff would have an additional incentive to reject the offer to "convince" the court that she has a high type.

but not its desirability when errors are made."[55] The benefit of using the appeals system hinges on its ability to harness the information of the litigants themselves. By getting the litigants to self-select, resources tend to be spent on cases where a mistake has already been made. Similarly, the appeals system is not a perfect substitute for random audits performed by the upper level courts.[56]

It is important to note that Shavell's (1995) upper-level court is not a Bayesian decision maker. Indeed, if the upper level court were fully rational it would realize that only "mistakes" are appealed, and would therefore rule in favor of the appellant. This would, of course, interfere with Shavell's self-selection equilibrium: if the upper level always found in favor of the appellant, then all losers in the lower court—correct and incorrect decisions alike—would find it in their interest to appeal.

Daughety and Reinganum (2000a) consider a Bayesian model of appeals where the upper court perceives the private decision to appeal as informative and tries to rule "correctly" given its posterior beliefs. Formally, the authors assume that the both the appeals court and the litigants themselves receive private signals that are correlated with the truth. Technically, their signals are "affiliated" random variables (Milgrom and Weber, 1982). In equilibrium, a losing party appeals if and only if his or her signal exceeds a threshold. The upper court subsequently finds for the appellant and overturns the lower court's decision if and only if their own private signal exceeds another threshold.[57]

3.3.2. Bifurcation

Legal systems often feature a sequence of decisions before a final judgment is reached. Appeals systems, mentioned previously, allow for the re-litigation of an issue if one of the litigants is dissatisfied with a lower court's decision. In many other settings, there are sequential decisions on different issues before the final judgment. In criminal procedures, guilt is typically established before hearings to determine the convicted defendant's sentence take place. In products liability settings, proof must first be offered that the defendant was indeed the manufacturer of the product that caused the plaintiff harm before issues of negligence and damages can be considered.

[55] Shavell (1995, p. 387) notes that, in addition to efficiently reducing lower-court error through self-selection, the appeals process may also improve accuracy by providing lower court judges with better incentives to make careful and well-reasoned decisions. The idea is that judges dislike being reversed on appeal, and therefore they will devote more effort to their decisions and show less favoritism than otherwise.

[56] See Spitzer and Talley (2000) for a formal model of judicial auditing. They argue that higher levels of auditing are warranted when the lower courts are: (1) less accurate and (2) more swayed by ideology than by the facts of the case.

[57] Daughety and Reinganum (1999b) extend this logic to a horizontal sequence of courts facing similar cases, where each court along the chain receives an affiliated signal. While no appeals court's decision is precedential in another circuit, judges may view previous decisions in other circuits as a source of persuasive influence. They show that a herding phenomenon can arise where the earlier courts place more weight on their own private signals, wile the later courts discount their private signals in favor of the earlier courts' rulings. Herding is also referred to as an "informational cascade." See the survey by Bikchandani, Hirschleifer, and Welch (1998).

Landes (1993) presents the first formal analysis of the incentives to file, settle, and spend in bifurcated versus unitary trials.[58] In a "bifurcated" trial, the court first establishes the defendant's negligence before the plaintiff's damages are considered. In a "unitary" trial, the court determines both issues at the same time. His is an optimism framework, where the defendant and the plaintiff have potentially different estimates of the probability that the plaintiff will prevail on liability and of the damage award conditional upon liability being established.

When the private costs of establishing liability and damages are exogenous and settlement is not possible, then, conditional upon the plaintiff filing suit, bifurcation leads to lower litigation costs than a unitary trial does. The reason is simple: once the defendant is absolved of liability then the case is over and no further costs are incurred. Landes points out, however, that as a consequence of these litigation cost savings the plaintiff will file more suits than she otherwise would. Even though the litigation costs per case will fall, the number of cases will rise so that the overall effect on litigation costs is ambiguous.[59]

At first glance, it appears that bifurcation would have an added advantage over unitary trials when the plaintiff and defendant can settle their claims. The argument would go something like this: suppose that the plaintiff and defendant have the same assessment of the plaintiff's damages although they disagree about the defendant's liability. In a bifurcated proceeding, the parties would surely settle on damages following a plaintiff victory on liability. Therefore the litigation costs to establish damages in a unitary trial would avoided. The fallacy in this reasoning, as Landes points out, is that the parties could settle their damage dispute before a unitary trial as well, transforming a unitary trial in which both liability and damages are determined into one where only liability is considered. Therefore we would not expect that sequential and unitary trials would differ significantly on the propensity of private parties to settle.

Chen, Chien, and Chu (1997) reconsider Landes' questions in a model with asymmetric information instead of mutual optimism. In their model, the defendant is privately informed about both the probability that he will be found liable and the damages (conditional upon a finding for liability). With a unitary trial, plaintiffs run into difficulties making low settlement offers to the defendant. Since the cost of proceeding to a unitary trial is large, the plaintiff must maintain credibility not to drop the case following the rejection of a low settlement offer (as in Nalebuff, 1987). With a sequential trial, however, it is easier for the plaintiff to maintain credibility. Since the plaintiff will have another opportunity to settle before the stage in which damages are determined, the plaintiff's cost of proceeding is lower than in the case of a unitary trial. Chen, Chien, and Chu show that the overall effect on the settlement rate is ambiguous.

Finally, interesting insights are obtained when the costs of litigation are assumed to be a choice variable for the two parties. Landes (1993) provides a nice discussion of these

[58] However, there is a related informal discussion in Schwartz (1967).

[59] White (2002), in her analysis of asbestos trials, shows bifurcation raises the plaintiffs' expected returns and increases the number of cases that are filed.

issues, and Daughety and Reinganum (2000b) analyze and explicitly model endogenous litigation costs. It is clear, conditional upon the plaintiff winning in the liability stage, that both litigants will spend more in the damages stage. Intuitively, in a unitary trial the stakes for damages are smaller because the damages are discounted to reflect that chance that the plaintiff will lose on liability. According to Landes (1993), bifurcation transforms what was a fixed cost in a unitary trial into a variable cost in a bifurcated trial. The overall effect on the expected costs spent establishing damages is ambiguous, however, since the higher costs are borne less often in the bifurcated trial.

Interestingly, with endogenous litigation expenditures, bifurcated trials tend to favor defendants over plaintiffs.[60] To see why, consider the incentives of the two litigants to spend money in the liability stage of the bifurcated trial. Looking forward, the defendant has larger stakes than the plaintiff since the defendant expects to pay $x + c_d$ in total if he loses on liability while the plaintiff expects to receive $x - c_p$. To put this somewhat differently, the anticipated litigation costs in the damages stage drive a wedge between the plaintiff's and defendant's stakes in the liability stage. Consequently, the defendant marginal return from spending an extra dollar in the liability stage is higher than the plaintiff's and therefore the outcomes will be biased in favor of the defendant.

3.3.3. Collateral estoppel

A set of related rules and doctrines say when and whether a decision in one case will bind on another case *when at least one party is involved in both litigations*.[61] Under the doctrine of *Res Judicata*, the same claim between the same two parties may not be "re-litigated" (although parties may be permitted to appeal the outcome if they believe the court was in error). In criminal law, double jeopardy holds that a defendant cannot be tried twice for the same offense. Furthermore, the appeal rights under double jeopardy are asymmetric: the defendant has the right to appeal a conviction, but the prosecution may not appeal an acquittal. Related rules may apply when a single defendant (say) is facing a sequence of similar cases. For example, there may be several victims in an accident caused by a negligent truck driver. A finding of negligence in the first victim's case may bind for the second victim's case as well.

To explore some of the economic issues surrounding these rules, consider the following stylized example. A defendant, D, is facing a sequence of two plaintiffs with damages x_1 and x_2. Suppose that the two plaintiffs were injured in a highway collision with a truck and are bringing independent suits against the trucking company. The issues in both cases involve: (1) whether the trucking company was negligent and (2) the level of damages suffered by each plaintiff. The probability that the defendant will be found negligent by the court is p. We will consider two types of rules. The first, "2-sided

[60] See Landes (1993, pp. 22 and 41) and Daughety and Reinganum (2000b).

[61] This differs from precedent (although many of the economic issues are similar). First, precedent may hold even when the sequence of lawsuits involves different litigants. Second, it is often at the discretion of the judge whether to follow prior precedent or not (and perhaps create new precedent).

collateral estoppel," says that a finding of negligence in the first case would preclude relitigating the negligence issue in the second case. On the flip side, a finding that the defendant was not negligent in the first case would preclude the second plaintiff from bringing a case at all.

2-sided collateral estoppel Under the 2-sided rule described above, if the first plaintiff prevails, then the issue in the second case is one of determining damages alone. If the first plaintiff loses, then the second case is necessarily dismissed or dropped. This form of collateral estoppel makes a great deal of sense when the court's decision in the first case is unbiased and accurate. Litigation is expensive, after all, and it is a waste of resources to revisit the negligence issue once it has already been decided.

Several authors, including Spurr (1991), Katz (1988) and Che and Yi (1993) have shown that this rule leads to distortions when the court's decision in the first case depends on the litigation expenditures of the private parties. The defendant, as the long-run player, has higher stakes in the first case than does the first plaintiff—in fact, if the two plaintiffs have equal damages, then the defendant's stakes are twice as high. Consequently, the defendant's marginal return from additional litigation spending in the first case is higher. The divergence between the stakes, and the unequal litigation spending that results, will tend to bias the trial outcomes towards the defendant. As a result, the defendant's incentives to take care to begin with may be compromised. Note that this logic also suggests that the defendant would be more likely to appeal an adverse decision, thus exacerbating the bias. If given the choice, the defendant would also choose to litigate against the weaker plaintiff first. A plaintiff with low damages will spend less than a plaintiff with high damages, allowing the defendant to cheaply establish a favorable early ruling.

This two-sided rule could, potentially, influence the defendant's incentive to settle the first case. If the first case settles, then the court will hold the defendant liable with probability p if the second case goes to trial. Consequently, the settlement range before the second trial is $\{px_2 - c_p, px_2 + c_d\}$ where c_p and c_d are the litigation costs of the plaintiff and defendant, respectively. If, on the other hand, the first case goes to trial, then one of two scenarios will hold. If the defendant loses the first case, then the second plaintiff is sure to win on liability and the settlement range is $\{x_2 - c_p, x_2 + c_d\}$. If the defendant wins in the first round, then the second plaintiff has no grounds to bring a suit. There is no reason to think, *a priori*, that the defendant would derive a strategic benefit by settling with the first plaintiff rather than bringing the first case to trial. Indeed, if the plaintiff and defendant have equal litigation costs and equal bargaining power, then the "expected" settlement with the second plaintiff is px_2, whether the first case settles or not.

This result is sensitive to the particular assumptions about the bargaining power and litigation costs. If the defendant has all of the bargaining power, then he would prefer to settle the first claim. With all of the bargaining power, he can enjoy a settlement amount at the very bottom of the settlement range. If the first case settles, the defendant will settle the second claim for $S_2 = px_2 - c_p$. If the first case goes to trial and the

defendant loses, the defendant subsequently offers to settle for $S_2 = x_2 - c_p$. The defendant clearly prefers the former scenario because the expected value of the latter is $p(x_2 - c_p)$. The opposite conclusion holds when the plaintiff has all of the bargaining power. The plaintiff in this instance enjoys a settlement at the very top of the settlement range. If the first case settles, the second plaintiff demands $S_2 = px_2 + c_d$ and the defendant accepts this demand. If the first case goes to trial and the defendant loses, the second case settles for $S_2 = x_2 + c_d$. In expectation, the defendant pays $p(x_2 + c_d)$ in this litigation scenario, which is less than if he settled the first case.[62]

Che and Yi (1993) explore the interaction between settlement incentives and asymmetric information with collateral estoppel. In their model, the two plaintiffs have private information about their damages and so bargaining vis-à-vis the second plaintiff will be inefficient: the second case will go to trial with positive probability. The 2-sided collateral estoppel rule has both a negative and a positive impact on the defendant's incentive to settle the first case.

First, if the defendant is found negligent in the first case, then the fundamental stakes of the second case will rise from px_2 to x_2. The severity of the asymmetric information problem will rise as well, since there is "more for the plaintiff and the defendant to disagree about."[63] The defendant suffers from the certainty of negligence in two ways: (1) he bears his litigation costs more often since negotiations are more likely to fail and (2) he shares more of the surplus with the second plaintiff (the plaintiff gets information rents). On the other hand, if the defendant is found not negligent in the first case, then the stakes of the second case drop to zero—the second case will not be brought. In this instance, there are no litigation costs incurred and the second plaintiff receives nothing in the way of information rents.[64]

1-sided collateral estoppel Suppose instead that the rule was one-sided: a finding of negligence would apply in the second case but a finding of adequate care would not. This rule clearly serves to benefit the later plaintiffs as the case against the defendant is "ratcheted up" over time. The defendant still has a long run stake in avoiding a finding of negligence in the first case as before. But now he lacks the benefit from a finding of no negligence in the first case. The over-spending effect mentioned above is still present although in a mitigated form (see Spurr, 1991).

[62] The astute reader will notice the litigation costs are being held constant as the stakes increase in this example. In reality, larger cases have larger costs of litigation associated with them. If the litigation costs were exactly proportional to the stakes, then there would be no difference between settlement and litigation from the defendant's perspective.

[63] Suppose that the probability of a finding of negligence is 1/2 and that the second plaintiff's damages are uniformly distributed on the interval [100, 200]. If the first case settles, the stakes of the second case are distributed uniformly on the interval [50, 100]. Since the range of disagreement is smaller than before the second case is more likely to settle.

[64] Che and Yi also characterize conditions under which the settlement rate in the first round will rise or fall. The assumptions underlying their analysis are quite strong, however.

A short note in the Harvard Law Review (Note, 1992) considers the settlement effects of one-sided rules. Under these rules, the defendant clearly has a strong incentive to settle the first case. If he settles the first case, the stakes of the second case are simply px_2. If he litigates the first case and wins, the stakes of the second case are unchanged. But if he litigates the first case and loses, the stakes jump to x_2. Even absent the typical surplus created by litigation costs, there is a strong incentive for settlement in the first round. The paper argues that if the early plaintiffs are forward-looking and have bargaining power they will be able to extort more from the defendant in the settlement negotiations. Consequently, this one-sided rule will benefit early as well as later plaintiffs.

3.3.4. Precedent

A feature of Anglo-American legal systems is that legal rules can be created and changed by judges over time. The presence of earlier rulings on particular issues provides judges with a reason for ruling in the same way when new cases arise with similar issues. This section will mainly focus on the roles and incentives of judges in making laws over time. Private litigants with long-run interests also influence the evolution of the common law, of course.[65]

At first blush, adherence to the precedent set by earlier cases can create value for two reasons. First, past decisions embody useful information for future decision making, and so precedent will tend to lead to more accurate court decisions. (This is especially true if judges lack the expertise or the time to make accurate decisions in isolation.) Second, economic actors value the predictability that accompanies strong precedents. Predictability in a legal system will facilitate the smooth operation of an economy because it reduces the scope for disagreement. Predictable laws should correspond to fewer disagreements over liabilities, rights, and obligations and therefore produce fewer legal disputes. Predictability can also create greater value *ex ante* since economic agents are more likely to engage in productive activities when property rights are well-defined and secure.

Landes and Posner (1976) interpret precedent as a productive capital stock, or productive input into future court decisions. As with other capital stocks, it is argued that the value of a body of precedent depreciates over time. In the words of Posner (1992), "... accident law that was developed to deal with collisions between horse-drawn wagons will be less valuable applied to automobile collisions." Depreciation may come from technological obsolescence (as with horse-drawn wagons) of economic activities over time.[66] The common law gives judges the flexibility to make new rules and laws

[65] The preceding subsection discussed related issues in the context of collateral estoppel. Hay (1993) presents an analysis of preclusion rules and the relitigation of claims.

[66] Landes and Posner (1976) look at citations to previous cases in a sample of 658 federal appeals court decisions. Although citations are surely not a direct measure of precedent, they arguably serve as a useful proxy for the influence of past cases. The authors found, among other things, that the capital stock of precedent depreciates slower when: (1) there is less statutory activity in the area (so the formal written laws are not

to respond to these changes. This flexibility raises many questions. For instance, does the common law evolve efficiently? Does private settlement of disputes prevent efficient evolution? What if judges are self-interested and lazy?

Cooter et al. (1979) present a relatively early formal model of legal evolution. Their model is one of a negligence regime where the courts learn about—and subsequently adjust—the standards of care for injurers and victims. In their framework, the courts can observe the social welfare function and make incremental adjustments to improve the state of affairs. The authors show that the incremental adjustment process will converge to social efficiency. Their analysis makes use of many explicit and implicit assumptions, some of which are quite strong. In particular, they assume that cases are constantly litigated and that the courts have the knowledge of the underlying economic model needed to make wise decisions. They also assume that the decision makers—the judges—are benevolent and have the interests of society in mind. Although they provide a good starting point, these assumptions are unlikely to hold all of the time in practice.

As discussed throughout this chapter, the vast majority of cases that are filed ultimately settle out of court. Others are resolved before the plaintiff files a case at all, either through settlement or through a decision by the plaintiff not to pursue the case at all. This latter case—where the plaintiff drops the suit or fails to file it—will typically occur when the costs of going to trial are significant and the expected return is small. In a nutshell, the cases that actually make it before a judge—and whose decisions could serve as precedent for the future—are a very select group of cases. Rubin (1977) argues that this sample selection would actually *work in favor of efficiency*. (See also the related arguments of Priest, 1977.) The reasoning behind this type of argument is that inefficient laws lead to dead-weight losses in future economic activity. Therefore, private parties have a greater incentive to bring these cases in order to change future laws. This is especially true of private parties with long-run interests in changing these laws.[67]

Landes and Posner (1976) discuss the potential problems associated with judges making socially inefficient decisions while in the pursuit of their own preferences and political agendas. They argue that such tendencies are kept in check by a simple mechanism: a judge who makes a socially inefficient (but privately desirable) decision is more likely to be overruled in the future. Being overruled can have important consequences for the long run because being overruled on one case may well undermine the weight given to the judge's other decisions, reducing his or her citations and influence.[68]

changing) and (2) when the prior rulings were from the Supreme Court (which presumably selects for cases with more general impact). In particular, the citations to Supreme Court cases were on average twice as old for other courts (20 years versus 10 years old).

[67] In a model with endogenous litigation expenditures, Katz (1988) also argues that parties with long-run interests will spend more money in pursuit of changing inefficient laws and mitigating the associated dead-weight losses.

[68] Miceli and Cosgel (1994) present a formal model where judges have preferences over two things: the outcomes of individual cases and their "reputations." They show that a judge may well stick with prior precedent

Rasmusen (1994) formalizes some of the interactions among a sequence of judges. Judges have personal preferences over laws, and want to establish precedents that will be followed by others in the future. He shows that there can be multiple equilibria in this dynamic framework. In one equilibrium, all judges pursue their own private preferences by overturning past precedents. This brings the judge private value in the short run but not in the long run. Other equilibria exist, however, where judges cooperate with each other over time through "trigger strategies." In these equilibria, judges follow past precedent closely because violations would lead to future breakdowns where their own precedents would be violated by others. Schwartz (1992) and Kornhauser (1992) consider the incentives and strategies of tribunals, or multiple decision makers, who are interacting with each other both in the short run (on a given case) and in the long run. Judges may well engage in strategic behaviors not unlike those observed in the political arena (congressmen trading votes and the like).

3.4. Allocating the costs of litigation

3.4.1. Loser-pays rules

Note that the expected judgment at trial, x (as defined in section 2) can be interpreted as the product of the probability that the defendant will be found liable and the plaintiff's damages. In this section, we can normalize the damages to 1, making then x simply the probability that the plaintiff will win. We previously assumed that each side paid for its own costs of litigation, a rule that applies to most litigation in the United States. In contrast, with the so-called English Rule the loser must pay for the winner's legal expenses. The plaintiff's payoff at trial in this instance may be written $x-(1-x)(c_p+c_d)$. The first term is the plaintiff's expected judgment at trial, x, and the second term reflects the plaintiff's expected litigation costs. With probability $(1-x)$ the plaintiff loses and is forced to pay the defendant's legal costs as well as her own. By analogy, the defendant's expected payments at trial may be written $x+x(c_p+c_d)$.

If there is complete information about all of the relevant variables, the bargaining range is simply $[\underline{S}^{ER}, \overline{S}^{ER}] = [x-(1-x)(c_p+c_d), x+x(c_p+c_d)]$. Notice that the size of the settlement range is exactly as it was for the American Rule in section 2.2: $\overline{S}^{ER}-\underline{S}^{ER}=c_p+c_d$. We will see that, compared with the American Rule, the English Rule has different implications for economic decisions. First, the English Rule changes the filing decisions of plaintiffs. Second, it affects the level of litigation spending. Fi-

against his personal preferences if the threat of reversal is sufficiently strong. He may deviate from precedent, however, if he views the outcome as sufficiently better for the case at hand or expects the decision to be upheld in the future. Levy (2005) presents a model where judges have career concerns and go against precedent to signal their abilities. See also recent work on precedent by Gennaioli and Shleifer (2005), Bustos (2006), Hylton (2006), Hadfield (2006), and Fon and Parisi (2003, 2004).

nally, it can change the litigation rate when the litigants' optimism and/or asymmetric information are taken into account.[69]

Filing decisions The English Rule changes the plaintiff's expected payoff at trial in a systematic way: it dilutes the value of low-probability-of-prevailing cases and enhances the value of high-probability-of-prevailing cases (Shavell, 1982a; Katz, 1990). This of course influences the plaintiff's decision about whether to pursue litigation to begin with. Assuming that all cases that are filed ultimately go to trial, a case is filed under the American Rule when the expected judgment exceeds the costs of litigation:

$$x > \tilde{x}^{AR} = c_p.$$

Under the English Rule, on the other hand, a case is pursued if and only if $x - (1 - x)(c_p + c_d) > 0$ or

$$x > \tilde{x}^{ER} = \frac{c_p + c_d}{1 + c_p + c_d}.$$

If $\tilde{x}^{AR} < \tilde{x}^{ER}$ then fewer cases are filed under the English Rule than under the American Rule; if $\tilde{x}^{AR} > \tilde{x}^{ER}$ then more cases are filed under the English Rule.[70]

To see this in a somewhat different way, notice that the plaintiff's expected litigation costs under the English Rule, $(1 - x)(c_p + c_d)$, exceed the litigation costs under the American Rule, c_p, if and only if $x < c_d/(c_p + c_d)$. This has important implications. At one extreme, when the plaintiff's probability of winning is low, then the English Rule can turn a viable case into a NEV proposition. In other words, the English Rule discourages low-probability-of-prevailing plaintiffs. At the other extreme, when x is high, the English Rule makes the plaintiff's case even stronger. In other words, the English Rule encourages high-probability-of-prevailing plaintiffs.[71]

Litigation spending Suppose that the litigation expenditures, c_p and c_d, affect the plaintiff's probability of success, x. Holding the defendant's expenditure fixed, the English Rule leads to greater litigation spending by the plaintiff for two reasons. First, there is an additional marginal benefit from greater spending since the stakes have increased from 1 to $1 + c_p + c_d$. Second, the marginal costs associated with spending are lower since the costs are partially externalized due to the fact that, under the English Rule, your opponent may be forced to pay your costs (Braeutigam, Owen, and Panzar, 1984; Hause, 1989; Katz, 1987).

[69] This discussion will focus on the theoretic literature. See Hughes and Snyder (1998) for a survey of the empirical literature in this area.

[70] $\tilde{x}^{AR} > \tilde{x}^{ER}$ if and only if $c_d < c_p^2/(1 - c_p)$.

[71] Kaplow (1993) and Polinsky and Rubinfeld (1996) discuss the normative implications of the English Rule and its effect on filing decisions. Polinsky and Rubinfeld discuss a more general set of rules that impose a penalty on losing plaintiffs and give a reward to winning plaintiffs.

Settlement behavior Suppose that the defendant is privately informed about x and, in particular, that he privately observes the probability that he will be found liable and that the uninformed plaintiff could make a take-it-or-leave-it settlement offer to him. As described in the section 2, any given settlement offer, S, corresponds to a cutoff value $\hat{x}$ where $S = \delta[\hat{x} + \hat{x}(c_p + c_d)]$. Defendant types whose liability is above the cutoff accept the offer and those whose liability is below the cutoff reject the offer and go to trail. The best screening offer corresponds to the cutoff, x^{ER}, that maximizes the plaintiff's expected payoff:

$$\int_{\underline{x}}^{x^{ER}} \delta[x - (1-x)(c_p + c_d)] f(x) dx + \left[1 - F\left(x^{ER}\right)\right] \delta\left[x^{ER} + x^{ER}(c_p + c_d)\right].$$

This gives the first-order condition:

$$\frac{1 - F(x^{ER})}{f(x^{ER})} = \frac{c_p + c_d}{1 + c_p + c_d}.$$

Comparing this expression to the first-order condition for the American Rule gives us a clear result: $x^{ER} > \hat{x}$. The cutoff under the English Rule is higher than the cutoff under the American Rule. It follows that more cases go to trial in equilibrium when the English Rule is adopted (Bebchuk, 1984). The intuition behind this result is simple. Trials occur in equilibrium because of asymmetric information about the outcome at trial—in this case, the defendant has private information about the probability that he will be found liable. This asymmetric information is exaggerated under the English Rule—now the parties are asymmetrically informed about who will bear the costs of litigation as well as the expected judgment at trial.[72] (Under the American Rule, the allocation of legal costs was common knowledge.) Intuitively, the English Rule creates more scope for disagreement between the two parties and the litigation rate consequently rises.[73]

The litigation rate will also rise with the English Rule when the informed defendant makes a take-it-or-leave-it offer to the plaintiff instead. As in section 2, we can construct a fully separating equilibrium with the following characteristic: the defendant signals his "type" through his offer of settlement, $S^{ER}(x) = \delta[x - (1-x)(c_p + c_d)]$ and the plaintiff randomizes between accepting and rejecting the offer. The interested reader can verify that an offer is accepted with probability:

$$\pi^{ER}(x) = e^{-(\bar{x}-x)\frac{(1+c_p+c_d)}{c_p+c_d}}.$$

[72] This result is driven by the fact that the defendant has private information about the probability of being held liable. If this probability were common knowledge and the parties were asymmetrically informed about the damages only, then the English Rule and the American Rule would lead to the same first-order conditions. Reinganum and Wilde (1986).

[73] Shavell (1982a) analyzes a model based on optimism. If parties are optimistic to begin with—so that the defendant estimates a lower value of x than the plaintiff estimates—then the English rule will exacerbate the problem. In this instance, the litigants are mutually optimistic about cost allocation in addition to damage awards.

Comparing this expression to the one for the American Rule shows that, given a defendant of type x, the acceptance probability is lower under the English Rule.[74] Again, the English Rule heightens the scope for disagreement: the defendant and the plaintiff now disagree about the allocation of costs in addition to the expected judgment at trial.

Several remarks are in order. First, the result that the English Rule raises the litigation rate may be reversed once litigation spending is taken into account. As mentioned above, we would expect the litigants to spend more money under the English Rule, and this will tend to create a greater incentive to settle. Second, the result is robust to changes in the information structure. If the plaintiff rather than the defendant were privately informed about his probability of prevailing, the litigation rate would still be higher under the English Rule. The selection of cases for trial would be reversed, of course: strong plaintiffs will tend to reject offers of settlement and go to trial instead. Polinsky and Rubinfeld (1998) point out that, in this case, the additional cases that go to trial under the English Rule all have lower probabilities of prevailing than those that go to trial under the American Rule. Finally, papers discussing the normative implications of the English Rule in a world of settlement include Hylton (2002), who analyzes a model with strict liability, and Spier (1997) who considers a negligence rule.

3.4.2. *Offer-of-judgment rules*

Another interesting class of fee-shifting rules bases the allocation of costs on the settlement offers that the litigants make to each other prior to trial. The most notable offer-of-judgment rule is Rule 68 of the United States Rules of Civil Procedure. Under this rule, if a plaintiff rejects a settlement offer made by the defendant and later receives a judgment that is less favorable than the offer, then the plaintiff is forced to bear the defendant's post-offer costs. Although Rule 68 is one-sided in the sense that it only applies to offers made by the defendant, other rules, such as California Code of Civil Procedure Rule 998, allow for two-sided cost shifting. Note that these offer-of-judgment rules are similar to the English Rule in the sense that they penalize the "loser" where the loser is defined relative to the settlement offers. Like the English Rule, offer-of-judgment rules can impact the likelihood of settlement; they can also affect the accuracy of settlements.

The litigation rate The effects of offer-of-judgment rules on litigation rates are interesting and subtle.[75] Spier (1994b) considers offer-of-judgment rules in a model where the defendant has private information. If the level of damages is common knowledge but there is disagreement over the probability of winning, then offer-of-judgment rules are very similar in theory to the English Rule. Since the settlement offer will typically

[74] This is a simple extension of Reinganum and Wilde (1986).

[75] Miller (1986) considers an optimism model. He showed that Rule 68 tends to be pro-defendant (which is not surprising since it is a one-sided rule) and has an ambiguous effect on the settlement rate.

lie between zero and the plaintiff's damage, the "loser" will end up picking up the expenses of the "winner." Not surprisingly, offer-of-judgment rules tend to increase the rate of litigation in these cases.[76]

Remarkably, this result may be reversed when liability is acknowledged but there is private information about damages: offer-of-judgment-rules discipline aggressive settlement offers and tend to lower the rate of litigation. To see why this is true, let us suppose as before that the defendant is privately informed about the expected judgment at trial, x.[77] The judgment at trial is noisy, however: $J = x + \varepsilon$ where ε is a noise term drawn from distribution $g(\varepsilon)$ with both mean and median equal to zero. Suppose that the plaintiff may make a single take-it-or-leave-it offer, S^J before trial. This offer will also be binding on the future allocation of costs: the defendant must bear the plaintiff's costs as well as his own if $J > S^J$ and the plaintiff will bear the defendant's costs if $J < S^J$. The defendant accepts the offer of judgment if and only if it is lower than what he would expect to pay at trial, $S^J < x + [1 - G(S^J - x)](c_p + c_d)$. This condition defines an implicit cutoff, $x(S^J)$ where defendants above the cutoff accept the offer and those below reject the offer and go to trial.[78] It is not difficult to extend the earlier analysis to show that the plaintiff's optimal offer satisfies the following first-order condition (Spier, 1994b):

$$1 - F\big(x\big(S^J\big)\big) - (c_p + c_d) f\big(x\big(S^J\big)\big) - (c_p + c_d) \int_{\underline{x}}^{x^J} g\big(S^J - x\big) f(x) dx = 0.$$

The intuition is straightforward. When the plaintiff raises the offer slightly there are both costs and benefits. $1 - F(x(S^J))$ represents the benefit: defendants with damages above $x(S^J)$ pay more in settlement. The next term represents a cost: when the offer rises, more costly trials occur. The third term represents an additional cost that is special to the offer-of-judgment rule: when the settlement offer is raised, it is more likely that the plaintiff will be forced to bear the costs should the case go to trial. It is through this third effect that offer-of-judgment rules discipline the plaintiff from making outrageous offers.

Settlement accuracy Bebchuk and Chang (1999) observe that offer-of-judgment rules can lead to greater accuracy in settlement. Their model is one of complete information—both parties observe the expected judgment at trial, x, and neither knows the realization of the noise term, ε. In their model, unlike the model of Spier (1994b), there are two

[76] Farmer and Pecorino (2000) show that this result does not necessarily hold when the level of damages is random. In this case, the analogy between the offer-of-judgment rule and the English rule no longer applies.

[77] Spier (1994b) considers an extensive form game where the plaintiff is privately informed instead. The identities of the informed and uninformed parties are reversed here to maintain consistency within the chapter. Spier also considers a more general mechanism design problem with two-sided private information and characterizes the fee-shifting rule and the bargaining game that maximizes the settlement rate. The fee-shifting rule has the same "flavor" as an offer-of-judgment rule.

[78] Note also that $x'(S^J) = 1$.

stages of bargaining: in the first stage, one of the two parties may make an offer of judgment, S^J, which is officially registered with the court. In the second stage, they arrive at the Nash bargaining solution (or, equivalently, they flip a coin to see who can make a take-it-or-leave it offer.

As a benchmark, recall that in the absence of fee-shifting, the bargaining range in the second stage would be $[x - c_p, x + c_d]$. With equal bargaining power the parties would settle at the midpoint: $S^* = x + (c_d - c_p)/2$. By backwards induction, this is what they would settle for in the first stage as well. Note that the party with the larger litigation cost is disadvantaged in the bargaining outcome. It is in this sense that the settlement does not accurately reflect the expected judgment at trial.

The outcome changes in a dramatic way with a two-sided offer-of-judgment rule. We will proceed by backwards induction. In the second stage, the most the defendant is willing to pay in settlement is $\overline{S} = x + [1 - G(S^J - x)](c_p + c_d)$ and the least the plaintiff is willing to accept is $\underline{S} = x - G(S^J - x)(c_p + c_d)$. With equal bargaining power, the parties would settle for an amount

$$\sigma\left(S^J\right) = x + \left[\frac{1}{2} - G(S^J - x)\right](c_p + c_d).$$

Working backwards to stage 1, suppose that the defendant must make the "offer of judgment," S^J. The defendant's equilibrium offer is clearly the value that satisfies $S^J = \sigma(S^J)$.[79] The same result would be obtained if the plaintiff rather than the defendant could make the offer of judgment. Using the assumption that the median $g(\varepsilon)$ is zero we have

$$S^J = x.$$

The parties settle for the expected judgment at trial, regardless of who has the higher litigation costs. The offer-of-judgment rule in this instance levels the playing field and generates a settlement outcome that accurately reflects the expected judgment at trial.

3.5. Negative expected value (NEV) claims and "frivolous litigation"

Many scholars and policy makers have expressed concerns about the presence of nuisance suits—that is, "frivolous" cases that are brought by aggressive plaintiffs for the sole purpose of extracting settlement offers from defendants. Despite the broad concern about nuisance suits and the importance of better understanding their sources and their ultimate control, there is little consensus about how a frivolous case should be defined. Some lawsuits, such as those where the plaintiff has suffered no damages or has no legal entitlement to recovery, are certainly frivolous. But not all frivolous cases are associated with a zero expected judgment—juries and judges may make mistakes and may

[79] If $S^J < \sigma(S^J)$ then the plaintiff would clearly reject S^J in the first round and subsequently settle for $\sigma(S^J)$ in the second stage.

grant an award despite the facts and the law. Furthermore, cases with very low expected judgments may, in fact, be socially valuable.

The law and economics literature has, for the most part, side-stepped this definitional problem by focusing instead on negative expected value (NEV) claims. We say that a plaintiff has an NEV claim *if the plaintiff's perceived return at the time of filing from taking the case all the way to trial is negative*. This corresponds to situations where the expected judgment is small when compared with the costs of filing, discovery, and litigation. At first blush, it does not appear that a plaintiff could profitably bring a lawsuit where the expected judgment at trial, x, is smaller than her costs of litigation, c_p. The plaintiff would surely choose to drop the case before trial! However, there are several reasons why NEV claims may in fact have positive settlement value. Also, there are several policy instruments that can affect these NEV claims.

3.5.1. Settlement of NEV claims

Bebchuk (1988) argues that the presence of asymmetric information—and the associated uncertainty—may make the plaintiff's threat to litigate the claim credible. In this model, as in Bebchuk (1984), the plaintiff privately observes the expected judgment at trial, x, filing is costless, and the defendant could make a take-it-or-leave-it offer before a costly trial. NEV cases occur with probability $F(c_p)$: the probability that x is smaller than the trial costs c_p. If $F(c_p)$ is small, the defendant will offer to settle for a positive amount: $S^* > 0$. Here, a plaintiff with a NEV claim benefits from the presence of other plaintiffs with positive expected value (PEV) suits. If $F(c_p)$ is large (so that there are many NEV claims) then the defendant will rationally refrain from settling—he offers $S^* = 0$ and all PEV plaintiffs go to trial.

Katz (1990) assumes that the plaintiff's costs are divided between the filing stage (k) and the trial stage ($c_p - k$) and that the defendant makes his settlement offer *after observing whether the plaintiff filed suit*. If $k < S^*$ then Bebchuk's (1988) result still holds: plaintiffs with NEV claims will file suit in the expectation—correct in equilibrium—that they will receive a favorable offer in the future. The equilibrium changes, however, when $k > S^*$. If the plaintiffs anticipated S^* as before, then no plaintiff would file suit with the intention of settling and, since only plaintiffs with PEV claims would find it worthwhile to file suit, the defendant would rationally make an offer above S^*. With filing costs taken into account, the defendant's offer becomes $S^{**} = k$. Plaintiffs with NEV claims are indifferent between filing suit and not and, in equilibrium, mix between these two options.[80]

Several papers have modeled the settlement of NEV claims in symmetric information environments. As discussed in section 2.2.1, Bebchuk (1996) shows that the divisibility of litigation costs can make the plaintiff's threat to litigate credible. Intuitively, the

[80] Although he considers this continuous case, Katz (1990) focuses on a two type example where the plaintiff is either injured or uninjured. In this example, the defendant's settlement offer is not deterministic as in the text but is a mixed strategy as well.

bulk of the costs have already been sunk once the case reaches the courthouse steps. Although the case may have begun with negative expected value it can be transformed into a positive expected value case when the costs of litigation are spread out over time. Rosenberg and Shavell (1985) show that plaintiffs may profitably extract settlement offers even when it is common knowledge that the plaintiff claims are NEV, and hence will be dropped before trial. Their argument rests not on divisibility but on an important timing assumption: the defendant had to sink some defense costs or risk a default or summary judgment before trial. The plaintiff is able to "hold up" the defendant in this scenario because the defendant is willing to pay *up to his defense costs* to settle the case.

3.5.2. Policy instruments

The ability of plaintiffs to extract settlements is affected in subtle ways by the fee shifting rules discussed earlier. First, suppose that a plaintiff's damages are substantial but the probability of winning is so low that the value of the case is only slightly above zero (it is a marginal PEV case). The English Rule dilutes the value of this case: the plaintiff will almost surely lose at trial and be forced to compensate the defendant for his litigation costs. This confirms the intuition the English rule serves to discourage low-probability-of-prevailing cases. Now suppose instead that the probability of winning is quite high but the damages are so small that the value of the case is slightly negative (it is a marginal NEV case). The English Rule will enhance the value of this case because the plaintiff's litigation costs will be shifted to the defendant at trial. Taken together, we see that the English rule can generate either more marginal cases or fewer of them, depending on the context.[81]

Some commentators have argued that contingent fees encourage frivolous claims and should therefore be prohibited. These arguments seem to be based on the idea that a plaintiff's threat to litigate is higher with contingent fees because it is the lawyer, not the plaintiff, who bears the direct costs of trials. In practice, however, the lawyer typically has more information about the case than the plaintiff and effectively controls the settlement decision. Since the lawyer will receive 33%, say, of the award but bear 100% of the costs, he would have an even greater incentive to avoid trials (Miller, 1987). Similarly, the lawyer would have even less of an incentive to represent an NEV case to begin with (Dana and Spier, 1993).

3.6. Contingent fees

In the United States, it is very common for plaintiffs to compensate their attorneys with contingent fees where the attorney receives a percentage of any settlement or judgment

[81] Bebchuk and Chang (1996) present an analysis of Rule 11 of the Federal Rules of Civil Procedure. In their framework, fee shifting is imposed not for winning or losing *per se*, but rather for deviations from case-specific thresholds for victory. They argue that Rule 11 may perform better at controlling frivolous lawsuits than alternatives.

but receives nothing if the case is lost. The typical contract involves a fixed percentage, often 33%, although there is variation. Some contracts specify that the percentage will decline with the amount of the award or the settlement, while others specify that a smaller percentage is received if the case is settled rather than litigated. See Rubinfeld and Scotchmer (1998) for an excellent discussion. Although contingent fees are most common in personal injury and medical malpractice cases, they appear in other types of litigation as well. It is much rarer to see contingent fees for defense attorneys, although they are occasionally adopted for tax cases. See the discussion in Dana and Spier (1993).

Contingent fees are, however, subject to restrictions and are often the focus of policy debates. In many European countries their use is totally prohibited. Despite their prevalence in the United States, the ABA Model Rules of Professional Conduct prohibit the outright purchase of cases by attorneys. The rationales behind these laws and restrictions tends to fall into one of two categories: (1) a paternalistic notion that unscrupulous attorneys can use contingent fees to take advantage of naïve plaintiffs or (2) the notion that contingent fees will lead to too high a level of litigation.[82]

In contrast to the views of policy makers, economists have tended to view contingent fees as a rational—and privately economically efficient—response to a variety of factors. First, contingent fees provide one way that liquidity-constrained plaintiffs can finance their cases. Without them, many plaintiffs would simply be priced out of the market. Second, contingent fees may provide for better risk sharing between plaintiffs and attorneys. This may be especially true if attorneys are "diversified" in the sense that they are representing many statistically-independent claims. Third, contingent fees help to overcome problems associated with moral hazard and asymmetric information. Fourth, contingent fees may (under some circumstances) serve as a valuable strategic commitment in negotiations. These latter issues are subtle and warrant a more detailed discussion.

3.6.1. Moral hazard

A moral hazard or hidden action problem exists when it is prohibitively costly for a principal (in this instance, the plaintiff) to directly monitor the effort chosen by the agent (here, the attorney). This type of problem is common in litigation since the plaintiff does not have the ability or the expertise to directly observe: (1) how much time the lawyer is spending on the case and (2) how hard the lawyer is working during those hours. Absent concerns for reputation or long-run play, paying the lawyer by the hour or by the job would clearly lead to insufficient attorney effort. Basing the attorney's pay on

[82] Santore and Viard (2001) have an alternative rationale for these limits: limits on the outright sale of claims generates more "rent" for the plaintiffs' attorneys as a group. The idea is this: absent prohibitions, attorneys will compete Bertrand-style and purchase claims for a large up-front fee and then receive 100% of the winnings. This scheme creates the efficient incentives for attorneys to work hard on the cases in the future, but gives them zero profits in an *ex ante* sense. With limits on the up-front transfer, the contingent percentage will fall short of 100%. Importantly, the individual rationality constraint will fail to bind.

the outcome of litigation, on the other hand, gives the attorney an incentive to work harder than he would otherwise. (See Schwartz and Mitchell, 1970 and Danzon, 1983 for early discussion of related issues.)

If attorney effort were the only concern, then an obvious economic solution exists: attorneys should "buy" cases from plaintiffs, paying an upfront fee to the plaintiff in exchange for 100% ownership of the settlement or judgment. Even if this were legal (and it is not), there would be practical reasons why this financial arrangement would not be universal. First, it is sometimes important to provide incentives to plaintiffs as well. If the plaintiffs were paid a lump-sum at the onset, they would have little incentive to cooperate with the lawyer or to put forth effort to maintain credibility on the stand. In short, the attorney and the plaintiff may be viewed as a "team," and sharing the outcome may be the second-best solution to the moral-hazard-in-teams problem. Second, we will see that the presence of *asymmetric information* may make selling the case undesirable for other reasons.

3.6.2. Asymmetric information

Plaintiffs and their attorneys typically have access to different information that is relevant to the case. The plaintiff will have first-hand knowledge of the extent of his or her injuries and the extent of contributory negligence. The attorney, on the other hand, knows more about his or her abilities and expertise in handling the case and knows more about the law that is relevant to the case. Suppose that the plaintiff can pinpoint the probability that the case will win at trial—it is either "high" or "low." The problem that the attorney faces in negotiating the contingent fee contract is analogous to the famous "Market for Lemons" problem: it is efficient for the attorney to "buy" the plaintiff's case (perhaps for risk-sharing or moral-hazard concerns) but he doesn't want to overpay for a "lemon" (i.e., a low-probability-of-winning case). Absent regulations on contingent fees, Rubinfeld and Scotchmer (1993) show that attorneys offer a menu of contracts in equilibrium: one with a high contingent percentage (and a relatively high purchase price) and the other with a low contingent percentage (and a lower purchase price). The client who believes that his case has a high chance of winning self-selects into the latter contract. Intuitively, the client signals the high quality of his case by his willingness to accept greater ownership of the case outcome. Rubinfeld and Scotchmer also characterize the equilibrium menu of contracts when the attorney has private information: the attorney signals his high quality through his willingness to accept contingent payment.

Dana and Spier (1993) show that contingent fees are the privately optimal financial arrangement when the attorney has better information than the plaintiff about the merits of the case. Intuitively, if the attorney were paid by the hour he would have little incentive to reveal to the client if the case lacked merit, and would pursue even claims with negative expected value.[83] Through the contingent fee, the plaintiff can be assured that

[83] Dana and Spier's (1993) model features equilibrium wages above the lawyers' opportunity cost of time. These attorney rents may result from fixed costs associated with overhead, case management, or perhaps

the attorney will make a more appropriate decision, pursuing *only cases that are more likely to win at trial*. This is an important insight and may be contrasted with the popular wisdom: in Dana and Spier (1993) contingent fees may lead to less (rather than more) litigation than otherwise.[84]

3.6.3. Settlement externalities

Using the same notation as in section 2, if the case goes to trial, the defendant will lose $x + c_d$ and the plaintiff and his attorney will (jointly) gain $x - c_p$.[85] As before, let x represent the plaintiff's expected judgment at trial and let c_p and c_d be litigation costs for the two sides. Absent agency issues, the settlement range would simply be $[\underline{S}, \bar{S}] = [x - c_p, x + c_d]$. The lower bound of the settlement range will change, however, depending upon the contingent fee received by the plaintiff's attorney, the allocation of costs, and who retains control over the settlement decision.

Formally, let α represent the attorney's fractional share of the joint cost (c_p) and let θ be the attorney's share of the judgment or settlement. If the case goes to trial, the lawyer's payoff will be $\theta x - \alpha c_p$ while the plaintiff's payoff will be $(1-\theta)x - (1-\alpha)c_p$. If the case settles out of court for S, on the other hand, the attorney's payoff is θS and the plaintiff's payoff is $(1-\theta)S$. The settlement offer that makes the attorney indifferent between settling out of court and going to trial is

$$\underline{S}^A = x - \left(\frac{\alpha}{\theta}\right)c_p,$$

while the offer that makes the plaintiff indifferent is

$$\underline{S}^P = x - \left(\frac{1-\alpha}{1-\theta}\right)c_p.$$

Clearly the lower bound on the settlement range depends crucially upon α, θ, and who retains control over the settlement decision.

First, suppose the proportion of the winnings received by the attorney is exactly equal to his share of the litigation costs, $\alpha = \theta$, then there is no conflict of interest between the plaintiff and his attorney: $\underline{S}^A = \underline{S}^P = \underline{S} = x - c_p$. The attorney will, for example, act in their joint interest when making settlement demands and responding to settlement offers. (This so-called "no-conflict" system, proposed by Polinsky and Rubinfeld,

education. Hourly fees would align the interests of lawyers and clients if they could be set to exactly equal the attorney's opportunity cost of time (see Emons, 2000). In that case, the lawyer would be indifferent between pursuing a case and not and there is an equilibrium where the lawyer makes the right decision on behalf of the client. If hourly fees could be fine-tuned in this way, they would be preferred to contingent fees.

[84] See also Miceli (1994). These theoretical results are consistent with the empirical evidence in Helland and Tabarrok (2003), who show that contingent fees are associated with higher-quality cases and a faster case resolution, and Danzon and Lillard (1983), who show a higher drop rate with contingent fees.

[85] We will abstract from the agency problem between the defendant and the defense attorney and assume that they always behave in their joint interest.

2003, also overcomes other agency problems, such as the effort problem identified in the discussion of moral hazard.)

More realistic, perhaps, is the case where $\alpha > \theta$. This would occur if the lawyer bears a disproportionate amount of the trial costs—$\alpha = 1$, perhaps—but only receives a third of the winnings—$\theta = 1/3$. It is easy to see in this case that the attorney's and the plaintiff's preferences diverge: $\underline{S}^A = x - 3c_p < \underline{S}$ and $\underline{S}^P = x > \underline{S}$. Their joint bargaining position is compromised if the attorney retains control over the settlement decision—the attorney is sorely tempted to settle for *too little*.[86, 87] (See Miller, 1987.) Giving the plaintiff control, on the other hand, enhances the bargaining outcome. Since the costs are "externalized" on the attorney, the plaintiff can credibly threaten to go to trial unless he receives an offer of at least $S = x$. (See Bebchuk and Guzman, 1996.) Put somewhat differently, contingent fees can sometimes serve as a bargaining tool in settlement negotiations.[88]

Hay (1997) analyzes a modified fee structure where the contingent fee associated with a settlement, θ_s, could differ from the fee associated with a judgment at trial, θ_t. In practice, it is not uncommon for the attorney to receive a greater percentage at trial: $\theta_t > \theta_s$. This fact is consistent with the idea that contingent fees are designed to both mitigate agency problems and to extract surplus during negotiations. The interested reader can easily verify that the lower bound of the settlement range, assuming the attorney retains control, is:

$$\underline{S}^A = \left(\frac{\theta_t}{\theta_s}\right)x - \left(\frac{\alpha}{\theta_s}\right)c_p.$$

If the plaintiff retains control, on the other hand, the lower bound is:

$$\underline{S}^P = \left(\frac{1-\theta_t}{1-\theta_s}\right)x - \left(\frac{1-\alpha}{1-\theta_s}\right)c_p.$$

If the plaintiff retains control over settlement decisions and has a great deal of bargaining power, then, Hay shows, the plaintiff will create a contract with a high contingent fee for trial winnings but low contingent fee for settlements. To take an extreme illustrative example, $\theta_t = 1$ creates a very powerful incentive for the attorney to work hard at trial. This raises the expected judgment at trial, and "scares" the defendant into accepting a high settlement offer. The plaintiff can benefit from this by setting a lower contingent fee for settlement. In the extreme, $\theta_s = 0$ allows the plaintiff to keep all of the money for himself! Interestingly, the same pattern emerges (although in modified form) when the attorney has all of the power at trial. θ_s cannot be too small, however, for if it were, then the settlement range would disappear.

[86] This discussion is assuming, of course, that the contingent fee contract is not renegotiated and that side payments between the attorney and client during settlement are impossible. In a frictionless world, they would certainly accept an offer of $S = x - c_p$.

[87] Thomason (1991) shows that plaintiffs who represent themselves in litigation have both higher settlement rates and higher settlement amounts, consistent with the theory.

[88] See also Choi (2003). Rickman (1999) who extends the dynamic asymmetric information model of Spier (1992a) to include the decisions of contingent-fee attorneys.

3.7. Tribunals

3.7.1. Judges and juries

Condorcet jury theorem Suppose there is a single defendant who is either innocent or guilty, and N jurors, each of whom privately receives an informative (but imperfect) signal about the defendant's guilt. Furthermore, suppose that the jurors have the same preferences: they would all agree on conviction versus acquittal if they each had access to all of the signals. More than 200 years ago, Condorcet (1785) proved that majority voting implements the jury's preferred outcome in the limit as N approaches infinity. An important underlying assumption in this analysis is that the jurors vote "sincerely" or "non-strategically"—they behave as if their vote, and only their vote, matters for the final outcome. This assumption has been maintained in most of the literature that followed.

In recent years, scholars have extended the analysis of jury decision making to include strategic behavior. To see why jurors would vote strategically, consider the following example, based on Austen-Smith and Banks (1996). There are three jurors who share a common prior that a defendant is in all likelihood innocent and would—absent additional evidence—unanimously vote to acquit. However, the jurors would change their minds and prefer to convict the defendant if (and only if) all three private signals indicate guilt. Suppose the first two jurors vote sincerely (a la Condorcet), entering a vote for acquittal when the signal is "innocent" and a vote for conviction when the signal is "guilty." The third juror, being rational and strategic, will not vote sincerely. With a majority rule, this third juror's vote is pivotal only when the votes of the first two jurors conflict. Therefore the third juror will vote to acquit even when his signal indicates guilt.

Strategic behavior among jurors can lead to unintended outcomes. Take, for example, the requirement in criminal courts that a defendant can be convicted only when the jurors are unanimous in their opinions. This rule is commonly thought to serve to reduce the probability of a type I error: the erroneous conviction of an innocent man. Although this is certainly true with sincere (non-strategic) voting, Feddersen and Pesendorfer (1998) have shown that the unanimity requirement can actually lead to more type I errors than a majority rule. Furthermore, in contrast to the limiting result in the Condorcet jury theorem, the probability of a type I error may actually increase with N, the number of jurors. Larger juries can also create a free-rider problem where jurors shirk in their individual responsibilities to pay attention and process information, reducing the accuracy of the ultimate decision (Mukhopadhaya, 2003).

Empirical work Casual observers tend to argue that judges may be preferred to juries because (1) juries are not capable of evaluating complex cases and (2) juries tend to be pro-plaintiff, granting inappropriately large awards for such things as punitive damages and pain and suffering. These premises have not been completely borne out in practice. Kalven and Zeisel (1966) survey judges who presided over civil jury trials and showed that the judges would have come to exactly the same decision as the jury in 78% of the

cases. Importantly, the divergence was unbiased: the probability that the judge would have chosen to find a defendant liable when the jury found him not liable was the same as the chance that the judge would have found the defendant liable when the jury declared him not so. In an empirical analysis of awards, win rates, and settlements, Helland and Tabarrok (2000) find some confirmation of jury bias: conditional upon a finding of liability in their overall sample, a judge's award was on average only 31% of a jury award. They did find, however, that the plaintiff win rate was lower with a jury trial, and that controlling for a variety of case characteristics and selection bias got rid of most (but not all) of the jury bias.[89]

3.7.2. Adversarial versus inquisitorial systems

Scholars have found it useful to distinguish between *adversarial legal systems*, such as the one found in the United States, and *inquisitorial legal systems*, such as those found in continental Europe, Japan, and most other non-English speaking countries. In adversarial systems, the two sides in the dispute (or their representative agents) gather and process information. The two sides are self-interested and selective in what they reveal to each other and to the court.[90] In an inquisitorial system, the gathering and processing of evidence is centralized and often presided over by a judge.[91] Ideally, the inquisitor would be unbiased and hardworking, striving to uncover the truth. This distinction is, of course, a caricature—most legal regimes have elements of both.[92] Nevertheless, this stark distinction is useful for identifying the relevant costs and benefits of the different systems.

At first blush, inquisitorial systems have some obvious advantages over adversarial systems. In an adversarial process, the two sides gather evidence—evidence that is both helpful and harmful to their own position—and then choose to present the helpful information and discard the harmful information. Inquisitorial systems, by virtue of being centralized, can avoid the duplication (and wasted) efforts inherent in the adversarial process. Inquisitorial systems also appear to lead to more accurate trial outcomes: inquisitors, being disinterested parties, have no incentive to hide relevant information or mislead the decision maker.[93] Closer inspection, however, reveals that these critiques

[89] See also Clermont and Eisenberg (1992).

[90] The fact that adversarial systems such as that in the United States condone the hiding of damning information may appear puzzling at first. Why not require the two sides to reveal all of the information that they find? The reason may be quite simple: a rule that requires both sides to reveal all information would require a very sophisticated system of enforcement. In particular, there would have to be sanctions for not revealing all relevant information. In practice, it would be very difficult or impossible for the court to accurately identify what evidence should have been revealed. The errors in enforcement would likely have a chilling effect on the primary economic activity. See the discussion in Shavell (1989a).

[91] According to Posner (2003, table 1.1) the ratio of lawyers to judges in the United States is 55 to 1 while in Germany it is 7 to 1.

[92] See the discussion in Parisi (2002).

[93] See the informal arguments of Tullock (1980).

may be overstated and that there may be some distinct advantages of the adversarial system.

Milgrom and Roberts (1986) present a persuasion game where the parties can strategically withhold evidence at trial. The two parties to the lawsuit have equal access to all of the relevant evidence and can costlessly and credibly disclose it at trial. They find that the full information decision arises in equilibrium, even when the court is uninformed and strategically unsophisticated. The intuition is straightforward: since any piece of evidence will favor one party or the other surely, one party will have the private incentive to reveal it. In equilibrium, no stone is left unturned. In sum, accuracy is not necessarily compromised by the private incentives to withhold evidence.[94]

This stark result may no longer hold when parties have asymmetric access to evidence or when evidence is costly to gather and disclose. Daughety and Reinganum (2000b) present a model where the two parties independently engage in a sequential search for evidence. Although the court's decision depends on the information presented at trial, the court is not modeled as a purely Bayesian player. Instead, Daughety and Reinganum consider decision rules that satisfy axiomatic principles. Equilibrium bias may arise for several reasons in this setting. First, the two parties may have asymmetric sampling costs or access to very different sets of information. Furthermore, the parties may have very different stakes in the future. If the case is ultimately appealed, for example, then the costs of staging an appeal drive a wedge between what the two parties will win, creating a pro-defendant bias in the earlier stages.[95]

Dewatripont and Tirole (1999) argue that adversarial systems are more efficient than inquisitorial systems in providing incentives for information gathering. Suppose that two parties, A and B, are engaged in a dispute. An impartial and socially responsible judge must rule in one of three possible ways: finding for party A, finding for party B, and an intermediate outcome. Many types of disputes may fit this rough characterization. Suppose that A has been injured in an accident with B and that the doctrine of comparative negligence applies. If B was negligent but A took appropriate care then the judge would place full liability on B. On the other hand, if party A was negligent as well, then the judge may split the liability between the two parties. Custody battles between two parents provide another example: the judge can either award sole custody to one parent (excluding the other), or can award joint custody where the parents share the child.

[94] Shin (1998) argues that the adversarial system may be even more accurate. In his model, the two parties each receive an independent signal of the evidence, while the inquisitor receives one signal. The adversarial system may have an advantage in that the court may be able to glean better information in the event that nothing about the state of nature is directly revealed at trial.

[95] Froeb and Kobayashi (1996) argue that the full information decision may be obtained in the presence of costly information gathering even if the decision maker is fundamentally biased in favor of one of the parties. Intuitively, the favored party has less of an incentive to invest in information gathering and "free rides" on the court's bias. Despite being given an head start, the favored party may well use the same stopping rule as before, leading to exactly the same decision as would be made by an unbiased court.

Dewatripont and Tirole compare two regimes. In the inquisitorial system, an impartial (but possibly lazy) "inquisitor" is responsible for gathering evidence for both sides. In the adversarial system, partial advocates for each side are responsible for gathering the evidence. There may be evidence supporting each side of the case, but this evidence must be gathered before trial. (In the comparative negligence scenario, the injured party may be able to find proof of his or her care level. In the child custody scenario, a parent may be able to find proof of his or her suitability as a guardian.) In particular, if the gatherer of the information spends K there is a probability "x" that favorable evidence will be found. If only evidence favorable to A comes to light, the judge will find for A. Similarly, if the only evidence offered is favorable to B, then the judge will find for B. If both types of evidence are presented, then the judge will implement the intermediate outcome.

A crucial assumption in Dewatripont and Tirole's paper is that incentive schemes can be conditioned only upon the final judgment at trial and not upon the inquisitor or advocate's efforts or the evidence, directly. Recall that the intermediate judgment can arise because either: (1) the inquisitor was lazy and did not put in any effort into gathering evidence or (2) the inquisitor put in the effort but was unsuccessful at finding the evidence. The inability of the reward scheme to distinguish between these two alternatives dampens the incentives of the inquisitor. The adversarial system, on the other hand, provides better incentives for the advocates to find evidence.

The advantages of the adversarial system are easily seen in the following example where it is assumed that the agents cannot manipulate or distort evidence once it is gathered. We will assume that the gatherer of information receives a reward, W, only if the court finds exclusively for one of the two parties. Assuming that the advocate for party B will put in effort K to find evidence, party A will put in effort when:

$$x(1-x)W \geq K.$$

That is, when the probability that Advocate A finds evidence, x, multiplied by the probability that Advocate B does not find evidence, $1-x$, multiplied by the reward is greater than the cost of gathering evidence. Indeed, when $W > K/x(1-x)$ it is a dominant strategy for both advocates to gather evidence and the first best outcome is obtained.

Suppose instead that the evidence is gathered by a single inquisitor. Two incentive compatibility constraints are relevant here. The first is that the inquisitor prefers to gather evidence for both A and B rather than gathering no evidence at all. By analogy to the advocate case above, this constraint may be written:

$$2x(1-x)W \geq 2K.$$

The second constraint is that the inquisitor must prefer to gather evidence for both A and B (a payoff of $2x(1-x)W-2K$) rather than for just one party (a payoff of $xW-K$):

$$x(1-2x)W \geq K.$$

This second constraint is harder to satisfy than the first. Here is the intuition. If the inquisitor were to succeed in gathering evidence for party A, then any further effort on

behalf of party B would be counterproductive. With evidence supporting both parties, the judge would rule for the "intermediate outcome" and the inquisitor would lose his reward, W.

It is not difficult to see that when $x \geq 1/2$ there does not exist a wage, W, that satisfies incentive compatibility for the inquisitor. In this case, the adversarial system is more efficient because it generates more information at trial. When $x < 1/2$, on the other hand, then it becomes necessary to pay rents to the inquisitor to get him to gather information favorable to both sides. Here, the adversarial system is more efficient because it generates the same information as the inquisitorial system but at lower cost to society. Dewatripont and Tirole confirm these same intuitions when the parties can conceal information from the court. Interestingly, when $x \geq 1/2$, then allowing the inquisitor to hide information (in order to avoid an intermediate ruling) creates better incentives—albeit at the expense of generating biased outcomes at trial.

The empirical work on this topic is scant. As part of a series comparing common- and civil-law countries, La Porta et al. (1997, 1998) provide evidence that common-law countries are better at protecting the rights of investors than civil-law countries. For example, they find that common-law countries had vastly more IPOs (initial public offerings) than those in the civil law tradition. The number of IPOs per million people was 2.2 for common-law countries, compared with .02 for countries in the French civil law tradition and .12 in the German tradition. Similarly, common-law countries had a much higher proportion of outsider-held stock as a fraction of GNP and many more firms per capita than their civil-law counterparts.[96]

3.7.3. Alternative dispute resolution (ADR)

Many cases are resolved outside of the legal arena through alternative dispute resolution (ADR). ADR is a very general term and encompasses both formal and informal proceedings that help parties resolve their disputes *outside* of formal litigation. Unlike settlement, which is typically achieved by the litigants themselves (and their agents—the lawyers—who represent them), ADR proceedings typically make use of third parties, including arbitrators, who offer opinions and/or advice.[97] ADR proceedings reflect two primary goals: (1) to reduce the transactions costs of reaching an agreement (these costs could include both the legal costs as well as the non-pecuniary costs to the litigants themselves) and (2) to come to "better" outcomes. Mnookin (1998) presents an excellent brief survey of ADR in practice and in theory. Although they tend to share common goals, ADR proceedings take on a wide variety of forms, each of which raises its own economic issues.

[96] Concentrated ownership and insider ownership may be good substitutes for legal protection.

[97] Mediation, where a third party may speak confidentially with the two sides and help them to find creative solutions to their dispute, is another common form of ADR. Since there has been little formal economic analysis of mediation, the topic has been downplayed here. See Ayres and Nalebuff (1997) for an economic model of mediation.

Some ADR systems are specified and designed by the parties themselves at either an *ex ante* or an *ex post* stage. Labor contracts, such as agreements between unions and employers, and commercial contracts, such as those between the wholesalers, dealers, and brokers of rough diamonds, include dispute resolution mechanisms to be used in the future if disputes arise.[98] Common features of these contracts include the ability of the parties to choose the arbitrators (the "third party") themselves and the binding nature of the third party's decision. Even if the parties did not specify the mechanism in the contract *ex ante* (perhaps because of drafting costs), they may still decide to adopt ADR *ex post* (after the dispute has arisen). Absent externalities and other market imperfections, the economist's view of *ex ante* agreements is that they are in the interest of society more broadly. Efficiency is served by dispute resolution mechanisms that reduce the *ex post* costs of disputes and facilitate economic activity *ex ante*.[99] This is not necessarily true for mechanisms that are chosen *ex post*, since *ex post* agreements would not naturally reflect the *ex ante* efficiency concerns. See Shavell (1994).

There are also many ADR systems that are adopted by the government instead of by the private parties. Many state courts in the United States, for example, have *mandatory* pretrial arbitration for automobile injury and medical malpractice cases. As with the private systems mentioned above, these court-annexed systems often allow the private parties to have some discretion in choosing the arbitrators and are less formal than trials.[100] Unlike the private systems, however, the court-annexed systems are typically non-binding: either party is free to reject the panel's decision and proceed to trial instead. Interestingly, some systems include fee-shifting provisions: if the litigant who rejected the arbitrator's decision receives a less favorable outcome at trial, then he or she will have to bear the opponent's post-arbitration legal costs. (See, for example, Farber and White, 1991.)

The empirical work on ADR has produced interesting and mixed results. Farber and White's (1991) study of medical malpractice claims suggests that pre-trial arbitration is informative to the private parties and that many cases settle subsequent to the arbitrator's decision. Yoon (2004) presents a time series difference-in-differences analysis of medical malpractice claims in Nevada and finds that fewer cases go to trial after the adoption of court-annexed ADR. This encouraging finding is dampened by Yoon's additional finding that ADR neither reduces the litigation costs nor shortens the delay to agreement. The relative dearth of research—both theoretical and empirical—makes ADR a ripe topic for further investigation.

[98] See the excellent case study and informal analysis of the New York Diamond Dealers Club in Bernstein (1992). In addition to specifying their own set of ADR procedures, the diamond industry also "opts out" of New York's contract law by specifying their own rules and codes.

[99] Dixit (2003) presents a model where an arbitrator—and expert—observes additional verifiable signals that are not observed by a court. The contracting parties benefit from this because they are able to write more complete contracts *ex ante*.

[100] See Bernstein (1993). Wittman (2003) uses California automobile cases to support his theory that litigants will choose arbitrators whose decision will reflect—and predict—the future outcome at trial. He finds little difference in the outcomes of cases that are arbitrated versus those that go to trial.

Finally, there is a small literature considering "final-offer arbitration" (FOA) where the arbitrator is bound to choose between final offers that are submitted by each side. FOA is particularly common in employer-union and baseball contract disputes. Farber (1980) characterizes the equilibrium strategies employed by the two sides when they share common uncertainty about the arbitrator's preferred outcome, exogenously given by $z \sim F(z)$. Given a final decision, x, the arbitrator's loss function is quadratic and given by $-(x - z)^2$. (If unconstrained, Farber's arbitrator would simply choose to implement his preferred outcome, $x = z$.) The players face a tradeoff when making their final offers: by making a more aggressive bid, the probability of acceptance is reduced (a cost) but the payoff conditional on the offer being chosen is higher (a benefit). In equilibrium, the plaintiff makes a higher offer than the defendant, but the average of the two offers equals the expected value of z.

Gibbons (1988) extended Farber's insights to include equilibrium learning by the arbitrator. In his normal learning model, there are two noisy signals of the underlying state of the world, z: one signal, s_A, is observed by the arbitrator and the other signal, s_P, is commonly observed by the disputants. As in Farber (1980), there is a separating equilibrium where the parties' final offers are *centered* on their private signal. The arbitrator consequently "learns" the disputants' signal from their offers and, together with his own signal, updates his beliefs about z.[101] Although not explicitly discussed by Gibbons, this model has an interesting implication for the accuracy of arbitration. In particular, although final offer arbitration allows the arbitrator to perfectly learn the signal s_P, it leads to an inefficient *ex post* decision since the arbitrator is bound to choose one of the final offers. With "conventional arbitration," however, the arbitrator is unconstrained. Gibbons shows that, as in final offer arbitration, there is a fully separating equilibrium where the arbitrator perfectly learns the parties' signal, s_P, and subsequently chooses the most efficient outcome. This, of course, begs the question of why we observe final offer arbitration to begin with.

3.8. Multiparty litigation

3.8.1. Class actions

When an injurer has harmed a group of victims, these victims may, under some circumstances, join their claims for the purpose of litigation and/or settlement. Consolidation has some obvious merits: the plaintiffs, the defendant, and the court may benefit from scale economies associated with common proceedings and legal representation (see Miller, 1998). This will obviously affect the private litigation incentives: plaintiffs who would otherwise have not pursued their claims are now able to do so, increasing the volume of litigation. The nature of settlement negotiations will change as well. A plaintiff

[101] See Farmer and Pecorino (2003) for a model that integrates FOA with settlement incentives. Ashenfelter et al. (1992) present experimental results comparing different arbitration systems and give other empirical references. See also Ashenfelter and Bloom (1984).

who joins a class will sacrifice, to a greater or lesser extent, her individuality. This manifests itself in a variety of ways. First, the plaintiff typically loses direct control over the attorney (who is now an agent for multiple parties) creating agency problems. Second, the outcome of a plaintiff's case typically rides on aggregate class characteristics or the characteristics of a "representative plaintiff" rather than her individual characteristics. Each of these issues will be discussed in turn.

Settling for coupons Many lawsuits over the years have settled for coupons as opposed to cash. This most commonly happens in antitrust suits, such as the price-fixing case against the airlines. There, consumers who could prove that they had traveled in the collusive period received (typically, non-tradable) coupons that could be applied to future flights. A main concern with these settlements is that relatively few of the coupons are ultimately redeemed in many of these suits. Consequently, the plaintiff class typically values the coupons at less than their face value. The lawyers representing the class, on the other hand, will often receive compensation that is based on the face value of the coupons. The coupons settlements could reflect an agency problem between the plaintiffs and the lawyer who is representing them. Since the class-action plaintiffs are not present at the bargaining table, the deal struck in settlement will naturally be susceptible to corruption.

In addition to the agency problems identified above, coupon settlements suffer from two additional flaws. First, as emphasized in Borenstein (1996) in a model with imperfect competition, the coupon settlement will typically affect the future pricing behavior of the defendants. Borenstein's main point can be illustrated in an example with undifferentiated Bertrand-style price competition. In the absence of coupons, the competitive price would settle at marginal cost and the defendants would make zero profits. If all consumers receive an abundant supply of coupons, the competitive price would be the marginal cost of production *plus the face value of a coupon*. At this price, firms earn zero profits. Note that the consumers are no better off with coupons than without: the competitive pricing has completely neutralized the effect of the coupons. With heterogeneous consumers, where some receive non-transferable coupons and others do not, we still see that consumers as a class do not benefit. The price of the product would rise to reflect that the coupon will be redeemed. Here, consumers who own the coupons will be better off than before while consumers who did not receive coupons will be worse off.

Polinsky and Rubinfeld (in preparation) zero in on an important welfare distortion arising from coupon settlements. Consumers are heterogeneous in their model, with stochastic per period demand that is uncorrelated across periods. Through a coupon settlement, consumers receive coupons in proportion to their earlier purchases during the so-called "injury period." In this context, several situations can arise. First suppose a consumer who had low demand in the injury period (and hence a small number of coupons) ends up with high demand later. This consumer will use all of his or her coupons and will also purchase additional tickets at full price: the consumer's purchase level does not hinge on the coupon's value. Suppose instead that a consumer who had

high demand in the injury period (and hence has many coupons) has low demand subsequently. For this consumer, the coupons will increase his purchase level since the marginal unit is now cheaper. A fundamental distortion can arise in this case: the quantity that this consumer purchases may be inefficiently high in the sense that the marginal cost of the unit is higher than the intrinsic value that the consumer derives from its use.[102]

Settling for taxes In 1997, "The Tobacco Resolution" settled the lawsuits brought against big tobacco companies by the states (the main plaintiffs) for an expected $13 Billion per year in future tax revenues on cigarettes. The $13 Billion in tax revenues would not just come out of Big Tobacco's pockets, however: basic economic theory tells us that the market price of cigarettes would rise to reflect the tax and the burden would be borne (at least in part) by consumers. In the extreme, if the tobacco industry were perfectly competitive, then the entire burden would be borne by consumers. Why would the states and the tobacco companies agree to this settlement? As argued by Bulow and Klemperer (1998), Tobacco Resolution served the interests of the parties controlling the litigation: the state coffers were enhanced, the lawyers received a contingent fee tied to the tax revenues, and the tobacco companies effectively received license to collude and raise their prices. Bulow and Klemperer argue that, at the end of the day, only about $1 Billion per year would actually be borne by the defendants in that lawsuit. Since the big tobacco companies are able to externalize liability on consumers themselves, the deterrent benefit of liability is compromised.

3.8.2. Private incentives to consolidate

Damage averaging Che (1996) assumes that plaintiffs who join a class enjoy economies of scale from consolidation: the per-plaintiff cost of pursuing litigation decreases in the number of plaintiffs who join the class. This may be due, at least in part, to the streamlined proceedings. Instead of scrutinizing the damages of each individual plaintiff, the court may make a judgment based on the group average or, equivalently, on the damages of a randomly chosen plaintiff representative. Instead of receiving an award that is fine-tuned to his individual characteristics, a plaintiff's award reflects the average damages of the entire class.

Absent settlement, it is clear that plaintiffs with weak cases are more likely to join a class. A weak plaintiff has a stronger incentive to forego an individual hearing in order to receive an average judgment instead. Che shows that this adverse selection problem is mitigated when plaintiffs are privately informed about their damages. In short, weak

[102] A related distortion is also present in Borenstein's (1996) model, although the author does not highlight it. Borenstein also has heterogeneous consumers: some with coupons and others without. The consumers with the coupons may be over-consuming the product. This should be more likely when: (1) the products are less differentiated (so prices are lower), (2) when the proportion of consumers with coupons is relatively small, and (3) when the face value of the coupons is relatively high.

plaintiffs have an incentive to remain independent, too, in an attempt to "signal" that they have strong cases and, in equilibrium, fewer weak plaintiffs join the class. Although the ability to join class actions is shown to reduce the overall litigation spending it does not necessarily create a Pareto improvement.

Information rent extraction The presence of asymmetric information between the defendant and the plaintiff class, and among the plaintiffs themselves, implies that class formation can play a valuable strategic role. Che (2002) argues that classes may form to increase the members' bargaining power via information aggregation. In his model, the author assumes that the plaintiffs jointly evaluate a settlement offer and accept only if each and every member can be made better off than by going to trial. The commitment to a joint decision changes the defendant's choice of settlement offer. Instead of thinking about the distribution of an individual plaintiff's type, the defendant instead considers the distribution of the total damages (the sum of the individual damages). The defendant may subsequently choose to make more generous offers to the class as a whole than when bargaining one-on-one with individuals.[103] In this model, the plaintiffs also have an incentive to exaggerate their types among themselves in order to capture a greater share of the class settlement. This incentive to exaggerate commits the class to be even tougher and induces the defendant to make more generous settlement offers.

3.8.3. Joint and several liability

Just as it is common for a single injurer to harm many victims through his actions, there are many cases where a single victim is harmed by the actions of many injurers. The issue of how to allocate responsibility among multiple injurers has been a challenge for policy makers and scholars alike. Common rules include non-joint liability, where each losing defendant is held responsible only for his own share of the damages, and joint and several liability, where a single losing defendant can be held responsible for the entire level of the plaintiff's damages. While some proponents of joint and several liability have argued that the rule is good for public policy, the economic effects—especially those on settlement outcomes—are quite subtle.[104]

Kornhauser and Revesz (1994a, 1994b) show that settlement effects hinge on several factors including: (1) the treatment of prior settlements when determining the liability of a non-settling defendant and (2) the degree of correlation between the defendant's cases. They analyze the so-called unconditional *pro tanto* setoff rule.[105] With this rule,

[103] Suppose, for example, that there are two plaintiffs and each plaintiff's type (the expected return at trial) is either \$0 or \$1 with equal probability. Taken together, the two plaintiffs have damages \$0 with probability 1/4, \$1 with probability 1/2, and \$2 with probability 1/4. A defendant who would have otherwise offered \$0 to each plaintiff individually may now find it in his interest to offer \$1 to the group.

[104] The survey of Kornhauser and Revesz (1998) also reviews the papers that abstract from the litigation process. How does joint and several liability perform in terms of deterrence? Other issues are addressed include the ability of defendants to collect from each other after an adverse judgment and the role of insolvency.

[105] Klerman (1996) shows that the results are actually sensitive to the formulation of the rule and compares some sensible alternatives.

the liability of a non-settling defendant is reduced, dollar for dollar, by the value of the previous settlements. Suppose, for example, that the plaintiff's damages are \$80 and there are two defendants. If the first defendant settles for S, the second defendant's liability is capped at $\$80 - S$. Kornhauser and Revesz show that joint and several liability encourages settlement when the two cases are positively correlated but discourages settlement when the two cases are independent.

We can see this in a concrete example. Let the unconditional probability of prevailing against each defendant be 50%. Suppose that the two cases are perfectly correlated: the two defendants will either lose together or win together at trial. Ignoring litigation costs, if both cases go to trial at the same time then the expected payment of each defendant is \$20. (They are held liable half the time and split the \$80 between them.) Now let's think about a settlement game. Suppose each defendant is presented with an offer to settle for $S = \$20$. If the first defendant accepts the offer then the second defendant's liability has changed: under the *pro tanto* setoff rule, the second defendant's liability is capped at $\$80 - \$20 = \$60$, which implies an expected judgment of \$30. The settlement decision of the first defendant has imposed a *negative externality* on the second. The plaintiff may be able to take advantage of this externality and induce the two defendants to settle out of court for more than \$20 apiece. The negative externality implies an additional incentive for settlement.[106]

Now suppose that the two cases are independent. If both cases go to trial, each defendant faces an expected judgment of \$30.[107] Putting aside the costs of litigation, suppose that the two defendants are presented with offers to settle for $S = \$30$. If the first defendant accepts the offer, then the second defendant's liability is capped at $\$80 - \$30 = \$50$, which implies an expected judgment of \$25. The settlement decision of the first defendant has imposed a *positive externality* on the second defendant. Note that the second defendant would refuse to settle on these terms and would demand (and receive) a discount. Kornhauser and Revesz show that, for low levels of correlation, the plaintiff will either reduce the settlement offers (if litigation costs are high) or forego settlement altogether and take both defendants to trial (if litigation costs are low).

This interesting theory has found some empirical support. Chang and Sigman (2000) test the results on settlement date for disputes between the Environmental Protection Agency (EPA) and Superfund defendants, the generators and transporters of hazardous waste and the owners of waste sites. They find that joint and several liability tends to promote settlement, and that the results are stronger for cases that are more highly correlated.

[106] Spier (1994a) argues that this may have bad normative implications as it may lead the defendant to overinvest in precautions *ex ante* and may encourage frivolous claims.

[107] A defendant loses alone and pays \$80 with probability $1/4$ and shares responsibility and pays \$40 with probability $1/4$.

3.8.4. Most-favored-nation clauses

Many settlement contracts in litigation involving multiple plaintiffs (or multiple defendants) include "most-favored-nation" (MFN) clauses. They work in the following way: if an early settlement agreement includes an MFN clause and the defendant settles later with another plaintiff for more money, the early settlers receive these terms, too.[108] It is important to emphasize that *these clauses typically apply to settlement payments only and not to judgments at trial*. This feature raises legitimate policy concerns. In the words of one critic, "Because [defendants] are 'straight-jacketed' by the most-favored-nations agreements with certain prior settling [plaintiffs], the strong public policies favoring complete settlement are being frustrated."[109]

Spier (2003a) argues that MFN clauses economize on delay costs when a single defendant makes repeated offers to plaintiffs who have private information about their prospects at trial. Without MFNs, recall that settlement negotiations resulted in cases settling on the courthouse steps (Spier, 1992a). These "11th hour" plaintiffs reject the defendant's early offers because they anticipate, correctly, that the defendant's offers could only improve with time. Through an MFN clause, the defendant induces these late-settling plaintiffs to accept early settlement offers instead.[110] While early settlement is socially desirable, there can sometimes be undesirable side effects of MFNs. In particular, the defendant may choose a more aggressive settlement strategy where more cases end up going to trial than before.

MFNs may also be used as an effective bargaining tool when future plaintiffs have the power to make offers as well. Intuitively, an MFN commits the defendant to be tough in future negotiations, placing an upper bound on what a future plaintiff can extract in settlement. The MFN allows the defendant and the early plaintiffs to capture a greater share of the future bargaining surplus. When committing to the MFN *ex ante*, the defendant and the early plaintiffs do not fully internalize the *ex post* cost of breakdowns, since at least part of that cost will be borne by the future plaintiffs. See Spier (2003b). In a model where early plaintiffs are also privately informed and signal their types through their settlement offers, Daughety and Reinganum (2004) show that MFNs make early settlement negotiations more efficient.[111] Taken together, the welfare effects of most-favored-nation clauses are ambiguous.

[108] These clauses have been included in prominent class action settlements (e.g., the 1999 Vitamins Antitrust case) as well as in settlement agreements with individual litigants (e.g., the tobacco settlements with the states of Florida, Mississippi, and Texas).

[109] In re Chicken Antitrust Litigation, 560 F. Supp 943 (Ga. 1979).

[110] These ideas are related to Butz (1990), where best-price provisions mitigate the time inconsistency problem that a monopolist faces when selling a durable good, thereby reducing social welfare. In contrast, here the reduction in delay improves social welfare.

[111] In essence, the potential for a future MFN payment lowers the incentive of an early plaintiff to inflate his or her settlement demand. This, in turn, reduces the need for the defendant to reject demands so as to ensure the separation of types.

3.8.5. Secret settlement

It is not uncommon for private litigants to settle their lawsuits quietly, where neither the existence of the suit nor the terms of the settlement are observable to the public. Secrecy may be facilitated through court-ordered sealing of the records, "gag orders," or through the parties themselves in their private contracts. Not surprisingly, secret settlements have attracted attention both in the policy arena and in academia. What are the advantages and disadvantages of allowing litigants to settle in secret? Secret settlements chill the flow of information to the public—information about the existence of legal remedies and the safety of products. There are important externalities at play: the value of this information accrues to future litigants and is not internalized by the settling parties.

Daughety and Reinganum (1999a, 2002) identify two types of information externalities. First, open settlements allow future plaintiffs to learn about the defendant's involvement in a case. A cancer patient, for example, may not know whether the cancer was a natural occurrence or whether the condition was caused by (or exacerbated by) a local waste site. This causal link among cases creates a "publicity effect" whereby an open (as opposed to secret) settlement makes it more likely that another plaintiff will file suit in the future. Second, an open settlement may provide future plaintiffs with information about the expected value of their claim—the "learning effect." Daughety and Reinganum (1999a) focus on the former publicity effect by assuming that the defendant's culpability—assumed to be private information of the defendant—is weakly correlated across plaintiffs. The uninformed plaintiff uses secret settlements as a screen: defendants who know that there are other victims are more likely to settle secretly. The plaintiff is able to extract "hush money" from these defendants in exchange for reducing the flow of information, money that is in effect coming out of the other victims' pockets. In this way, early plaintiffs can enrich themselves at the expense of later plaintiffs.[112] Importantly, secrecy can compromise firms' behavior and product safety choices in a market setting. Daughety and Reinganum (2005) present a signaling model where the average safety of consumer products is higher when firms can commit *ex ante* to settle all future disputes in the open.

3.9. Additional topics

3.9.1. Plea bargaining

Most criminal cases in the United States are resolved before trial through a process known as plea bargaining. The prosecutor and the defendant typically negotiate a guilty plea in exchange for a lesser charge (which is associated with a lighter sentence for

[112] Daughety and Reinganum (2002) introduce the learning effect by assuming strong correlation of culpability across cases. This complicates matters because there is an offsetting effect: as in section 2, defendants who are more culpable are more likely to settle and this creates an adverse inference on the part of future plaintiffs.

the defendant). In many ways, the economic approach to plea bargaining is similar to that of civil settlement. Many themes—including the avoidance of direct litigation costs and risks, the role of private information, and the feedback effects on deterrence and social welfare more broadly—are common to both. They diverge, however, in their assumptions about the prosecutor's preferences. In civil cases, the damages paid by the defendant are typically received by the plaintiff. In criminal cases the prosecutor does not "receive" the prison sentence directly.

The papers in the plea bargaining literature vary in their assumptions regarding the prosecutor's payoff function. Landes (1971), in the first formal analysis of plea bargaining, assumes that the prosecutor maximizes the sum of expected sentences subject to a resource constraint. His approach is more aligned with the models of civil litigation and settlement than the literature that follows. Several subsequent papers, including Grossman and Katz (1983) and Reinganum (1988), assume that the prosecutor represents the interests of society more broadly with a payoff function that includes type I and type II errors in addition to litigation costs.

Grossman and Katz assume that the defendant privately observes his guilt, which is correlated with the probability that he will be convicted at trial. The uninformed prosecutor can make a single take-it-or-leave-it offer of a reduced sentence in exchange for a guilty plea. Although they abstract from the direct costs of litigation, both the defendant and prosecutor are assumed to be risk averse and so, all else equal, they prefer to settle before trial. As in Bebchuk's (1984) analysis of civil settlement, the plea bargain offered by the prosecutor screens among the different defendant types: the guilty defendant is more likely to accept the offer than the defendant who is convinced of his innocence.[113]

Reinganum (1988) extends Grossman and Katz to allow the prosecutor to have better information about the probability of conviction at trial, a variable that is correlated with the defendant's innocence or guilt.[114] The prosecutor's offer of a plea bargain serves as a signal to the defendant who would subsequently update his beliefs about trial. The partial pooling equilibrium has the following form: when the probability of conviction is below a cutoff, then the prosecutor drops the case (formally, the sentence offered is zero). When the probability is above the cutoff, then the offered sentence perfectly reveals the prosecutor's private information and is increasing in the probability of conviction. As in the model of Reingaum and Wilde (1986), the defendant mixes between accepting and rejecting the prosecutor's offer with high sentences being rejected more often.[115] An important implication of this model, one that distinguishes it from Grossman and Katz (1983), is that trials are more likely when the defendant is guilty.

[113] Notice that the prosecutor knows that defendants who reject the plea are more likely to be innocent; thus, he has a greater incentive to drop charges after the plea offer is rejected. This observation is analogous to Nalebuff (1987). Baker and Mezzetti (2001) extend the model to take this into account.

[114] Only the prosecutor's private information is payoff-relevant to the defendant at trial, although the defendant's private information is relevant for the social welfare function (which corresponds to the prosecutor's payoff).

[115] This is an implication of incentive compatibility for the prosecutor.

Should prosecutors be granted full discretion in the offers that they make to defendants before trial? Reinganum (1988) shows that the answer may be "no," even when the prosecutor has the interests of society in mind *ex post*.[116] Reinganum compares the equilibrium described above to a regime where the prosecutor's offer cannot be fine-tuned to the signal received—in other words, a forced "pooling" offer. The equilibrium that results is akin to Grossman and Katz (1983): the pooling offer screens among the defendants where the innocent go to trial and the guilty accept the offer. If there are many guilty defendants in the overall population, then limiting prosecutorial discretion is socially desirable—with discretion, these guilty defendants are associated with costly trials. If there are many innocent defendants, on the other hand, then discretion is preferred. The benefit of the equilibrium described above is that cases against innocent defendants are likely to be dropped.[117]

3.9.2. Case selection and the 50% hypothesis

The cases that end up in trial are the "tip of the iceberg"—the vast majority of filed cases are settled before trial and even more are never filed to begin with. The cases that end up in the courtroom result from bargaining failures—failures that can arise for many reasons: asymmetric information, divergent expectations, the long-run interests of the parties (such as the need to establish precedent in a civil rights case), or a variety of externalities. In sum, the cases that go to trial are an unusual sample of cases and are likely to differ—perhaps systematically—from the cases that never reach the courtroom.

In 1984, Priest and Klein presented a "divergent-expectations" model of litigation and settlement with a striking prediction: for cases that go to trial, the probability of the plaintiff winning tends towards 50%. In contrast, for cases that settle out of court, the probability that the plaintiff would win (had they gone to trial instead) could be systematically higher or lower than 50%.[118] This paper has received a great deal of attention in the law and economics literature for (presumably) several reasons. First, they illustrate that litigated cases are unrepresentative of the broader case population. (Bebchuk, 1984, and others, of course, share this feature.) Second, the "50% hypothesis" is consistent with many people's rough intuition: if the evidence in a case *clearly* favors one party over the other, then the case should settle; if it is *unclear* who should win, there is greater scope for disagreement. Finally, the 50% hypothesis readily lends itself to empirical testing.

Shavell (1996) argues that any plaintiff win rate at trial is possible under more general assumptions. Most obviously, if both parties firmly believe that the probability that the

[116] Reinganum (2000) presents a signaling model where an informed defendant makes an offer to an uninformed prosecutor. In this model, the prosecutor's incentives are not aligned with society's incentives.

[117] To put it somewhat differently, the prosecutor may be the victim of his own discretion. The signaling equilibrium that results from discretion is socially inefficient. The prosecutor is better off if his hands are tied.

[118] Waldfogel (1998) gives a clear presentation of Priest and Klein's assumptions and results and surveys the related empirical literature. Asymmetric stakes, including situations where one side has a stronger interest in the case than the other, would change the results.

plaintiff will prevail is bounded between, say, 0 and 1/3, then the plaintiff win rates for both settled and litigated cases must be below 1/2. Less obviously, one can construct examples with asymmetric information that generate: (1) any given win rate at trial, p^T, and (2) any given (implied) win rate for settled cases p^S. To see why this is true, suppose that the distribution of plaintiff win- rates, $f(p)$, has full support on [0, 1]. We know from the earlier presentation of the basic framework that if the plaintiff has private information about her winning probability then plaintiffs with strong cases (high plaintiff win rates) are more likely to go to trial than those with weak cases (low plaintiff win rates). This information structure is consistent with examples where $p^S < p^T$. One can construct a distribution $f(p)$, litigation costs, c_p and c_d, and an extensive form game consistent with this outcome. If, on the other hand, the defendant privately observes the probability winning, then the selection goes in the other direction: defendants with strong cases (low plaintiff win rates) are more likely to go to trial than those with weak cases (high plaintiff win rates). This information structure is consistent with $p^S > p^T$ and, as above, consistent examples can be constructed.

Priest and Klein's 50% hypothesis was a limiting result where the errors in observation become increasingly precise—in the limit, parties beliefs converge and trial rates approach zero. Given the strong assumptions needed to generate the 50% result, it is not surprising that there is little empirical result for the strong form of the hypothesis.[119] More generally, however, Priest and Klein's framework suggests that trial rates may be systematically related to plaintiff win rates. Empirical work along these lines has been more fruitful. Waldfogel (1995), for example, empirically documents the relationship between litigation rates and plaintiff win rates in a structural model using data from broad variety of cases (including tort, contract, property, and civil rights) and argues that his results are consistent with the theory. See also Kessler et al. (1996) and Waldfogel's (1998) survey of the literature.

Although most of the literature on the selection of cases for trial uses the divergent expectations frameworks, some papers have investigated the relationship between trial rates and plaintiff win rates using asymmetric information. Froeb (1993) considers plea bargaining involving a privately informed defendant where guilty defendants are more likely to accept pleas. In his theoretical framework, factors that would increase the number of cases tried (such as a fall in prosecution costs) should increase the average guilt of defendants who go to trial. This pattern is documented in Froeb's sample of criminal cases.

3.9.3. Decoupling

The preceding discussion has assumed that any judgment against a defendant is automatically awarded to the plaintiff. Tying the defendant's liability to the plaintiff's award

[119] Many authors have documented systematic deviations from this number. See the references in Waldfogel (1998).

is also not necessarily optimal in theory. Polinsky and Che (1991) present a simple framework where a defendant chooses his level of precautions to reduce the probability of an accident, and an injured plaintiff would bring suit only if his or her expected award, a, exceeded the cost of litigating suit, c_p (which is distributed according to density $f(c_p)$ in the plaintiff population). Suppose the defendant's liability is l. There are two potential sources of inefficiency in this framework: (1) inefficient precautions by the defendant and (2) the wasted resources associated with litigation. The first-best outcome is not achievable when the defendant's liability is tied to the plaintiff's award. The award/liability level that achieves zero process costs, $a = l = 0$, provides the defendant with no incentives for care. (Conversely, the level that provides incentives for care leads to a deadweight loss at trial.) The optimal decoupled scheme makes the plaintiff's award, a, very small so that only a handful of cases are brought, and, at the same time, it makes the defendant's liability l very large so that his expected future liability equals the social harm that his actions cause. Decoupling creates a strong incentive for settlement because it creates a wedge between the most that the defendant is willing to pay and the least that the plaintiff is willing to accept: $[a - c_p, l + c_d]$.

Decoupling liability also appears in practice. Several states, including Iowa and Indiana, have adopted split-award statutes where the government keeps a percentage of punitive awards.[120] Kahan and Tuckman (1995) observe that these statutes can lead to an uneven playing field since the defendant's stakes are so much larger than the plaintiff's. Consequently, the defendant's incentives for care may be compromised by the statute because the plaintiff's incentives to invest in the case following an accident are reduced. See also Choi and Sanchirico (2004). Incentives are further compromised because the negotiated settlement will reflect, to a greater or lesser extent, the plaintiff's award at the bottom of the bargaining zone. See Daughety and Reinganum (2003).

3.9.4. Patent litigation

Suppose that a patentee and an imitator are negotiating prior to trial. If the case goes before a judge, there is a chance that the patent will be invalidated and the imitator will subsequently compete on equal footing with the innovator. The private settlement of these cases can be a legal way for competitors (or potential competitors) to collude and preserve market power at the expense of consumers.

Collusion through settlement is perhaps most obvious when the settlement between an owner of a patent and a current (or potential) imitator involves a lump-sum fee along with an agreement not to compete in the future. The patentee and the imitator can divide the monopoly profits between them through the lump-sum payment according to their relative bargaining strengths. Collusion can be achieved in more subtle ways, including

[120] In Iowa, for example, the plaintiff keeps only 25% of the punitive damage award. This example and others are discussed in Daughety and Reinganum (2003). See Landeo, Nikitin, and Babcock (2007) for a recent experimental analysis.

joint ventures (where one supplies the other) and a variety of royalty arrangements. It is not surprising that the Department of Justice and the Federal Trade Commission have watched these settlements with interest and have investigated many of them. See Shapiro (2003) for a good discussion of these mechanisms, the relevant cases, and proposed criteria for judicial approval of patent settlements.[121]

Competitors may also be able to use the threat of patent litigation to soften competition in a patent race. Marshall, Meurer and Richard (1994) describe a scenario where the loser of a patent race retains the option of bringing an action against the winner of the race.[122] Since the loser may succeed in this action, the possibility of patent litigation serves to reduce the differential value between winning the race and losing the race. The competitors subsequently spend less on research than otherwise, something that is typically in their joint interest. The overall impact of this collusive strategy on society more broadly (which includes consumer welfare) could be either positive or negative. See Reinganum (1989) for an excellent survey of the patent race literature.

Choi (1998) explores strategic decision making and externalities in repeated patent litigation. When a patentee is deciding whether or not to pursue an imitator, he must think carefully about the impact that a legal decision will have on future entrants. If the patent is deemed invalid and subsequently revoked, then the floodgates will open and industry profits will be dissipated. (On the other hand, if the patent is deemed valid, then the patentee will enjoy greater protection from the court's decision.) The patentee's decision to litigate the case depends upon the nature of these externalities.

Choi shows that when the patent is very likely to be upheld, then the patentee has an incentive to litigate. Less obviously, when the patent is likely to be overturned, then the patentee still has an incentive to litigate. The upside, the off-chance that the current imitator will be eliminated from the market, is stronger than the downside, the increase in entry once the patent is invalidated. The key to this last result is that entry will occur with or without a formal invalidation when the patent is very weak to begin with. When the probability of patent invalidation is in an intermediate range, however, the patentee chooses to tolerate early infringement and imitation.

Lanjouw and Schankerman (2001) document interesting correlations between litigation decisions and the characteristics of the patents that are involved. Many of the empirical findings are consistent with the predictions of the theory. First, they show that a patent is more likely to be litigated if it serves as the "base of a cumulative chain"—in other words, there are more rents to captured from future innovators. Second, they

[121] Meurer (1989) argues that patent cases may lead to more rather than less litigation when the settlements are closely regulated. To take an extreme case, when a regulator can force the patentee and the imitator to compete duopoly style following a settlement, it would be in their joint interest to go to trial in the hopes of establishing patent validity and the chance of monopoly rents.

[122] The patent example is informally described in the conclusion of their paper. Their formal model is one of procurement, where the loser of the auction can take action to reverse the outcome. The possibility of future legal action softens competition between the bidders during the auction. Intuitively, the differential value between winning and losing is smaller than before, since the loser may succeed in winning upon protest or appeal.

find support for the idea that firms establish reputations for protecting their intellectual property through litigation.

3.9.5. Insurance contracts

It is common for liability insurance contracts to: (1) place a dollar limit on coverage and (2) impose on the insurer a duty to defend the case, including the authority over settlement and the obligation to bear the legal costs of defense. We will see that these contracts create a conflict of interest between the insurer and the holder of the policy in the event of an accident, that these contracts potentially impose negative externalities on the victims of accidents, and that there is scope for legal intervention in these contracts.

Suppose that a victim—the plaintiff—has been harmed by an insured defendant. The defendant has an insurance policy, provided by a competitive insurance market, with a coverage limit of L. The plaintiff expects a random return at trial, x, from distribution $f(x)$ with expectation $E(x)$. If the case goes to trial and $x > L$ then the insurance company pays L and the defendant bears the residual, $x - L$. Suppose further that the insurer and the plaintiff would each bear litigation costs c_p and c_d if the case went to trial. (This is consistent with the insurance company's so-called "duty to defend.") If the case went to trial, the plaintiff would receive $E(x) - c_p$ in expectation—the expected damages at trial minus her litigation costs. When taken jointly, the insurance company and the defendant expect to pay $E(x) + c_d$. If they were jointly negotiating with the plaintiff, the case would surely settle for some amount in this range.

Conflicts of interest between the defendant and the insurer arise when the insurance company holds the authority to settle (Meurer, 1992; Sykes, 1994). If the case goes to trial, the insurer's expected losses are:

$$\int_0^L xf(x)dx + \int_L^\infty Lf(x)dx + c_d < E(x) + c_d.$$

The coverage limit makes the insurer tough in negotiations, lowering the most that he is willing to pay in settlement. This is because the defendant bears the downside associated with adverse judgments above the policy limit: $\int_L^\infty (x - L) f(x)dx$. Consequently, the insurance company may choose to litigate the case even though it is in the joint interest of the insurance company and the defendant to settle the case, instead.

Why would the defendant and the insurer ever write a contract that creates these *ex post* conflicts? Meurer (1992) argues that the insurance contract has strategic value for the defendant: it is a strategic commitment to be "tough" in settlement negotiations. By reducing the most that the insurer is willing to pay in settlement, the contract serves to extract value from the plaintiff by lowering the plaintiff's settlement demand. To put it somewhat differently, even though the conflict of interest may end up harming the defendant *ex post*, it may be in his advantage to sign such a contract *ex ante*. Indeed, a risk-neutral defendant with no liquidity constraints has an incentive to buy an insurance contract with these features. The value is coming not from risk aversion but from the strategic effects in settlement negotiations.

These insurance contracts are problematic from a social welfare perspective when neither the defendant nor the insurance company can foresee the distribution $f(x)$ at the time of contracting. Settlement negotiations can break down on the equilibrium path in this instance, leading to the *dead-weight loss* of trials $(c_p + c_d)$. In theory, this is similar to Aghion and Bolton's (1987) analysis of liquidated damages clauses. There, supply contracts with penalty clauses for breach have private strategic value because they encourage entrants to price more aggressively but create a social loss when they prevent entry altogether. Here, the insurance contract induces the plaintiff to accept a lower settlement amount but may lead to a breakdown in negotiations altogether.[123]

The *ex post* conflicts of interest between defendants and their liability insurers are mitigated in practice by the duty of liability insurers to act in good faith. A famous case, *Crisci v. Security Insurance Co.*, laid out the obligations of insurers in litigating and settling cases.[124] In particular, the court held that "the insurer must give the interests of the insured at least as much consideration as it gives its own interests."[125] Formally, this may be interpreted as forcing the insurer to equally weigh the defendant's expected payments at trial, $\int_L^\infty (x - L) f(x) dx$. This would be in the interest of public policy, as well. Crisci removes the insurer's incentive to play tough in negotiations and achieve a settlement in the range $[E(x) - c_p, E(x) + c_d]$.

3.9.6. Financial distress

Spier (2002) considers the role of settlement externalities in negotiations between one defendant and two plaintiffs when the defendant's wealth is insufficient to cover the total level of damages should both plaintiffs win at trial. Negotiations can break down when the probabilities that the plaintiffs will win at trial are sufficiently positively correlated. This may be seen in a simple example. Suppose that two plaintiffs, each of whom has suffered damages of \$80, are suing the same defendant. The defendant's wealth is \$80, sufficient to cover exactly one claim. Suppose that each plaintiff will prevail with probability $1/2$ and that the claims are perfectly correlated. If both plaintiffs go to trial, they each will receive \$20 in expectation (they win with probability $1/2$ and then split the defendant's wealth between them). But the plaintiffs will refuse to settle for \$20 each: if one plaintiff accepts the \$20, the other would rather go to trial with the expectation of winning the remaining \$60 with probability $1/2$! When the plaintiffs are acting individually, the defendant would have to offer a hefty settlement premium to get both to accept.

Suppose instead that the plaintiffs' cases are uncorrelated. The expected return for each plaintiff is \$30 if both cases go to trial. Perhaps surprisingly, the defendant may

[123] Note that the social cost of trials may be avoidable here if the defendant can engage in negotiations as well and contribute to the settlement. In practice, transactions costs and information asymmetries may interfere with renegotiation along these lines. See Sykes (1994).

[124] 66 Cal. 2d 425, 58 Cal Rptr. 13, 426 P.2d 173 (1967).

[125] 66 Cal. 2d at 429.

succeed in coercing the plaintiff to settle for less than this amount. If one plaintiff settles out of court for $30, then $50 remains in the defendant's coffers. Since the non-settling plaintiff wins at trial half the time, his expected return is only $25. In general, when the degree of correlation between the cases is low, then the defendant's insolvency creates a "rush to collect" by the plaintiffs and may allow the defendant to successfully chisel his offers.

Spier and Sykes (1998) explore how the capital structure of a defendant can affect negotiations with a single plaintiff when a civil judgment may bankrupt the firm. Not surprisingly, the outcome of settlement negotiations depends critically on a firm's debt level and on the priority afforded the debtholders in bankruptcy. Most interestingly, they show that even junior debt can have strategic value by making the shareholders into tougher negotiators. The intuition rests on the fact that while any settlement that is paid will come out of the shareholders' pockets, a large award at trial will (eventually) come out of the debtholders' pockets. Although the presence of junior debt does not directly dilute the value of the tort claim, it does so indirectly through its impact on the bargaining range.

3.9.7. Strict liability and negligence rules

There is a large law and economics literature that compares strict liability to negligence.[126] Under strict liability, a defendant is held responsible for a plaintiff's damages regardless of his level of care—the defendant is forced to compensate the plaintiff even if he took appropriate precautions to avoid the accident. Under the negligence rule, the defendant is held liable only if he failed to take due care. At first glance, strict liability appears to have higher transactions costs because every single accident involving the defendant creates a meritorious legal claim. Under the negligence rule, on the other hand, a meritorious claim arises only if the defendant was negligent. There are countervailing factors, however. First, a given case brought under a negligence rule may be more expensive to try: the plaintiff must establish the defendant's negligence in addition to causation. Second, cases brought under a negligence rule may be less likely to settle since there is more scope for disagreement because the plaintiff and defendant may have different perceptions or information concerning the defendant's level of care. Again, this would lead to higher transactions costs than otherwise. Finally, the negligence rule can lead the defendant to engage in higher levels of the risky economic activity, creating more accidents and potentially more lawsuits. When these additional factors are taken into account, the negligence rule may actually be more expensive than strict liability.

[126] See Shavell (2004) for a survey of this literature. In a frictionless world, both rules can lead to socially optimal levels of care although they differ in the effects on the level of economic activity.

3.9.8. Scheduled damages

Another strand of the literature focuses on the accuracy of legal rules. In a model without settlement, Kaplow and Shavell (1994, 1996) argued that legal rules that fine-tune liability to the plaintiff's actual level of harm lead to social waste. It is expensive for the litigants (and for the court) to verify the precise level of damages at trial. The "scheduling" of damages, or standardizing awards for injuries that fall into particular categories (as in workers' compensation), can reduce the transactions costs of litigation. Scheduling can be valuable when settlement is possible as well (Spier, 1994c). If the plaintiff has private information about his damages, for example, then negotiations will sometimes break down and litigation costs will be borne. Scheduling makes the future outcome of the case more transparent—there is less to argue about—and can help to promote settlement and save on litigation costs. There may be a downside of damage schedules, however: schedules may be less effective in encouraging potential injurers to take precautions to reduce the magnitude of harm. Spier (1992b) makes a related point in the context of contract design. In a risk-sharing model, it is shown that the benefits associated with optimal risk sharing may be outweighed by the transactions costs of enforcing these contracts (e.g., costly state verification and equilibrium litigation activity).

4. Conclusion

The purpose of this chapter has been to survey the vast literature on litigation and settlement and to synthesize its main themes. Although the set of issues raised in this chapter are very broad and far reaching, the approach taken in this chapter was quite focused. The first underlying premise was that those engaged in litigation—the litigants, their attorneys, the judges, and other interested parties—are rational, self-interested, and strategic. A second premise was that the primary purpose of the court system should be to facilitate value-creating economic activity and deter value-destroying economic activity.

Section 2 of this chapter laid out the basic economic framework of a potential lawsuit involving a single plaintiff and a single defendant. It carefully examined the plaintiff's decision to pursue litigation and the decision of the plaintiff and defendant to settle their case out of court. The positive and normative implications that arise in this setting are surprisingly subtle. We saw that the plaintiff and defendant's litigation strategies depend critically upon the timing of moves and on the information available to them. We also saw that the private interests of the plaintiff and the defendant in litigation are not necessarily aligned with the interests of society, more broadly. The plaintiff may bring "too many" lawsuits, choosing to hire an expensive attorney to battle a defendant in court even in situations when the transfer of funds has little corresponding social benefit. The plaintiff may also pursue "too few" lawsuits, particularly if the anticipation of legal action is a public good, giving potential defendants incentives to take precautions.

Since an individual plaintiff does not personally capture the social value of deterrence in this instance, she may not sue often enough, from a societal perspective. Similarly, the level of expenditures by the plaintiff and the defendant and their private incentives to settle their dispute are not necessarily aligned with the interests of society. Section 3 showed how this framework has been productively applied in the law and economics literature.

Economic approaches to studying litigation will likely continue to be popular in the future. It is clear that a deeper understanding of the economic issues is necessary and critical for those who are creating public policy. The subtleties—and excitement—of these issues will also continue to capture the imagination of talented young economists and legal scholars for years to come. One way that the frontiers are being pushed forward is through the introduction of new methods and approaches. The law and economics field more broadly has experienced a growing influence of behavioral assumptions and experimental techniques in addition to the ongoing growth of empirical work. These approaches are very valuable, allowing us to resolve theoretical ambiguity and pinpoint the magnitude of economic effects. On the theoretical side, there is a definite need to explore more fully the normative implications of settlement and litigation. There has been relatively little work, for example, on how the transactions costs of litigation affect the design of contracts, the organization of economic activity, and the boundaries of the firm.

References

Aghion, P., Bolton, P. (1987). "Contracts as a barrier to entry". American Economic Review 77, 388–401.

Andreoni, J. (1991). "Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime?" RAND Journal of Economics 22, 385–395.

Ashenfelter, O., Bloom, L. (1984). "Models of arbitrator behavior: Theory and evidence". American Economic Review 74, 111–124.

Ashenfelter, O., Currie, J., Farber, H.S., Spiegel, M. (1992). "An experimental comparison of dispute rates in alternative arbitration systems". Econometrica 6, 1407–1433.

Aumann, R. (1976). "Agreeing to disagree". Annals of Statistics 4, 1236–1239.

Austen-Smith, D., Banks, J. (1996). "Information aggregation, rationality, and the Condorcet jury theorem". American Political Science Review 90, 34–45.

Ayres, I., Nalebuff, B. (1997). "Common knowledge as a barrier to negotiation". UCLA Law Review 80, 323–401.

Baker, S., Mezzetti, C. (2001). "Prosecutorial resources, plea bargaining and the decision to go to trial". Journal of Law, Economics & Organization 17, 149–167.

Bar-Gill, O. (2006). "Evolution and persistence of optimism in litigation". Journal of Law, Economics & Organization 22, 490–507.

Bebchuk, L.A. (1984). "Litigation and settlement under imperfect information". RAND Journal of Economics 15, 404–415.

Bebchuk, L.A. (1988). "Suing solely to extract a settlement offer". Journal of Legal Studies 17, 437–450.

Bebchuk, L.A. (1996). "A new theory concerning the credibility and success of threats to sue". Journal of Legal Studies 25, 1–25.

Bebchuk, L.A., Chang, H. (1996). "An analysis of fee shifting based on the margin of victory: On frivolous suits, meritorious suits, and the role of Rule 11". Journal of Legal Studies 25, 371–403.

Bebchuk, L.A., Chang, H. (1999). "The effect of offer-of-settlement rules on the terms of settlement". Journal of Legal Studies 28, 489–513.

Bebchuk, L.A., Guzman, A. (1996). "How would you like to pay for that? The strategic effects of fee arrangements on settlement terms". Harvard Negotiation Law Review 1, 53–63.

Bebchuk, L.A., Kaplow, L. (1992). "Optimal sanctions when individuals are imperfectly informed about the probability of apprehension". Journal of Legal Studies 21, 365–370.

Becker, G. (1968). "Crime and punishment: And economic approach". Journal of Political Economy 76, 169–217.

Bernardo, A.E., Talley, E., Welch, I. (2000). "A theory of legal presumptions". Journal of Law, Economics, & Organization 16, 1–49.

Bernstein, L. (1992). "Opting out of the legal system: Extralegal contractual relations in the diamond industry". Journal of Legal Studies 21, 115–157.

Bernstein, L. (1993). "Understanding the limits of court-connected ADR: A critique of federal court-annexed arbitration programs". University of Pennsylvania Law Review 141, 2169–2233.

Bikchandani, S., Hirschleifer, D., Welch, I. (1998). "Learning from the behavior of others: Conformity, fads, and informational cascades". Journal of Economic Perspectives 12, 151–170.

Borenstein, S. (1996). "Settling for coupons: Discount contracts as compensation and punishment in antitrust lawsuits". Journal of Law and Economics 39, 379–404.

Braeutigam, R., Owen, B., Panzar, J. (1984). "An economic analysis of alternative fee shifting systems". Law and Contemporary Problems 47, 173–204.

Bulow, J.I., Klemperer, P.D. (1998). "The tobacco deal". Brookings Papers on Economic Activity: Microeconomics, 323–394.

Bulow, J.I., Geanakoplos, J.D., Klemperer, P.D. (1985). "Multimarket oligopoly: Strategic substitutes and complements". Journal of Political Economy 93, 488–511.

Bustos, A.E. (2006). A Dynamic Theory of Corporate Common Law, Princeton University Doctoral Dissertation.

Butz, D.A. (1990). "Durable goods monopoly and best-price provisions". American Economic Review 80, 1062–1076.

Chang, H.F., Sigman, H. (2000). "Incentives to settle under joint and several liability: An empirical analysis of superfund litigation". Journal of Legal Studies 24, 205–236.

Che, Y.-K. (1996). "Equilibrium formation of class action suits". Journal of Public Economics 62, 339–361.

Che, Y.-K. (2002). "The economics of collective negotiations in pretrial bargaining". The International Economic Review 43, 549–576.

Che, Y.-K., Yi, J.G. (1993). "The role of precedents in repeated litigation". Journal of Law, Economics, and Organization 9, 399–424.

Chen, K.-P., Chien, H.-K., Chu, C.Y.C. (1997). "Sequential versus unitary trials with asymmetric information". The Journal of Legal Studies 26, 239–258.

Cho, I.-K., Kreps, D. (1987). "Signalling games and stable equilibria". Quarterly Journal of Economics 102, 179–221.

Choi, J.P. (1998). "Patent litigation as an information-transmission mechanism". American Economic Review 88, 1249–1263.

Choi, A. (2003). "Allocating settlement authority under a contingent-fee arrangement". The Journal of Legal Studies 32, 585–610.

Choi, A., Sanchirico, C.W. (2004). "Should plaintiffs win what defendants lose? Litigation stakes, litigation effort, and the benefits of decoupling". Journal of Legal Studies 22, 323–354.

Clermont, K., Eisenberg, T. (1992). "Trial by jury or judge: Transcending empiricism". Cornell University Law Review 77, 1124–1177.

Condorcet, M. de (1785). Essais Sur L'appication de L'analyse a la Probabilite des Decisions Rendues a la Puralite des Voix, Paris (Trans. 1994 by Iain McLean and Fiona Hewitt).

Cooter, R., Kornhauser, L., Lane, D. (1979). "Liability rules, limited information, and the role of precedent". The Bell Journal of Economics 10, 366–373.

Cooter, R.D., Rubinfeld, D.L. (1989). "Economic analysis of legal disputes and their resolution". Journal of Economic Literature 27, 1067–1097.
Cooter, R.D., Rubinfeld, D.L. (1994). "An economic model of legal discovery". Journal of Legal Studies 23, 435–464.
Dana, J.D., Spier, K.E. (1993). "Expertise and contingent fees: The role of asymmetric information in attorney compensation". The Journal of Law, Economics, & Organization 9, 349–367.
Danzon, P.M. (1983). "Contingent fees for personal injury litigation". Bell Journal of Economics 14, 213–223.
Danzon, P., Lillard, L. (1983). "Settlement out of court: The disposition of medical Malpractice Claims". Journal of Legal Studies 12, 345–378.
Daughety, A.F. (2000). "Settlement". In: Bouckaert, B., De Geest, G. (Eds.), Encyclopedia of Law and Economics, vol. 5. Edward Elgar Publishing Co, Cheltenham, pp. 95–158.
Daughety, A.F., Reinganum, J.F. (1993). "Endogenous sequencing in models of settlement and litigation". Journal of Law, Economics and Organization 9, 314–349.
Daughety, A.F., Reinganum, J.F. (1994). "Settlement negotiations with two-sided asymmetric information: Model duality, information distribution and efficiency". International Review of Law and Economics 14, 283–298.
Daughety, A.F., Reinganum, J.F. (1995). "Keeping society in the dark: On the admissibility of pretrial negotiations as evidence in court". RAND Journal of Economics 26, 203–221.
Daughety, A.F., Reinganum, J.F. (1999a). "Hush money". RAND Journal of Economics 30, 661–678.
Daughety, A.F., Reinganum, J.F. (1999b). "Stampede to judgment: Persuasive influence and herding behavior by courts". American Law and Economics Review 1, 158–189.
Daughety, A.F., Reinganum, J.F. (2000a). "Appealing judgments". RAND Journal of Economics 31, 502–525.
Daughety, A.F., Reinganum, J.F. (2000b). "On the economics of trials: Adversarial process, evidence, and equilibrium bias". Journal of Law, Economics, & Organization 16, 365–394.
Daughety, A.F., Reinganum, J.F. (2002). "Information externalities in settlement bargaining: Confidentiality and correlated culpability". RAND Journal of Economics 33 (4), 587–604.
Daughety, A.F., Reinganum, J.F. (2003). "Found money? Split-award statutes and settlement of punitive damages cases". American Law and Economics Review 5, 134–164.
Daughety, A.F., Reinganum, J.F. (2004). "Exploiting future settlements: A signaling model of most-favored-nation clauses in settlement bargaining". RAND Journal of Economics 35 (3), 467–485.
Daughety, A.F., Reinganum, J.F. (2005). "Secrecy and safety". American Economic Review 95, 1074–1091.
Demougin, D., Fluet, C. (2006). "Preponderance of evidence". European Economic Review 50, 963–976.
Dewatripont, M., Tirole, J. (1999). "Advocates". Journal of Political Economy 107, 1–39.
Dixit, A. (1987). "Strategic behavior in contests". American Economic Review 77, 891–898.
Dixit, A. (2003). "Arbitration and information", Princeton University Working Paper.
Emons, W. (2000). "Expertise, contingent fees, and insufficient attorney effort". International Review of Law and Economics 20, 21–33.
Farber, H.S. (1980). "An analysis of final offer arbitration". Journal of Conflict Resolution 35, 683–705.
Farber, H.S., White, M.J. (1991). "Medical malpractice: An empirical examination of the litigation process". Rand Journal of Economics 22, 199–217.
Farmer, A., Pecorino, P. (1994). "Pretrial negotiations with asymmetric information on risk preferences". International Review of Law and Economics 14, 273–281.
Farmer, A., Pecorino, P. (2000). "Conditional cost shifting and the incidence of trial: Pretrial bargaining in the face of a Rule 68 offer". American Law and Economics Review 2, 318–340.
Farmer, A., Pecorino, P. (2003). "Bargaining with voluntary transmission of private information: Does the use of final offer arbitration impede settlement?" Journal of Law, Economics & Organization 19, 64–82.
Feddersen, T., Pesendorfer, W. (1998). "Convicting the innocent: The inferiority of unanimous jury verdicts". American Political Science Review 92, 23–35.
Fon, V., Parisi, F. (2003). "Litigation and the evolution of legal remedies: A dynamic analysis". Public Choice 116, 419–433.
Fon, V., Parisi, F. (2004). "Judicial precedents in civil law systems: A dynamic analysis", George Mason Law & Economics Research Paper No. 04-15.

Fournier, G.M., Zuehlke, T.W. (1996). "The timing of out-of-court settlements". RAND Journal of Economics 27, 310–321.
Froeb, L.M. (1993). "Adverse selection of cases for trial". International Review of Law and Economics 13, 317–324.
Froeb, L.M., Kobayashi, B.H. (1996). "Naive, biased, yet bayesian: Can juries interpret selectively produced evidence?". Journal of Law, Economics, & Organization 12, 257–276.
Gennaioli, N., Shleifer, A. (2005). "The evolution of precedent", NBER Working Paper No. W11265.
Gibbons, R. (1988). "Learning in equilibrium models of arbitration". American Economic Review 78, 896–912.
Gould, J. (1973). "The economics of legal conflicts". Journal of Legal Studies 2, 279–300.
Grossman, S.J. (1981). "The informational role of warranties and private disclosure about product quality". Journal of Law and Economics 24, 461–483.
Grossman, G.M., Katz, M.L. (1983). "Plea bargaining and social welfare". American Economic Review 73, 749–757.
Hadfield, G.K. (2006). "The quality of law in civil code and common law regimes: Judicial incentives, legal human capital, and the evolution of law", working paper. (Available at http://law.bepress.com/alea/16th/art40.)
Hause, J. (1989). "Indemnity, settlement, and litigation, or I'll be suing you". Journal of Legal Studies 18, 157–180.
Hay, B.L. (1993). "Some settlement effects of preclusion". Illinois Law Review 1993, 21–57.
Hay, B.L. (1994). "Civil discovery: Its effects and optimal scope". Journal of Legal Studies 23, 481–517.
Hay, B.L. (1995). "Effort, information, settlement, trial". Journal of Legal Studies 24, 29–62.
Hay, B.L. (1997). "Optimal contingent fees in a world of settlement". Journal of Legal Studies 26, 259–278.
Hay, B.L., Spier, K.E. (1997). "Burdens of proof in civil litigation: An economic perspective". Journal of Legal Studies 26, 413–433.
Hay, B., Spier, K.E. (1998). "Settlement of litigation". In: Newman, P. (Ed.), The New Palgrave Dictionary of Economics and the Law, vol. 3. Macmillan Reference Limited, London, pp. 442–451.
Helland, E., Tabarrok, A. (2000). "Runaway judges? Selection effects and the jury". Journal of Law, Economics & Organization 16, 306–333.
Helland, E., Tabarrok, A. (2003). "Contingency fees, settlement delay, and low-quality litigation: Empirical evidence from two datasets". Journal of Law, Economics, and Organization 19, 517–542.
Hua, X., Spier, K.E. (2005). "Information and externalities in sequential litigation". Journal of Institutional and Theoretical Economics 161, 215–232.
Hughes, J.W., Snyder, E.A. (1998). "Allocation of legal costs: American and English rules". In: Newman, P. (Ed.), The New Palgrave Dictionary of Economics and the Law, vol. 1. Macmillan Reference Limited, London, pp. 51–56.
Hylton, K.N. (2002). "Welfare implication of costly litigation under strict liability". American Law and Economics Review 4, 18–42.
Hylton, K.N. (2006). "Information, litigation, and common law evolution". American Law and Economics Review 8, 33–61.
Jost, P.J. (1995). "Disclosure of information and incentives for care". International Review of Law and Economics 15, 65–85.
Judicial Business of the United States Courts (2001). Annual Report 2001. Available online at: http://www.uscourts.gov/judbus2001/contents.html.
Kalven, H. Jr., Zeisel, H. (1966). The American Jury. Little Brown, Boston.
Kahan, M., Tuckman, B. (1995). "Special levies for punitive damages". International Review of Law and Economics 15, 175–185.
Kaplow, L. (1993). "Shifting plaintiffs' fees versus increasing damage awards". RAND Journal of Economics 24, 625–630.
Kaplow, L. (1998). "Accuracy in adjudication". In: The New Palgrave Dictionary of Economics and the Law, vol. 1. Macmillan Reference Limited, London, pp. 1–7.

Kaplow, L., Shavell, S. (1992). "Private versus socially optimal provision of ex-ante legal advice". Journal of Law, Economics, and Organization 8, 306–320.
Kaplow, L., Shavell, S. (1994). "Accuracy in the assessment of liability". Journal of Law and Economics 37, 1–16.
Kaplow, L., Shavell, S. (1996). "Accuracy in the assessment of damages". Journal of Law and Economics 39, 191–209.
Katz, A.W. (1987). "Measuring the demand for litigation: Is the English Rule really cheaper?". Journal of Law, Economics, and Organization 3, 143–176.
Katz, A.W. (1988). "Judicial decisionmaking and litigation expenditure". International Review of Law and Economics 8, 127–143.
Katz, A.W. (1990). "The effect of frivolous lawsuits on the settlement of litigation". International Review of Law and Economics 10, 3–27.
Kim, J.-Y., Ryu, K. (2000). "Pretrial negotiation behind closed doors versus open doors: Economic analysis of Rule 408". International Review of Law and Economics 20, 285–294.
Kessler, D., Meites, T., Miller, G.P. (1996). "Explaining deviations from the 50% rule: A multimodal approach to the selection of cases for litigation". Journal of Legal Studies 25, 233–259.
Klerman, D. (1996). "Settling multidefendant lawsuits: The advantage of conditional setoff rules". Journal of Legal Studies 25, 445–461.
Kornhauser, L.A. (1992). "Modeling collegial courts, 2: Legal doctrine". Journal of Law, Economics, and Organization 8, 441–470.
Kornhauser, L.A., Revesz, R.L. (1994a). "Multidefendant settlements: The impact of joint and several liability". Journal of Legal Studies 23, 41–76.
Kornhauser, L.A., Revesz, R.L. (1994b). "Multidefendant settlements under joint and several liability: The problem of insolvency". Journal of Legal Studies 23, 517–542.
Kornhauser, L.A., Revesz, R.L. (1998). "Joint and several liability". In: Newman, P. (Ed.), The New Palgrave Dictionary of Economics and the Law, vol. 2. Macmillan Reference Limited, London, pp. 371–376.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R. (1997). "Legal determinants of external finance". Journal of Finance 52, 1131–1150.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R. (1998). "Law and finance". Journal of Political Economy 106, 1113–1155.
Landeo, C.M., Nikitin, M., Babcock, L. (2007). "Split-awards and disputes: An experimental study of a strategic model of litigation". Journal of Economic Behavior and Organization 63, 553–572.
Landes, W.M. (1971). "An economic analysis of the courts". Journal of Law and Economics 14, 61–107.
Landes, W.M. (1993). "Sequential versus unitary trials: An economic analysis". Journal of Legal Studies 22, 99–134.
Landes, W.M., Posner, R.A. (1976). "Legal precedents: A theoretical and empirical analysis". Journal of Law and Economics 19, 249–307.
Lanjouw, J.O., Schankerman, M. (2001). "Characteristics of patent litigation: A window on competition". RAND Journal of Economics 32, 129–151.
Levy, G. (2005). "Careerist judges". RAND Journal of Economics 36, 275–297.
Lewis, T., Poitevin, M. (1997). "Disclosure of information in regulatory proceedings". Journal of Law, Economics & Organization 13, 50–73.
Loewenstein, G., et al. (1993). "Self-serving assessments of fairness and pretrial bargaining". Journal of Legal Studies 22, 135–158.
Marshall, R.C., Meurer, M.J., Richard, J.-F. (1994). "Litigation settlement and collusion". Quarterly Journal of Economics 109, 211–239.
Meurer, M. (1989). "The settlement of patent litigation". RAND Journal of Economics 20, 77–91.
Meurer, M. (1992). "The gains from faith in an unfaithful agent: Settlement conflict between defendants and liability insurer". Journal of Law, Economics and Organization 8, 502–522.
Miceli, T. (1994). "Do contingency fees promote excessive litigation?". Journal of Legal Studies 23, 211–224.
Miceli, T., Cosgel, M.M. (1994). "Reputation and judicial decision making". Journal of Economic Behavior and Organization 23, 31–51.

Milgrom, P., Roberts, J. (1986). "Relying on the information of interested parties". RAND Journal of Economics 17, 18–32.
Milgrom, P., Weber, R. (1982). "A theory of auctions and competitive bidding". Econometrica 50, 1089–1122.
Miller, G.P. (1986). "An economic analysis of Rule 68". Journal of Legal Studies 15, 93–125.
Miller, G.P. (1987). "Some agency problems in settlement". Journal of Legal Studies 16 (1), 189–215.
Miller, G.P. (1998). "Class Actions". In: Newman, P. (Ed.), The New Palgrave Dictionary of Economics and the Law, vol. 1. Macmillan Reference Limited, London, pp. 257–262.
Mnookin, R.H. (1998). "Alternative dispute resolution". In: Newman, P. (Ed.), The New Palgrave Dictionary of Economics and the Law, vol. 1. Macmillan Reference Limited, London, pp. 56–60.
Mnookin, R.H., Wilson, R. (1998). "A model of efficient discovery". Games and Economic Behavior 25, 219–250.
Mnookin, R.H., Kornhauser, L. (1979). "Bargaining in the shadow of the law: The case of divorce". Yale Law Journal 88, 950–997.
Mukhopadhaya, K. (2003). "Jury size and the free rider problem". Journal of Law, Economics & Organization 19, 24–44.
Myerson, R. (1979). "Incentive compatibility and the bargaining problem". Econometrica 47, 61–73.
Myerson, R., Satterthwaite, M. (1983). "Efficient mechanisms for bilateral trading". Journal of Economic Theory 29, 265–281.
Nalebuff, B. (1987). "Credible pretrial negotiation". RAND Journal of Economics 18, 198–210.
Neeman, Z., Klement, A. (2005). "Against compromise: A mechanism design approach". Journal of Law, Economics, & Organization 21, 285–314.
Note (1992). "Exposing the extortion gap: An economic analysis of the rules of collateral estoppel". Harvard Law Review 105, 1940–1960.
Ostrom, B.J., Kauder, N.B., LaFountain, R.C. (2001). Examining the Work of the State Courts, 1999–2000. National Center for State Courts, Williamsburg, VA.
Parisi, F. (2002). "Rent seeking through litigation: Adversarial and inquisitorial systems compared". International Review of Law and Economics 22, 193–216.
P'ng, I.P.L. (1983). "Strategic behavior in suit, settlement, and trial". RAND Journal of Economics 14, 539–550.
P'ng, I.P.L. (1986). "Optimal subsidies and damages in the presence of judicial error". International Review of Law and Economics 6, 101–105.
Polinsky, A.M., Che, Y.-K. (1991). "Decoupling liability: Optimal incentives for care and litigation". Rand Journal of Economics 22, 562–570.
Polinsky, A.M., Rubinfeld, D.L. (1988a). "The deterrent effects of settlements and trials". International Review of Law and Economics 8, 109–116.
Polinsky, A.M., Rubinfeld, D.L. (1988b). "The welfare implications of costly litigation for the level of liability". Journal of Legal Studies 17, 151–164.
Polinsky, A.M., Rubinfeld, D.L. (1996). "Optimal awards and penalties when the probability of prevailing varies among plaintiffs". RAND Journal of Economics 27, 269–280.
Polinsky, A.M., Rubinfeld, D.L. (1998). "Does the English Rule discourage low-probability-of-prevailing plaintiffs?". Journal of Legal Studies 27, 519–535.
Polinsky, A.M., Rubinfeld, D.L. (2003). "Aligning the interests of lawyers and clients". American Law and Economics Review 5, 165–187.
Polinsky, A.M., Rubinfeld, D.L. (in preparation). "The deadweight loss of coupon remedies for price overcharges". Journal of Industrial Economics.
Polinsky, A.M., Shavell, S. (1989). "Legal error, litigation, and the incentive to obey the law". Journal of Law, Economics, and Organization 5, 99–108.
Posner, R.A. (1973). "An economic approach to legal procedure and judicial administration". Journal of Legal Studies 2, 399–458.
Posner, R.A. (1992). Economic Analysis of Law, 4th edn. Little, Brown, Boston.
Posner, R.A. (2003). Law and Legal Theory in England and America (Clarendon Law Lecture). Oxford University Press, New York.

Priest, G. (1977). "Common law process and selection of efficient rules". Journal of Legal Studies 6, 65–82.
Priest, G., Klein, B. (1984). "The selection of disputes for litigation". Journal of Legal Studies 13, 1–55.
Rasmusen, E. (1994). "Judicial legitimacy as a repeated game". Journal of Law, Economics, and Organization 10, 63–83.
Reinganum, J.F. (1988). "Plea bargaining and prosecutorial discretion". American Economic Review 78, 713–728.
Reinganum, J. (1989). "The timing of innovation: Research, development, and diffusion". The Handbook of Industrial Organization 1, 849–908.
Reinganum, J.F. (2000). "Sentencing guidelines, judicial discretion, and plea bargaining". RAND Journal of Economics 31, 62–81.
Reinganum, J., Wilde, L. (1986). "Settlement, litigation, and the allocation of litigation costs". RAND Journal of Economics 17, 557–568.
Rickman, N. (1999). "Contingent fees and litigation settlement". International Review of Law and Economics 19, 295–317.
Rosenberg, D., Shavell, S. (1985). "A model in which lawsuits are brought for their nuisance value". International Review of Law and Economics 5, 3–13.
Rubin, P. (1977). "Why is the common law efficient?" Journal of Legal Studies 6, 51–63.
Rubinfeld, D.L., Sappington (1987). "Efficient awards and standards of proof in judicial proceedings". RAND Journal of Economics 18, 308–315.
Rubinfeld, D.L., Scotchmer, S. (1993). "Contingent fees for attorneys: An economic analysis". RAND Journal of Economics 24, 343–356.
Rubinfeld, D.L., Scotchmer, S. (1998). "Contingent fees". In: Newman, P. (Ed.), The New Palgrave Dictionary of Economics and the Law, vol. 1. Macmillan Reference Limited, London, pp. 415–425.
Sanchirico, C.W. (1997). "The burden of proof in civil litigation: A simple model of mechanism design". International Review of Law and Economics 17, 431–447.
Sanchirico, C.W. (2001). "Relying on the information of interested—and potentially dishonest—parties". American Law and Economics Review 3, 320–357.
Santore, R., Viard, A.D. (2001). "Legal fee restrictions, moral hazard, and attorney rights". Journal of Law & Economics 44, 549–572.
Schrag, J. (1999). "Managerial judges: An economic analysis of the judicial management of legal discovery". RAND Journal of Economics 30, 305–323.
Schwartz, E. (1992). "Policy, precedent, and power: A positive theory of supreme-court decision making". Journal of Law, Economics, and Organization 8, 219–252.
Schwartz, W. (1967). "Severance—A means of minimizing the role of burden and expense in determining the outcome of litigation". Vanderbilt Law Review 20, 1197–1214.
Schwartz, M.L., Mitchell, D.J.B. (1970). "An economic analysis of the contingent fee and personal injury litigation". Stanford Law Review 22, 1125–1162.
Schweizer, U. (1989). "Litigation and settlement under two sided incomplete information". Review of Economic Studies 56, 163–177.
Shapiro, C. (2003). "Antitrust limits to patent settlements". RAND Journal of Economics 34, 391–411.
Shavell, S. (1982a). "Suit, settlement and trial: A theoretical analysis under alternative methods for the allocation of legal costs". Journal of Legal Studies 11, 55–82.
Shavell, S. (1982b). "The social versus the private incentive to bring suit in a costly legal system". Journal of Legal Studies 11, 333–339.
Shavell, S. (1989a). "Optimal sanctions and the incentive to provide evidence to legal tribunals". International Review of Law and Economics 9, 3–11.
Shavell, S. (1989b). "The sharing of information prior to settlement or litigation". RAND Journal of Economics 20, 183–195.
Shavell, S. (1993). "Suit versus settlement when parties seek nonmonetary judgments". Journal of Legal Studies 22, 1–14.
Shavell, S. (1994). "Alternative dispute resolution: An economic analysis". Journal of Legal Studies 24, 1–28.

Shavell, S. (1995). "The appeals process as a means of error correction". Journal of Legal Studies 24, 379–426.
Shavell, S. (1996). "Any probability of plaintiff victory at trial is possible". Journal of Legal Studies 25, 493–501.
Shavell, S. (1997). "The fundamental divergence between the private and the social motive to use the legal system". Journal of Legal Studies 26, 575–613.
Shavell, S. (2004). Foundations of Economic Analysis of Law. Belknap Press of Harvard University Press, Cambridge.
Shepherd, G. (1999). "An empirical study of the effects of pretrial discovery". International Review of Law and Economics 19, 245–263.
Shin, H.S. (1994). "The burden of proof in a game of persuasion". Journal of Economic Theory 64, 253–264.
Shin, H.S. (1998). "Adversarial and inquisitorial procedures in arbitration". Rand Journal of Economics 29, 378–405.
Sobel, J. (1985). "Disclosure of evidence and resolution of disputes: Who should bear the burden of proof?" In: Roth, A. (Ed.), Game Theoretic Models of Bargaining. Cambridge University Press, New York.
Sobel, J. (1989). "An analysis of discovery rules". Law and Contemporary Problems 52, 133–159.
Spier, K.E. (1992a). "The dynamics of pretrial negotiation". Review of Economic Studies 59, 93–108.
Spier, K.E. (1992b). "Incomplete contracts and signaling". RAND Journal of Economics 23, 432–443.
Spier, K.E. (1994a). "A note on joint and several liability: Insolvency, settlement, and incentives". Journal of Legal Studies 23, 559–568.
Spier, K.E. (1994b). "Pretrial bargaining and the design of fee-shifting rules". RAND Journal of Economics 25, 197–214.
Spier, K.E. (1994c). "Settlement bargaining and the design of damage awards". Journal of Law, Economics, & Organization 10, 84–95.
Spier, K.E. (1997). "A note on the divergence between the private and social motive to settle under a negligence rule". Journal of Legal Studies 26, 613–623.
Spier, K.E. (2002). "Settlement with multiple plaintiffs: The role of insolvency". Journal of Law, Economics, & Organization 18, 295–323.
Spier, K.E. (2003a). "The use of most-favored-nation clauses in settlement of litigation". The RAND Journal of Economics 34, 78–95.
Spier, K.E. (2003b). "Tied to the mast: Most-favored-nation clauses in settlement contracts". Journal of Legal Studies 32, 91–120.
Spier, K.E., Sykes, A.O. (1998). "Capital structure, priority rules, and the settlement of civil claims". The International Review of Law and Economics 18, 187–200.
Spitzer, M., Talley, E. (2000). "Judicial auditing". Journal of Legal Studies 24, 649–683.
Spurr, S.J. (1991). "An economic analysis of collateral estoppel". International Review of Law and Economics 11, 47–61.
Statistical Abstract of the United States (2001). U.S. Department of Commerce, Economics and Statistics Administration, US Census Bureau, Washington, D.C.
Sykes, A.O. (1994). "Bad faith' refusal to settle by liability insurers: Some implications of the judgment-proof problem". Journal of Legal Studies 23, 77–110.
The 2001 Survey of Law Firm Economics (2001). Altman Weil, Inc., Altman Weil Publications, Newton Sq., Pa.
Thomason, T. (1991). "Are attorneys paid what they are worth? Contingency fees and the settlement process". Journal of Legal Studies 20, 187–233.
Tullock, G. (1980). Trials on Trial. Columbia University Press, New York.
Waldfogel, J. (1995). "The selection hypothesis and the relationship between trial and plaintiff victory". Journal of Political Economy 103, 229–260.
Waldfogel, J. (1998). "Selection of cases for trial". In: Newman, P. (Ed.), The New Palgrave Dictionary of Economics and the Law, vol. 3. Macmillan Reference Limited, London, pp. 419–424.
Watanabe, Y. (2005). Learning and Bargaining in Dispute Resolution: Theory and Evidence from Medical Malpractice Litigation. Northwestern University mimeo.

White, M. (2002). "Explaining the flood of asbestos litigation: Consolidation, bifurcation, and bouquet trials", National Bureau of Economic Research Working Paper, W9362.
Wittman, D. (2003). "Lay juries, professional arbitrators, and the arbitrator selection hypothesis". American Law and Economics Review 5, 61–93.
Yildiz, M. (2003). "Bargaining without a common prior—an immediate agreement theorem". Econometrica 71, 793–811.
Yildiz, M. (2004). "Waiting to persuade". Quarterly Journal of Economics 119, 223–248.
Yoon, A. (2004). "Mandatory arbitration and civil litigation: An empirical study of medical malpractice litigation in the west". American Law and Economics Review 6, 95–134.