JEL: K41

LITIGATION AND SETTLEMENT[*]

Bruce L. Hay[**]
Kathryn E. Spier[***]

Abstract

Perhaps the most widely discussed topic in the economics of legal procedure is the relation between litigation and settlement. This paper offers a critical introduction to the voluminous economic literature on this topic. This literature considers questions such as why the parties to a dispute choose to go to court, rather than settling the matter privately; what determines the likelihood that a case will settle, and the terms on which it is settled; how various legal instruments or rules influence parties' decision to settle or go to court; and how the choice between settlement and litigation affects social welfare. These issues have turned out to be surprisingly complex, and continue to receive extensive attention as commentators employ increasingly refined models to examine the many nuances of the problem. The present paper attempts to set forth in a concise and informal manner the major results of this work and some directions for future research.

[*]Forthcoming in the *New Palgrave Dictionary of Economics and the Law*. The authors would like to thanks Steven Shavell for his comments and suggestions.
[**] Harvard Law School.
[***] Northwestern University.

Perhaps the most widely discussed topic in the economics of legal procedure is the relation between litigation and settlement. A rich literature has developed on such questions as why the parties to a dispute choose to go to court, rather than settling the matter privately; what determines the likelihood that a case will settle, and the terms on which it is settled; how various legal instruments or rules influence parties' decision to settle or go to court; and how the choice between settlement and litigation affects social welfare. These issues have turned out to be surprisingly complex, and continue to receive extensive attention as commentators employ increasingly refined models to examine the many nuances of the problem.

The essay to follow will selectively review some of the major results of this literature, and discuss promising directions for future research. The essay is organized as follows. The first part examines the positive theory of litigation and settlement, focusing on the parties' equilibrium behavior. The second part considers the relation between private settlement behavior and social welfare. The third part explores how the legal system may affect settlement behavior — in particular, the likelihood or terms of settlement — by adjusting the rules of legal procedure in various ways, for example by manipulating the rewards and costs of litigation or by employing alternative dispute resolution techniques.

The Positive Theory Of Litigation And Settlement

The Choice Between Litigation and Settlement of Disputes. Consider a lawsuit in which the plaintiff seeks monetary damages from the defendant.

Normally, the parties to the dispute have the choice of either having the case resolved in a court trial or instead settling the matter between themselves. From the standpoint of the parties, a private settlement has the advantage of saving them the costs of litigating the case, which may be quite substantial. Litigation may entail extensive out-of-pocket attorneys fees, and in addition may be highly time-consuming. Any agreement to settle the case will generate a surplus for the parties — in the form of saved litigation costs — that the parties can divide between themselves. (To be sure, for some parties the prospect of a trial is desirable, perhaps as a forum for self-expression or publicity.) In addition, to the extent the outcome of a trial is uncertain, litigation is risky, making settlement still more attractive if the parties are risk averse.

For these reasons, one would expect that there is generally some range of settlement amounts leaving both parties better off than they would be in going to trial. For example, suppose that if the case goes to trial, the probability that the defendant will be held liable is 75 percent, and that if he is held liable the likely damage award will be $100,000. The plaintiff's expected recovery from trial — we will call this the "expected judgment" — is thus .75 × $100,000 = $75,000. Suppose, in addition, that if the case goes to trial, each party will incur litigation costs of $10,000. Then if the parties are risk neutral, any settlement figure between $65,000 and $85,000 makes both parties better off than going to trial. The plaintiff's net expected recovery at trial is given by $75,000 - $10,000 = $65,000; the defendant's net expected loss at trial is $75,000 + $10,000 = $85,000. Hence any settlement amount between these figures leaves both parties wealthier,

in expected terms, than they would be going to trial. (Of course, if the item at stake in the litigation is indivisible, as in child custody litigation, no such compromise may be possible; see Shavell (1993); however, compromise will normally be possible in disputes over money.) Moreover, if the parties are risk averse, the range of settlements that make both parties better off than going to trial will be even wider. For example, the risk averse plaintiff may be willing to settle for any amount exceeding $50,000; the risk averse defendant may be willing to settle for any amount below $100,000.

This simple logic presumably explains why the vast majority of disputes settle out of court. Of filed lawsuits in America, over 95 percent settle before trial; and of course there are many more disputes that settle without suit being filed in the first place. See Shavell (1997); Kritzer (1991). For students of settlement behavior, examination of this logic naturally raises two types of question. First, why do cases ever go to trial, given the benefits the parties can reap from settlement? What factors determine which cases settle and which go to trial? Second, if the parties do settle, what amount do they settle for? Assuming there is a range of feasible settlement amounts, what determines the amount on this range that is actually agreed to by the parties?

Determinants of the Choice Between Litigation and Settlement. To settle, the parties must identify some settlement amount that makes them both better off, in their view, than going to trial; a party will not agree to an amount that makes her worse off than going to trial. An important determinant of settlement, therefore, is what the parties expect to gain or lose from litigating the case. The

better a plaintiff expects to do at trial — the more likely she believes she is to win, or the greater the amount she expects to recover in the event of victory — the greater the settlement amount she will have to be paid as a condition for agreeing to settle. Likewise, the better the defendant expects to do at trial, the less she will be willing to pay to settle the case.

Accordingly, divergent party beliefs about the likely outcome of trial may prevent the parties from settling. In particular, if both parties are sufficiently optimistic about their prospects in court — roughly speaking, if each expects to prevail in the litigation — then there may be no mutually acceptable settlement amount. Returning to our earlier numerical example, suppose that the plaintiff believes the probability of a verdict against the defendant 75 percent, while the defendant believes that this probability is only 25 percent. Then the plaintiff will not accept anything less than $65,000 to settle the case; yet the defendant will not pay anything more than $35,000. (For clarity, we assume in this and all future numerical examples that the parties are risk neutral.) Hence settlement is infeasible. Though each party would benefit from settling the case, thereby saving litigation costs, there is no mutually agreeable settlement amount.

This insight concerning party expectations was at the heart of the earliest models of the choice between settlement and litigation. See Landes (1971); Posner (1973); Gould (1973); see also Shavell (1982); Danzon and Lillard (1983); Priest and Klein (1984). The existence of divergent party expectations concerning trial remains the most influential account of why cases may fail to settle. The immediate problem it raises, however, is *why* the parties may have divergent

beliefs. More precisely, the question is why a party would maintain his own optimistic assessment of his chances at trial, once he becomes aware of his opponent's very different assessment. Rational choice theory tells us that each party should revise downward his optimistic assessment of his own chances once he discovers — during the course of bargaining — how optimistic his opponent is. See Aumann (1976). As a result, we should expect the parties' beliefs to converge during the negotiation process, eventually making settlement possible. (Perhaps this is especially true if lawyers, being experts in the law, are involved.) The task facing students of the settlement has been to explain why such a convergence does not occur.

The most widely accepted economic explanation, stemming from Bebchuk (1984), is that the parties may have different information about the likely outcome of the case. For example, the plaintiff may be privately informed about the severity of her damages; or the defendant may be privately informed about whether he behaved negligently. A large body of literature has explored the equilibrium outcomes of settlement bargaining under such conditions of asymmetric information. The central finding of this literature has been that the presence of asymmetric information yields a positive probability that a case will fail to settle. Put otherwise, given a set of cases in which the parties are asymmetrically informed about the probable outcome of trial, it is likely that a positive fraction of these cases will fail to settle. For surveys of the literature building on this result, see Cooter and Rubinfeld (1989); Kennan and Wilson (1991); Kritzer (1991).

The intuition behind this result can be captured by the following simple model. Assume that a plaintiff's claim can be either "strong" or "weak"; there are an equal number of claims of each type. If the claim is strong, the expected judgment at trial is $100,000; if the claim is weak, the expected judgment is $50,000; each side has trial costs of $10,000. Assume that the plaintiff knows whether her claim is strong or weak; the defendant, however, knows only the distribution of claim types, without knowing the type of claim involved in her own case. Finally, assume that settlement bargaining consists of a single take-it-or-leave-it offer by the defendant.

Using the techniques of game theory, it is straightforward to show that, in this model, cases involving a strong claim will fail to settle. A weak claim holder will accept an offer of at least $40,000; but a strong claim holder will refuse any offer of less than $90,000. Hence, because the defendant does not know what type of plaintiff she faces, the only offer that will guarantee a settlement is an offer of $90,000. However, it is not in the defendant's interest to offer that much; she is better off making an offer of $40,000. The expected cost to the defendant of making the higher offer is $90,000 (because the plaintiff will definitely accept); the expected cost of making the lower offer is (.5 × $110,000) + (.5 × $40,000) = $75,000 (because the plaintiff will accept only if she has a weak claim). In equilibrium, therefore, the defendant will offer $40,000; if the case involves a strong claim, the offer will be rejected — leading to litigation.

Despite its simplicity, the implications of this model are quite robust. The central result — that under asymmetric information some cases will fail to settle

— obtains even in more complex bargaining environments, in which the parties make offers and counter-offers. In such an environment, the defendant may be able to draw inferences — from the plaintiff's bargaining behavior — about the nature of the privately-informed plaintiff's claim. This makes it possible that some cases involving a strong claim will settle. Nonetheless, there is no equilibrium in which *all* cases settle. Essentially, the reason is that weak claim holders will rationally attempt to pass themselves off as strong claim holders. The defendant, knowing this, will rationally refuse the settlement demands of a positive fraction of plaintiffs who state they are strong claim holders. Yet some of the plaintiffs who are turned away will, in fact, be strong claim holders. No matter how long the series of offers and counter-offers, some cases will, in equilibrium, fail to settle. For analysis of this point, see Spier (1994a).

This model takes as given the existence of private information about the claim. A natural question, however, is why the privately informed party would not voluntarily reveal the information to her adversary. Notice that in the model, it is the relatively high-value claims that fail to settle. Holders of these claims have a natural incentive to disclose the value of their claims, so that the defendant will make a settlement offer reflecting that value. Disclosure of this sort makes settlement possible in cases that would not otherwise settle. See Shavell (1989). However, in some instances it may not be feasible for the privately informed party to credibly disclose her information. (To illustrate, the plaintiff's losses may not be provable until after experts have appraised them.) In other instances, the costs of disclosure may be too high to make it worthwhile — for example, when the

private information may be exploited for its surprise value at trial. For these reasons, voluntary disclosure is unlikely to eliminate all of informational asymmetries that inhibit settlement.

Another difficulty with the asymmetric information explanation of litigation is the existence of "discovery" rules, which compel parties to disclose before trial the evidence in their possession. See Shavell (1989); Sobel (1989). Assuming these rules are effectively enforced, the parties will be symmetrically informed about the evidence that will be introduced at trial. However, the value of the plaintiff's claim is generally a function not only of the content of the evidence, but also of a party's investment in preparation of the case; and information about a party's preparation is generally exempt from discovery. This exemption from discovery suffices to explain why cases fail to settle even with discovery rules in place. For in equilibrium, a party will be uncertain about what preparation measures her opponent has taken; and this uncertainty will lead to litigation in some cases.

This point may be seen with the following modification of the example just used. Assume that the plaintiff can make a "light" investment in pretrial case preparation, yielding an expected award at trial of $50,000, or can (for an additional $10,000) make a "heavy" investment, yielding an expected award of $100,000. Assume, in addition, that her investment occurs during the discovery process, and is not observable by the defendant. Assume, finally, that settlement bargaining consists of a single offer by the defendant. In such a setting, it is readily verified that the parties will pursue mixed strategies in equilibrium: the

plaintiff will sometimes make a light investment and sometimes a heavy one; the defendant will sometimes make a low offer and sometimes a high one. Litigation will occur when the defendant makes a low offer to a plaintiff who has made a heavy investment. Hence, even with discovery rules compelling disclosure of the evidence, we should expect to find cases failing to settle in equilibrium. See Hay (1995); see also Shavell (1989) for further insights on this issue.

The theoretical literature has developed additional predictions concerning the choice between litigation and settlement under asymmetric information. Some work has emphasized comparative statics, investigating the factors in a case that bear on the likelihood of settlement. See, for example, Bebchuk (1984). The main results of this work are as follows. First, settlement becomes more likely when the trial costs are larger. Essentially, this is because the incentive to settle increase as trial costs go up; it is thus less likely that disagreement about the outcome of trial will prevent the parties from finding a settlement amount that makes both sides better off than going to trial. Second, the settlement becomes more likely when the information of the litigants is more closely aligned. The less private information a party has about the expected outcome of the case, the more likely it is that the parties will agree on the expected outcome, making it easier to find mutually acceptable settlement terms. Third, in cases where the parties are mutually optimistic about their chances of trial, the likelihood of settlement decreases as the amount at stake in the case increases (all else being equal). Essentially, the reason is that high stakes magnify the optimism: if the plaintiff is sure she will win a lot, settlement will be more difficult than if the

plaintiff thinks she will win only a modest amount at trial. Finally, the odds of settlement increase if one or both parties are risk averse.

Other work in the asymmetric information literature has examined the dynamics of pretrial negotiations, explicitly modeling the parties' bargaining strategies over time. When there is asymmetric information about the outcome at trial, then a strong deadline effect emerges, in which settlement is often delayed until the last possible moment. Intuitively, the willingness to settle the lawsuit early would be interpreted as a sign of weakness, and would consequently put the litigant at a strategic disadvantage. Waiting until the last moment to settle makes the litigant seem tougher, and is a strategic advantage (even though negotiations sometimes break down.) This finding may explain the observed phenomenon that many cases settle "on the courthouse steps," immediately before trial. See Spier (1992), which also shows that there may be a "U-shaped" pattern of settlement, in which the likelihood of settlement is greater at the beginning and end of the pretrial period than in the middle.

Apart from explanations emphasizing asymmetric information about the value of the claim, three basic alternative theories have been advanced to explain why cases may fail to settle. One is that the parties may fail to agree on a division of the surplus from settling. Suppose, for example, that the expected judgment in a case is $100,000, and that each party faces trial costs of $10,000; and suppose these figures are common knowledge to the parties. Any settlement amount between $90,000 and $110,000 will make both parties better off than they would be by going to trial. Yet to settle, they must agree on an amount (equivalently, on

a procedure for selecting an amount) on that range. In effect, their problem is to agree on a distribution of the $20,000 surplus (in saved litigation costs) they will jointly capture by avoiding trial. Each party has an incentive to seek the lion's share of this surplus: the defendant will rationally press for a settlement amount of about $90,000; the plaintiff will counter with a proposal to settle for an amount closer to $110,000. Each party may dig in her heels, expecting the other to relent; if each misjudges the other's willingness to compromise, they may fail to settle. See Cooter, Marks and Mnookin (1982).

This explanation is most plausible when understood as a variant of the asymmetric information theory. A party's bargaining strategy may in part be determined by factors that are private information to her. Suppose, in this example, that a party offers to settle for a given amount. Her willingness to "stick" to this position and turn away counter-offers may depend on factors such as her risk preferences (see Farmer and Pecorino (1994)) or her reputational objectives. If these are known only to her, then her opponent will in some cases misjudge the her bargaining strategy, perhaps yielding a bargaining impasse. This idea can easily be incorporated into the model presented earlier: we might suppose, for example, that there are "tough" parties and "weak" parties, distinguished by their willingness to adhere to an aggressive settlement demand. If a party's type is unobservable to her adversary, then we might expect negotiations to fail in cases in which two tough parties are pitted against each other.

Another theory focuses on the diverging interests of lawyer and client. This explanation rests on the simple observation that lawyers are not necessarily benefited by the quick resolution of disputes. This is most obvious in the case of a lawyer who is paid solely on the time he has spent on handling the dispute, for example according to an hourly fee; for such a lawyer, protracted litigation is more profitable than an early settlement. (The prospect of free advertising through a publicized trial may furnish a similar incentive.) To the extent this lawyer can act on this incentive, litigation may occur even though it would be in the client's interest to settle. Obviously, the likelihood of this occurrence depends on the degree of control the lawyer exerts over the settlement decision; even if the client formally controls the decision, the lawyer may have substantial influence over the client, for example by regulating the information available to the client. In addition, the lawyer's incentives regarding settlement depend on her fee arrangement with the client; some fee structures are more likely than others to discourage the lawyer from seeking an early settlement. See Miller (1987). Lawyers are not necessarily barriers to settlement in any event; sometimes their participation may facilitate settlement by furnishing information to their clients, see Menkel-Meadow (1984), or by providing credibility to party communications in negotiation, see Gilson and Mnookin (1994).

Finally, another theory relaxes the conventional assumptions concerning the rationality of economic actors. This account points to experimental evidence suggesting that litigants' make "self serving" estimates of their chances of success in court; all else being equal, a party tends to give herself better odds than her

opponent gives her. The implication is that even if the parties are symmetrically informed, each may be relatively optimistic about winning the case, so that there is no mutually acceptable settlement amount. See Loewenstein et al. (1993); Mnookin (1993). One question this account raises is whether parties would take steps to correct for biases of this sort; one way of doing so might be to engage a lawyer who is skilled at making unbiased assessments (though, as we have seen, delegation to lawyers raises problems of its own). Development of this theory is a promising area for future research.

Terms of Settlement. What determines the terms on which the parties agree to settle a dispute? As the model sketched above suggests, a major determinant is the expected judgment at trial. Neither party will agree to a settlement unless its terms leave her at least as well off as she would expect to be if there were no settlement. As an initial approximation, therefore, we might anticipate that cases will generally settle for an amount roughly equal to the expected judgment at trial. That, however, is at best a very crude approximation; several other factors may have a significant bearing on the terms of settlement. Four such factors are worth highlighting here: (1) the costs of litigation to the parties; (2) the parties' risk preferences; (3) the parties' relative bargaining power in dividing the surplus from settlement; and (4) the allocation between lawyer and client of control over negotiations.

Consider litigation costs. As we have seen, avoiding these costs is a major motivation for settling. Yet even if they are not directly incurred – in the sense that the parties successfully avoid litigation – these costs may nonetheless make

themselves felt in the terms of settlement. For example, suppose that the expected judgment (in both parties' estimate) is $50,000, and that each faces litigation costs of $10,000. In principle, the case could settle for as little as $40,000 (the plaintiff's net expected recovery from trial), so that the plaintiff indirectly "bears" the costs of litigation – in that she is scarcely better off than if she had in fact gone to trial. Similarly, the case could settle for as much as $60,000, leaving the defendant in much the same position as she would have been in if she had incurred the costs of litigation. Hence, the costs of litigation may push the settlement amount well above or well below the value of the expected judgment.

Moreover, it is straightforward to show that, all else being equal, the greater the plaintiff's litigation costs, the lower the likely settlement amount. Too see this, assume that the parties settle at a point *p* along the settlement range, where α is defined as a fraction of the distance between the plaintiff's net expected recovery or and the defendant's net expected loss. (If the parties settle at the midpoint of this range, then α is simply .5; but in principle *p* may have any value from zero to one.) Assuming, reasonably, that α is unaffected by the magnitude of a party's litigation costs, then raising the plaintiff's costs has the effect of lowering the likely settlement amount. By the same token, all else being equal, the greater the defendant's litigation costs, the higher the likely settlement amount.

A much-discussed topic concerning the effect of litigation costs on settlement terms is the settlement of suits that are unprofitable for the plaintiff to take to trial. For many claims, the costs of going to trial will exceed the expected

judgment, either because the probability of a favorable verdict is low (as in the case of a "frivolous" claim), or because the plaintiff's damages are small. In such cases, the plaintiff appears unable to make a credible threat that she will litigate the case if the defendant does not pay a satisfactory settlement. Yet in the absence of a credible threat of litigation, the defendant has no reason to pay anything to settle the case. Hence, in such cases we might expect the plaintiff to come away empty-handed, even though the expected judgment would be positive if the case had gone to trial.

Commentators have, however, pointed out that depending on the manner in which litigation costs are incurred over time, it may be possible for the plaintiff to successfully extract a positive settlement amount from the defendant, even though both parties know that going to trial would be unprofitable for the plaintiff. Bebchuk (1996) points out that this result is possible when litigation expenditures and settlement negotiations occur over multiple periods in a given case. Suppose that, in the *nth* period, most of the plaintiff's litigation costs are sunk, so that she can at that point credibly threaten to go to trial. Then the defendant will rationally pay to settle the case. Yet if both parties know this, the case will not in fact proceed to the *nth* period: the defendant will offer to settle (at the latest) in period *n*-1. By the logic of backward induction, it can be shown that if the parties' litigation costs are sufficiently divisible over time, the case will settle for a positive amount at the outset – even though the plaintiff's overall costs of going to trial would exceed the expected judgment.

Another theory focuses on the relative timing of the plaintiff's and defendant's litigation expenditures. Suppose that filing suit is comparatively inexpensive for the plaintiff, and that the filing of suit forces the defendant to incur response costs. Then the plaintiff will, by simply filing (or threatening to file) suit, induce the defendant to settle for some amount up to the value of her response costs. See Rosenberg and Shavell (1985). An important implication of these studies, then, is that the outcome of settlement negotiations may be highly sensitive to litigation costs – not only their magnitude, but how they are incurred over time.

Now consider the parties' risk preferences. In nearly all cases, the outcome of trial is uncertain; the uncertainty may be over whether the plaintiff will win, how much she will recover in the event she wins, or both. For risk averse parties, the effect of this uncertainty is similar to making trial more costly. All else being equal, the more risk averse the plaintiff, the lower the settlement amount necessary to make settlement more attractive than trial; likewise, the more risk averse the defendant, the more she will be prepared to pay to avoid trial. Hence, as with litigation costs, the existence of party risk aversion may push the amount of settlement above or below the expected judgment; and all else being equal, the more risk averse a party is, the more unfavorable the likely settlement will be for that party.

Now consider the issue of bargaining power. In most cases, the settlement range – the set of amounts that make both parties better off than going to trial – will typically have many points on it. Our last example, in which the expected

judgment is $50,000 and each party faces trial costs of $10,000, had a settlement range stretching from $40,000 to $60,000. To settle, the parties must choose a point α on this range. (This is equivalent to saying they must agree on a division of the surplus from settling.) The value of α will be determined by the parties' relative bargaining power, which can be roughly equated with the ability to credibly make a "final offer" – that is, to credibly state that the offeror will bargain no further, and will take the case to trial if the offer is not accepted.

The nature of such bargaining power has not received systematic attention in the economic literature on settlement; Genn (1988) and Kritzer (1991) contain useful informal treatments. One intuitively plausible source of bargaining power is reputation: if, in our example, the defendant expects to be involved in a series of similar lawsuits in the future, then it may be in his interest to walk away from the table if her offer of (say) $41,000 is not accepted, because this may establish a reputation for "toughness" that will induce plaintiffs to make concessions in future cases. The present plaintiff, knowing this, will (assuming she has no comparable reputational incentive) rationally accept this relatively low offer. Analytically, this account is equivalent to saying that a party's reputational benefits from going to trial offset to some extent her trial costs, so that the settlement range is in effect shifted. Notice, however, that the reputational benefits are a function of what happens in the bargaining process itself; the benefits flow from the fact that the defendant walked away from the bargaining table.

Finally, consider the allocation of control between lawyer and client. Under commonly used litigation finance arrangements, the interests of lawyer and client may diverge, so that the settlement amount a given side agrees to depends on who on that side controls the decision. To take a simple example, suppose the plaintiff's lawyer will be paid pursuant to a contingent fee that gives him one-third of the amount recovered from the defendant. Assume that the expected judgment is $50,000, and that trial will cost the lawyer $10,000 in time and expense. From the client's standpoint, the minimum acceptable settlement amount is $50,000. From the lawyer's standpoint, however, the minimum acceptable settlement amount is $20,000. (If the case goes to trial, the lawyer gets (1/3 H $50,000)-$10,000 = $6,666; a settlement of $20,000 would yield her the same amount.) The reason for this difference is that the lawyer in this example bears all the costs of going to trial. Hence, if the lawyer controls the decision, the "bottom" of the settlement range – the minimum amount acceptable to the plaintiff's side – will be lower than if the client controls the decision. See Miller (1987).

There may, however, be countervailing effects. For example, the lawyer may be less risk averse than the client. The claim may be the client's only asset; yet it may be part of a portfolio of claims for the lawyer. Likewise, the lawyer may have greater bargaining power against the defendant than does the client. The lawyer, expecting to be involved in future cases, will be concerned about her reputation in future settlement negotiations; the client has less at stake in this regard. Hence, lawyer control may in some instances lead to greater, not smaller,

settlement recoveries than would be obtained under client control of the negotiations. Similar complexities – arising out of the crosscutting incentives created by fee arrangements, reputational concerns, and risk preferences – can be expected on the defendant's side. These complexities make it difficult to generalize about how the allocation of control will influence the terms of settlement, beyond observing that the allocation is likely to be important. For further discussion, see Kritzer (1991); Kritzer (1990).

Settlement Decisions and Social Welfare

To what extent to parties' settlement decisions coincide with the interests of society as a whole? Is it the case that settlement behavior that is privately efficient (in the sense of benefiting the parties) is also necessarily socially efficient? Or does society have an independent stake in the parties' settlement decision, which the parties may fail to take into account?

Settlement decisions produce at least two types of costs and benefits that are not internalized by the parties. The first is the costs of operating the court system in cases that fail to settle. The parties' time and litigation expenses typically constitute only a fraction of the costs incurred in adjudicating their dispute; the rest of these costs are borne by the public. The second is the deterrent effects on primary behavior (such as precautions against accidents) that may be generated by the anticipated resolution of a case. Because disputes typically arise after the primary behavior has occurred, the parties have no incentive to take such effects into account at the time they decide whether to settle and for how much.

As a result of these external effects, there is no necessary intersection between the parties' welfare and social welfare in the settlement decision.

A simple example may illustrate the point. Consider a hypothetical accident scenario in which the prospective injurer can reduce the probability of an accident by taking some specified precaution. The average injury causes losses of $100,000; if the defendant expects to pay at least that amount in the event of an injury, she will, we may assume, take the precaution. Finally, assume that if an injury occurs and the case is taken to trial, the defendant will be compelled to compensate the plaintiff for her full actual losses.

Social welfare is maximized in this model if the parties settle accident cases rather than taking them to trial, provided that the parties settle for the "right amount" – more precisely, provided that anticipated settlements are large enough to induce prospective injurers to take precautions. To see this, observe that if settlements are sufficiently great – for example, if all cases simply settle for $100,000 – then settlement is socially preferable to trial, since a trial will generate litigation costs but have no effect on the accident rate. Yet if settlement amounts are low – for example, if all cases settle for less than $100,000 – then social welfare would be improved if the settlement amount were increased, since this would reduce the accident rate.

With these observations, it is possible to observe how the parties' settlement decisions may diverge from what is socially desirable in this model. On one hand, the parties may fail to settle at all. Suppose, for example, that the plaintiff is privately informed about the actual losses she sustained in the accident.

Such private information about the likely outcome of trial will, as we have seen, prevent settlement in a positive fraction of cases. In essence, trial will occur because the parties disagree about the plaintiff's actual losses. This result is socially wasteful, in the sense that trial costs could be saved without affecting the accident rate if all cases settled for an amount at least equal to the *average* plaintiff's losses. See Kaplow (1994); Shavell (1997).

On the other hand, even if the parties successfully settle the case, they may do so on socially undesirable terms. Suppose, for example, that going to trial is very costly to the plaintiff, who for that reason is willing to settle for much less than the expected judgment at trial. If the plaintiff acts on that incentive (perhaps because the defendant has greater bargaining power), cases may settle for well under the amount of the plaintiff's losses – so that the average settlement amount is well under $100,000. Prospective injurers, knowing this to be true, will have no incentive to take precautions against harm. This result too is socially wasteful, in that the accident rate could be lowered without affecting litigation costs if the settlement amount were raised. If for some reason a higher settlement amount cannot be achieved, it may indeed be socially preferable for cases to go to trial rather than settling. See Hay (1994) and Spier (1997) for exploration of this idea.

This simple model highlights two basic points. First, the socially optimal result will normally involve a settlement rather than a trial, in that there is in principle *some* settlement amount such that, if the parties settle for it, trial costs will be saved without raising any other social cost. (The exception would be in cases where trial itself has some social value, such as setting a precedent to help

decide future cases.) Yet as we have seen, for a variety of reasons rational behavior on the part of the parties may prevent them from settling. Hence there is a rough prima facie argument for some governmental encouragement of settlement, if the correct terms can be arrived at.

Second, some settlement amounts are socially preferable to others, in that certain settlements will have undesirable effects on primary behavior. Settlements that are too low may underdeter accidents; settlements that are too high may discourage socially beneficial activities. Yet because the dispute comes after the primary behavior has occurred, parties will not take such behavioral effects into account when deciding on a settlement amount. Hence, there is also a prima facie argument for some governmental intervention into the choice among settlement amounts. See Shavell (1997); Polinsky and Rubinfeld (1988).

Effects of Legal Rules on the Settlement of Litigation

Policymakers frequently attempt to design legal rules to influence, in one way or another, litigants' choice between litigation and settlement. For example, one common motivation of policymakers is to increase the settlement rate by adopting rules that enhance litigants' incentives to settle. Another common motivation is to prevent litigants from compelling their adversaries to accept settlements that do not reflect the underlying merit of the claim – as when, for example, a plaintiff with a frivolous claim extracts a substantial sum from the defendant by threatening the defendant with costly litigation. A natural line of inquiry therefore is how different legal rules affect litigants' decision to settle.

In a sense, virtually any legal rule will have some effect on the settlement of litigation. It is difficult to imagine a legal rule that does not have some effect on either the costs or expected outcome of litigation, which as we have seen are major determinants of the likelihood and terms of settlement. Our focus, however, will be on legal rules that are used explicitly as (or are commonly understood to be) instruments for intervening in the settlement process. These include fee shifting and related financial devices; rules governing the litigation process that indirectly influence settlement; and alternative dispute resolution procedures such as arbitration and mediation that are designed directly to affect parties' settlement decisions.

Fee Shifting and Related Instruments. One subject that has received extensive economic analysis is how rules allocating the parties' litigation costs affect settlement behavior. Two aspects of this issue, in particular, have attracted considerable attention. The first is the choice between the so-called "American" and "English" rules for allocating legal expense: under the former rule, each party bears her own litigation costs, while under the latter the loser pays the winner's litigation costs. The second is the design of rules – sometimes called "offer-of settlement" rules – that base the allocation of costs upon the settlement offers made before trial. Under an offer-of-settlement rule, if a litigant rejects a settlement offer prior to the trial and later receives a less favorable judgment at trial, then she must compensate her opponent for some portion of her litigation costs. The major lesson emerging from economic analyses of these rules is that

their effect on settlement behavior is highly sensitive to the litigation context, making generalizations difficult.

Let us begin with the choice between the American and English rules. Suppose, first, that the parties share common beliefs about the probability of the plaintiff prevailing at trial. In this case, the size of the settlement range is the same under the two rules. This is easiest to see if we imagine that one party (say, the plaintiff) is certain to win at trial. The plaintiff's minimum demand (the least she will accept to settle) is higher under the English rule than under the American rule, because she expects her litigation costs to be borne by the defendant; but the defendant's maximum offer (the most she will pay to settle) goes up by the same amount, because she expects to bear the plaintiff's costs. This is true even if the parties do not agree on the likely damage award in the event of a plaintiff victory. Thus, when the parties agree on the likelihood of a plaintiff victory we should expect the frequency of settlement to be the same under the two rules.

Now suppose, second, that the parties do not agree on the plaintiff's probability of winning at trial. In particular, suppose that they are mutually optimistic – each thinks she is likely to be the winner. In this scenario, more cases will settle under the American Rule than under the English Rule. In essence, the reason is that the English rules raises the stakes in the case – since not only the plaintiff's damages, but also the parties' litigation costs, are up for grabs at trial – which, as we have seen, has the effect of making settlement less likely. An intuitive way of seeing the point is as follows. If both litigants are optimistic about their prospects at trial, then adopting the English rule will tend to decrease

each litigant's subjective assessment of his own expected costs (and raises the assessment of the opponent's expected costs). Litigation looks like a "better deal" relative to settlement, the settlement range narrows and settlement becomes less likely. See Shavell (1982). In contrast, when the parties are mutually pessimistic, the choice between the American and English rules is unlikely to affect the chances of settlement, since there will be a positive settlement range under either rule.

These basic results are confirmed in more refined game-theoretic models that explicitly take account of the parties' information and bargaining strategies. On the one hand, when one party has private information about the plaintiff's probability of success, the English rule decreases the likelihood of settlement. See, for example, Bebchuk (1984). Intuitively, when there is asymmetric information about the probability of winning, then the parties disagree about their expected legal costs in addition to the expected award at trial. The English rule thus magnifies the effect of asymmetric information, making settlement more difficult. On the other hand, if the parties have symmetric information about the probability of a plaintiff victory at trial, then the English rule has no effect on the likelihood of settlement, whether or not the parties agree on what the damage award will be. See Reinganum and Wilde (1986).

It would appear, then, that the English rule, relative to the American rule, either has no effect on the likelihood of settlement (when the parties agree on the likelihood of a plaintiff victory) or reduces it (when they assess that probability differently, perhaps because of asymmetric information). This generalization,

however, is subject to an interesting qualification. If litigation costs are choice variables for the parties, then the English rule may increase the chances of settlement. The reason is that under the English rule, each litigant view her legal expenses as coming, to some extent, out her opponent's pocket, this is particularly true when the parties are mutually optimistic. Each will therefore spend more on the litigation than she would otherwise. The upshot is that the settlement range will widen, making settlement more likely. This effect may offset the settlement-inhibiting features of the English rule described above. See Braeutigam, Owen and Panzar (1984); Katz (1987); Hause (1989). Hence, the English rule's ultimate effect on the likelihood of settlement is ambiguous.

Regarding the terms of settlement, the effects of the English rule in relation to the American rule are also unclear. On one hand, if the parties agree on the plaintiff's chances of winning at trial, then the English rule has the effect – all else being equal — of favoring the party who has the greater chance of success at trial. If, for example, the plaintiff is highly likely to win at trial, the English rule will increase both the plaintiff's minimum demand and the defendant's maximum offer – with the result that the likely settlement amount is greater than under the American rule. Similarly, if the defendant is highly likely to win at trial, then English rule will push the settlement range downward, making the likely settlement amount lower.

On the other hand, if the parties do not agree on the plaintiff's chances of success, then the English rule's impact on the terms of settlement is less obvious. In cases of mutual optimism (where each party believes she is likely to win), the

English rule frequently increases the plaintiff's minimum demand while at the same time decreasing the defendant's maximum offer. Which effect predominates – whether the midpoint of the settlement range moves up or down — will vary across cases. Similarly, in cases of mutual pessimism, the English rule will frequently decrease the plaintiff's minimum demand and at the same time increase the defendant's maximum offer. Whether this translates into higher or lower settlement amounts is impossible to say a priori.

Let us turn now from the choice between the English and American rules to the design of offer-of-settlement rules. These rules make the allocation of costs dependent upon the settlement offers made before the trial. For example, if litigant rejects a settlement proposal prior to the trial and later receives a less favorable judgment at trial (as compared to the settlement proposal), then she must compensate his opponent for some portion of her costs. (The portion varies in different settings and is, of course, a choice variable for the policymaker.) The purpose of such rules is to encourage settlement, and in particular to encourage settlements that are roughly commensurate with the expected judgment at trial. How effective are they?

An early study revealed the pitfalls of assuming that inducements of this sort will serve their intended goals. See Miller (1986). Consider a one-sided offer-of-settlement rule (widely used in America), in which the plaintiff can be held responsible for the defendant's costs, but not vice versa. Such a rule raises the plaintiff's costs of going to trial, and thus lowers her minimum demand. Yet it also lowers the defendant's cost of going to trial, and thus lowers his maximum

offer. Combined, these effects may simply shift the settlement range downward, without necessarily making it wider. Thus, the main impact of the rule may be simply to reduce the settlement amount in cases that settle, without increasing the settlement rate. Moreover, there is no necessary reason to expect that the settlement amount will be closer to the expected judgment than it would have been in the absence of the rule.

More recent work refines this analysis by explicitly considering the parties' information and bargaining strategies under a one-sided offer-of-settlement rule. Under conditions of asymmetric information, the effects of such a rule on the settlement rate are complex. In cases where the informational asymmetry concerns the amount the plaintiff will recover in the event she prevails at trial, then the rule tends to encourage the defendant to make a higher settlement offer than she would otherwise make. As a result, in such cases the rule tends to *increase* the odds of settlement. In contrast, when the informational asymmetry concerns the plaintiff's chances of prevailing at trial, the rule does not encourage higher defendant settlement offers. As a result, it turns out that in such cases the rule actually *decreases* the likelihood of settlement. The reason, in essence, is that in such cases the rule – like the English rule of cost allocation – tends to increase the stakes at trial, thereby discouraging settlement. In such cases, then, the rule is prone to backfire in its attempt to promote settlement. See Spier (1994a), which also considers a more general class of mechanisms.

Another line of inquiry has been the investigation of two-sided offer-of-settlement rules, under which either the plaintiff or the defendant may be held

responsible for the other's costs if she rejects a settlement proposal and then fails to do "better" at trial. Using a model in which the parties are symmetrically informed about the case, Bebchuk and Chang (1997) show that such a rule leads the parties to settle for exactly the amount of the expected judgment, even if the parties have very different litigation costs. Thus, such an offer-of-settlement rule may negate the distorting effect of litigation costs, which (absent the rule) may push the settlement amount a good distance away from the expected judgment. This striking result depends, however, on the assumption that the parties are symmetrically informed, as well as on certain assumptions about the shape of the distribution of cases.

Procedures Governing Adjudication. The procedures for adjudicating lawsuits are likely to have an impact both on the likelihood that a case will settle, and on the terms on which it settles. This is, quite simply, because legal procedures shape the bargaining environment in which settlement negotiations occur. See generally Mnookin and Kornhauser (1979). Analytically, these effects may be conveniently divided into two major categories.

First, legal procedure can influence parties' settlement decisions by affecting what will happen if the case goes to trial. More precisely, procedure affects the costs and expected outcome of going to trial, as well as its riskiness; these factors, in turn, influence the parties' decisions whether and on what terms to settle. Consider, for example, burden of proof rules, which allocate among the parties the responsibility for gathering and presenting evidence to the court. By determining the distribution of anticipated litigation costs between the parties, as

well as the odds of winning a trial, these rules will significantly influence what the parties will demand or offer to settle. See Mnookin and Kornhauser (1979); Sobel (1985); Hay and Spier (1997). Similarly, preclusion rules, which specify the circumstances in which judgments can be reopened, determine the magnitude of parties' expected litigation costs as well as their chances of winning, and hence strongly influence the amount they are willing to settle for. See Hay (1993).

A second effect of procedure is on the parties' *information* about what will happen at trial. For example, consider again the example of discovery rules, which compel the parties to exchange before trial the evidence in their possession. The effect is to reduce the amount of private information a party can have about what will happen at trial, thus reducing informational barriers to settlement. See Cooter and Rubinfeld (1994); Shavell (1989). Another example is the use of devices such as summary judgment and bifurcated trial procedures, by which courts adjudicate the issues in a case sequentially, rather than all at once. The effect of these devices is to reduce party uncertainty about the outcome of a case, thus potentially facilitating settlement. To illustrate, suppose the court's policy is first to adjudicate the question of liability, then to determine the amount of damages. After the first (liability) phase of the trial is over, party uncertainty about the outcome of the case will be less than it was before; all that remains for the parties to disagree about is the level of damages. See Landes (1993). A final example is damage "schedules" that specify in advance the sum to be paid for a specific injury; by making the victim's actual losses irrelevant at trial, these rules

eliminate an important potential informational barrier to settlement. See Spier (1994b).

An interesting insight to emerge from the study of these matters is that procedures designed to improve adjudication in some way may prove counterproductive, when effects on settlement are taken into account. For example, permitting extensive discovery enables the parties to unearth evidence that might otherwise go undetected; in this way it may improve the accuracy of trial verdicts by increasing the amount of information available to the court. Yet it also enables a party to inflict costs on her opponent, and as a result may be used as a weapon with which to extract settlements that do not reflect the merits of the case; the opponent will rationally make substantial settlement concessions in order to avoid the costs of discovery. Hence a discovery rule that leads to greater accuracy in adjudication may lead to less accuracy in settlement. See Hay (1994). Similarly, a burden of proof rule that increases accuracy in adjudication may, because of the costs or risks it creates for one or the other party, decrease the accuracy of negotiated settlements. See Mnookin and Kornhauser (1979). Effects of this sort remain relatively unexplored and are a fruitful path for future research.

Alternative Dispute Resolution. In recent years, policy makers have introduced various legal procedures designed expressly to increase or accelerate the settlement of disputes. These may be divided into two basic categories. One is mediation, in which a neutral actor attempts to bring about a negotiated resolution of the dispute by, for example, facilitating communication between the

parties. The other is arbitration, in which the parties briefly present their cases to a judge or some other decision maker, who renders a nonbinding "verdict" in the case; either party is free to reject the verdict and go to trial. Of course, parties can make use of these procedures on their own, without prodding by the legal system; on the economics of this decision, see Shavell (1994). It has long been common for litigants, especially in large-stakes cases, to hire a mediator or arbitrator to assist them in resolving their dispute out of court. What is novel in recent policy developments is the legal system's attempt to encourage (through subsidies) or require their use. To what extent do these "alternative dispute resolution" procedures increase or accelerate settlement?

Theory suggests that these procedures are likely to help the parties overcome certain types of barrier to settlement. The procedures may be of some predictive value concerning the likely outcome of trial; thus, to the extent party disagreement over expected trial outcomes impede settlement, we should expect these procedures to make settlement easier. The procedures may also help overcome certain principal-agent impediments to settlement, particularly if the parties rather than just their lawyers participate. See Posner (1986); Mnookin (1993).

The procedures are unlikely, however, to reduce asymmetric-information barriers to settlement. Suppose that, in the absence of these procedures, a privately-informed party will rationally choose to keep her information private (perhaps in order to conserve its surprise value at trial). Compelling her to participate in a mediation or arbitration proceeding will not change that incentive;

she will, if permitted to, simply refrain from disclosing the information in the proceeding. Interestingly, the incentive to do this is compounded by rules that penalize a party who, having rejected an arbitrator's proposed verdict, fails to do better at trial. Such a rule encourages a party to make a poor showing in the arbitration proceeding. See Bernstein (1993). Hence, compulsory alternative dispute resolution may in some instances frustrate its own goal of encouraging settlement. The design of procedures that desirably and effectively calibrate the parties' incentives to settle will doubtless remain a topic of considerable practical and theoretical interest for some time to come.

Bibliography

Aumann, R. 1976. Agreeing to disagree. *Annals of Statistics* 4: 1236-39.

Bebchuk, L. 1984. Litigation and settlement under imperfect information. *RAND Journal of Economics* 15: 404-415.

_____. 1996. A new theory concerning the credibility and success of threats to sue. *Journal of Legal Studies* 25: 1-25.

_____ and Chang, H. 1997. The effect of offer-of-settlement rules on the

terms of settlement. Working Paper 97(7), University of Southern California, Los Angeles, CA..

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.

Braeutigam, R, Owen, B. and Panzar, J. 1984. An economic analysis of alternative fee shifting systems. *Law and Contemporary Problems* 47(4): 173-204.

Cooter, R., Marks, S. and Mnookin, R. 1982. Bargaining in the shadow of the law: a testable model of strategic behavior. *Journal of Legal Studies* 9: 225-51.

Cooter, R. and Rubinfeld, D. 1989. Economic analysis of legal disputes and their resolution. *Journal of Economic Literature* 27: 1067-97.

_____. 1994. An economic model of legal discovery. *Journal of Legal Studies* 23: 435-64.

Danzon, P. and Lillard, L. Settlement out of court: the disposition of medical Malpractice Claims. *Journal of Legal Studies* 12: 345-78.

Farmer, A. and Pecorino, P. 1994. Pretrial negotiations with asymmetric information on risk preferences. *International Review of Law and Economics* 14: 273-81.

Genn, H. 1988. *Hard Bargaining: Out of Court Settlement in Personal Injury Actions.* Oxford: Oxford University Press.

Gould, J. 1973. The economics of legal conflicts. *Journal of Legal Studies* 2: 279-300.

Gilson, R. and Mnookin, R. 1994. Disputing through agents: cooperation and conflict between lawyers in litigation. *Columbia Law Review* 94: 509-66.

Hause, J. 1989. Indemnity, settlement, and litigation, or I'll be suing you. *Journal of Legal Studies* 18: 157-80.

Hay, B. 1993. Some settlement effects of preclusion. *Illinois Law Review* 1993: 21-57.

_____. 1994. Civil discovery: its effects and optimal scope. *Journal of Legal Studies* 23: 481-517.

_____. 1995. Effort, information, settlement, trial. *Journal of Legal Studies* 24: 29-62.

_____ and Spier, K. 1997. Burdens of proof in civil litigation: an economic perspective. *Journal of Legal Studies* 26: 413-33.

Kaplow, L. 1994. The value of accuracy in adjudication: an economic analysis. *Journal of Legal Studies* 23: 307-402.

Katz, A. 1987. Measuring the demand for litigation: is the English rule really cheaper? *Journal of Law, Economics & Organization* 3: 43-76.

Kennan, J. and Wilson, R. 1991. Bargaining with private information. *Journal of Economic Literature* 31: 381-99.

Kritzer, H. 19901. *The Justice Broker: Lawyers and Ordinary Litigation.* New York: Oxford University Press.

_____. 1991. *Let's Make A Deal: Understanding The Negotiation Process In Ordinary Litigation.* Madison: University of Wisconsin Press.

Landes, W. 1971. An economic analysis of the courts. *Journal of Law and Economics* 14: 61-107.

_____. 1993. Sequential versus unitary trials: an economic analysis. *Journal of Legal Studies* 22: 99-134.

Loewenstein, G. et al. 1993. Self-serving assessments of fairness and pretrial bargaining. *Journal of Legal Studies* 22: 135-58.

Menkel-Meadow, C. 1984. Toward another view of legal negotiation: the structure of problem solving. *UCLA Law Review* 31: 754-842.

Miller, G. 1986. An economic analysis of rule 68. *Journal of Legal Studies* 15: 93-125.

_____. 1987. Some agency problems in settlement. *Journal of Legal Studies* 16: 189-215.

Mnookin, R. 1993. Why negotiations fail: an exploration of barriers to the resolution of

conflict. *Ohio State Journal on Dispute Resolution* 8: 235-56.

_____ and Kornhauser, L. 1979. Bargaining in the shadow of the law: the case of

divorce. *Yale Law Journal* 88: 950-97.

Polinsky, A. and Rubinfeld, D. 1988. The deterrent effects of settlements and trials. *International Review of Law and Economics* 8: 109-24.

Posner, R. 1973. An economic approach to legal procedure and judicial administration. *Journal of Legal Studies* 2: 399-458.

_____. 1986. The summary jury trial and other methods of alternative dispute resolution:
some cautionary observations. *University of Chicago Law Review* 53: 366-91.

Priest, G. and Klein, B. 1984. The selection of disputes for litigation. *Journal of Legal Studies* 13: 1-55.

Reinganum, J. and Wilde, L. 1986. Settlement, litigation, and the allocation of litigation costs. *RAND Journal of Economics* 17: 557-68.

Rosenberg, D. and Shavell, S. 1985. A model in which lawsuits are brought for their nuisance value. *International Review of Law and Economics* 5: 3-13.

Shavell, S. 1982. Suit, settlement and trial: a theoretical analysis under alternative methods for the allocation of legal costs. *Journal of Legal*

Studies 11: 55-82.

_____. 1989. The sharing of information prior to settlement or litigation. *RAND Journal of Economics* 20: 183-95.

_____. 1993. Suit versus settlement when parties seek nonmonetary judgments. *Journal of Legal Studies* 22: 1-14.

_____. 1994. Alternative dispute resolution: an economic analysis. *Journal of Legal Studies* 24: 1-28.

_____. 1997. The fundamental divergence between the private and the social motive to use the legal system. *Journal of Legal Studies* 26: 575-613.

Sobel, J. 1985. Disclosure of evidence and resolution of disputes: who should bear the burden

of proof? In *Game Theoretic Models of Bargaining*, ed. A. Roth. New York: Cambridge

University Press.

_____. 1989. An analysis of discovery rules. *Law and Contemporary Problems* 52(1): 133-59.

Spier, K. 1992. The dynamics of pretrial negotiation. *Review of Economic Studies* 59: 93-108.

_____ 1994a. Pretrial bargaining and the design of fee-shifting rules. *RAND Journal of Economics* 25: 197-214.

_____. 1994b. Settlement bargaining and the design of damage awards. *Journal of Law, Economics, & Organization* 10: 84-95.

_____. 1997. A note on the divergence between the private and social motive to settle under a negligence rule. *Journal of Legal Studies* 26: 613-23.