May 18, 1994

Nancy L. Schwartz Memorial Lecture

Kellogg Graduate School of Management

Negotiation with Private Information: Litigation and Strikes

Robert Wilson *

I am glad to contribute this twelfth annual lecture celebrating the memory of Nancy Schwartz. Many of us remember the years in the 1970s when the most exciting developments in the theory of industrial organization were in the series of articles by the prolific team of Morton Kamien and Nancy Schwartz. Nancy's work with Mort initiated a renaissance based on explicit analysis of strategic behavior in dynamic contexts. Subsequent work on entry, limit pricing, and durable goods has continued from their seminal studies; and their results (and important survey article) on the effects of market structure on technological innovation remain classics. We were directors of doctoral programs at the same time, and I can attest too that she was an effective leader in that role. All who knew her are sad that she is not with us now. In keeping with her interest in strategic behavior, my theme today examines negotiations where private information hinders efforts to settle disputes efficiently. I develop the hypothesis that informational differences are likely an important cause of the costly impasses that afflict litigation and wage negotiations. I think she would have welcomed the use of game theory to address this practical problem; after all, her studies of competitive battles examine a similar kind of bargaining with limited means of credible communication.

Acknowledgements. My studies of wage negotiations are conducted jointly with John Kennan of the University of Wisconsin, and of litigation, with Robert Mnookin of Harvard Law School. I have also benefitted greatly from discussions with Peter Cramton of the University of Maryland. Kenneth Arrow, Ian Ayres, Lee Ross, and Amos Tversky at the Stanford Center on Conflict and Negotiation have been steady sources of inspiration. I am deeply indebted to these colleagues. The Hewlett Foundation and the National Science Foundation provided generous support for research.

* Stanford Business School, Stanford, CA 94305-5015.

Introduction

I have chosen negotiation to illustrate the effects of private information in markets. Bilateral bargaining about the terms of a transaction is the simplest context for studies of price formation. We can examine the parties' strategic behavior without the complications present in multilateral trading. To be specific, I focus on two cases. One is litigation in which a plaintiff and a defendant negotiate the terms of a settlement to avoid trial. In the other, a union and a firm negotiate the base wage in an employment contract.

In these specific contexts, I emphasize that informational disparities impose limits on the efficiency of the outcome. In particular, some settlement negotiations are likely to fail, requiring a costly trial on the merits; and some wage negotiations are likely to entail costly strikes. My theme is that self-interested behavior based on private information runs the risk of an outcome that is inferior for both parties. This theme is part of a broad hypothesis that costly disputes reflect fundamental barriers to successful negotiation. Some barriers stem from strategic behavior; others from cognitive biases, social processes, and institutional factors. This hypothesis may seem counter to the enthusiasm for 'alternative dispute resolution' procedures that encourage cooperative problem solving based on mutual understanding and goodwill. Nevertheless, understanding the root causes of failure can aid the design of better negotiation procedures.

Some History

To introduce the issues, I first recount some key insights. For my purpose here, the main ideas are in the work of George Akerlof (1970) on adverse selection, and Michael Spence (1973) on signaling. They were among the first to demonstrate how informational disparities impair efficiency.

Akerlof's Analysis of Adverse Selection

Akerlof studied a hypothetical market for used cars. He supposed that the seller knows more than the buyer does about the quality of her car. The significant conclusion is that they may fail to trade, even if they both know they could gain from trading. This inefficiency stems from the asymmetry of information: not knowing the quality, the buyer doesn't know which prices are mutually advantageous. Any particular price will be rejected by the seller if her car's quality is correspondingly higher. The buyer anticipates therefore that the seller's acceptance of a deal indicates that the quality is low compared to the price — and in that case the buyer may judge the price too high. Indeed, there may be no price acceptable to both parties.

Nowadays Akerlof's analysis is conducted using the mathematical theory of incentive compatibility. To apply this theory, one writes down the necessary conditions implied

by self-interested strategic behavior in *any* negotiation procedure, and then shows that these conditions are violated by all strategies that might result in trade. Or when trade is incentive-compatible, these conditions prescribe limits on the efficiency attainable.

This style of analysis reveals that the crucial ingredient is a feature popularly called the *winner's curse*. The quality affects each party's valuation of the car, and therefore the range of prices acceptable to each. Not knowing the actual quality, the buyer infers the likely range of qualities from the prices acceptable to the seller. Trade fails when the buyer realizes that every price, if it were acceptable to the seller, implies too low a range of possible qualities for it to be acceptable to him too.

The winner's curse is an important feature in other trading contexts. It is most familiar in auctions. Often the value of the item is inherently similar for all bidders, and each bidder bases his bid on an estimate of that value. In this case, each bidder must adjust his bid downward to guard against the likelihood that he wins only when his estimate is higher than others' estimates. That is, winning reveals that his estimate is biased high — even though it is unbiased beforehand. [1]

The same feature afflicts negotiations during litigation. Both the plaintiff and defendant encounter a version of the winner's curse that discourages generous settlement offers, and thereby raises the likelihood of an impasse and recourse to a trial.

Spence's Analysis of Signaling

Spence elaborated Akerlof's analysis by adding the possibility that the seller advertises or *signals* her car's quality. The buyer is still unable to verify quality directly, so the seller must offer a credible signal to persuade the buyer. A typical example of a signal that credibly connotes high quality is a credential, such as a mechanic's certification or a warranty, that would be more expensive if the quality were low, indeed too expensive to justify the expenditure. Thus, a credible signal is one that the seller would find unprofitable to imitate if her car's quality were any lower than it actually is. [2] The conclusion from Spence's analysis is that the seller can overcome the informational barrier to trade by providing a credible signal about its quality. Nevertheless, the outcome is inefficient if the expenditure on the signal does not improve the quality, or does not improve it enough to justify the cost.

Nowadays the analysis of signaling is a refined art form. Modern versions address dynamic contexts in which signaling continues throughout extended negotiations, or is invoked continually during repeated encounters. For instance, sustained signaling can be used to build and maintain a reputation. In some cases the predictions parallel Spence's analysis: credible signaling enables trade that would otherwise be thwarted by the buyer's fear of adverse selection. Akerlof's conclusions persist in other cases where credible signals are too expensive or the parties' interests are opposed. Entry deterrence provides examples of both

cases: as a warning to potential entrants, an incumbent firm might use its price to signal its cost; or, it might meet entry with aggressive pricing to sustain a reputation for ferocity that deters subsequent entrants. [3]

Subsequent analyses emphasize that the buyer has options too. The buyer's *screening* strategy is dual to the seller's signaling strategy. Screening is usually implemented via price discrimination, in which the buyer offers a menu of options among which the seller can choose. A typical menu offers several prices depending on the extent of the warranty obligation accepted by the seller. Though they differ in who takes the initiative, screening and signaling can result in similar outcomes in static contexts. But dynamic scenarios imply substantially different predictions. When the seller is impatient, the buyer can screen dynamically by offering a sequence of prices increasing over time, provided there is a fixed and long interval between price increases. That is, the seller prefers to accept a profitable offer immediately rather than wait for a higher price later. This kind of price discrimination is not very profitable for the buyer if the interval is short, because the seller does not incur a significant cost of delay from waiting for a higher price. In contrast, signaling by the seller is unaffected by the length of the interval between offers: she can signal credibly by waiting longer than she could do profitably were her quality or reservation price lower than it actually is. Thus, the buyer screens by imposing a costly delay on the seller for rejecting each serious offer; whereas the seller signals by incurring a delay cost sufficient to make her claim credible, even without serious offers in the interim.

Optimal strategies for signaling and screening depend sensitively on the details of specific situations. Nevertheless, several features are true generally. Their advantage is credible communication, which allows mutually profitable trade when otherwise it might be prevented by informational disparities. Their disadvantage is wasteful expenditure on signals, or costly delays in reaching agreement on the price.

Signaling and screening are intrinsic features of other negotiations affected by informational disparities. Later we interpret costly delays in wage negotiations in these terms. Strikes, lockouts, and work slowdowns arguably fit patterns predicted by theories of signaling and screening. [4] Although inefficient, enduring curtailed production long enough signals credibly that the firm cannot afford a high wage. The screening scenario is similarly palatable to the union: it exploits its monopoly power to price discriminate by lowering its wage demand gradually. It is less important to classify strikes according to whether they reflect signaling or screening tactics than to recognize that in either case they are the predictable consequence of informational differences.

It is well to recognize, however, that this hypothesis cannot be verified conclusively. Disputes have many causes, and informational disparities are peculiarly inaccessible to outside observers. One can attempt at most to check whether aggregate data fit predicted

profiles of strike incidence, duration, and wage settlements.

Litigation

I turn now to consideration of negotiations between a plaintiff and a defendant to settle a suit. In the event of an impasse, the default is a trial.

Several features distinguish litigation from wage disputes. One is that neither party can force the other to curtail production or employment; rather, the main cost of an impasse is the cost of preparing and conducting a trial. Delay costs are less important than the lump-sum cost of trial, and in any case delays are limited by the plaintiff's right to a timely trial. Another is that settlement involves only a cash transfer from the defendant to the plaintiff, whereas the consequences of an impasse are uncertain for both parties. Apart from expenses, the trial judgment is also a cash transfer in the amount of the damages awarded the plaintiff, if any. It is the likelihood and magnitude of the judgment that is most uncertain.

A model of litigation requires two ingredients to capture these features. The first is that the parties' total gain from settling the suit is the sum of their costs of a trial. The second is that each relies on an estimate of the judgment, usually in the form of a probability that the defendant will be found liable, and if so, the amount of the damages. An attorney usually provides these estimates of the trial cost and the magnitude of the judgment. The attorney provides expert assessment of these dollar amounts, as well as counsel during settlement negotiations and the preparation and conduct of a trial.

These ingredients reveal several aspects. One is that litigants 'bargain in the shadow of the law.' That is, the range of possible settlements is constrained by the range of possible judgments and the trial cost. [5] In many cases, this range is narrow because knowledgeable attorneys make similar predictions. If both know the facts of the case, and the applicable statutes, precedents, and rules of evidence are clear, then each attorney is bound to advise the client to settle rather than waste the cost of a trial. Asymmetries might affect how the avoided trial costs are divided between the parties, but an impasse is unlikely. Indeed, ensuring that most trials have predictable outcomes is a major aim of legislators, judges, and the bar. The social value of the attention paid to appellate review, the role of precedent, and rigorous standards of procedure and evidence stems partly from this consideration. The judiciary's success is reflected in high settlement rates for most kinds of civil suits. For instance, only about 5% of tort cases are tried.

The second aspect is that a settlement negotiation resembles the used-car market studied by Akerlof. The plaintiff has a suit that the defendant might buy at the price of a settlement. If she sells, the plaintiff foregoes the *difference* between the judgment and her cost of the trial. If he buys, the defendant gains the *sum* of the judgment and his cost

of the trial. The key feature is that both parties' valuations of a settlement are affected by a common factor about which they are both uncertain. That is, in litigation the trial judgment plays the same role that unobservable quality plays in other transactions.

The parties' mutual uncertainty about a component affecting both their valuations is the hallmark of the winner's curse. Because they are on opposite sides of the market, as seller and buyer of the suit, the risk of adverse selection is severe and affects both parties. In practice, this risk appears in the guise of reactive devaluation and regret. Reactive devaluation occurs when an offer deemed generous by one side is interpreted by the other as evidence of weakness, justifying demands for even more favorable terms. Similarly, regret occurs when one's offer is accepted, indicating that a less generous offer might have been acceptable. The prevalence of these emotional reactions from the involved parties indicates that often the give and take of settlement negotiations is best done privately by their attorneys, who can identify a compromise from a frank exchange on views on the merits of the case. [6]

The third aspect is that prospects for settlement depend crucially on the parties' information about the judgment. Two extreme cases illustrate the possibilities. In the first, one or both parties have direct knowledge about components of the judgment. For instance, the defendant knows the trial will reveal that he is liable, or the amount of damages due to his negligence; or analogously, the plaintiff knows the value of her injury provable in court. In the second, neither has such privileged information, and the best that each can obtain is an unbiased estimate of the judgment. One's attorney is usually the most accurate estimator, because a complete assessment of the prospects at trial requires thorough familiarity with the facts of the case *and* expert knowledge of the law.

Differences in Material Information

The first case is addressed by rules of civil procedure. Each side can depose the other's witnesses, and judges often grant rights of discovery to a party claiming an informational disadvantage regarding material facts. Further, each party must disclose in advance its line of argument and its evidence. The main motive for these rules is to ensure a fair trial, but an important consequence is to promote settlements. The beneficial effect is predicted by strategic models of negotiation. The settlement rate is lower when one or both parties know material information than when they must divulge this information.

The reasoning that underlies these predictions is a direct analog of Akerlof's analysis of adverse selection. When the defendant has superior information about his liability, for instance, the plaintiff anticipates acceptance of her settlement offer only when her prospects at trial are good. Indeed, there may be no offer that would not provoke regret after taking account of the implications of its acceptance by the other side. This is another instance of the

winner's curse: the mere acceptance of one's settlement offer reveals significant information about material facts known to the other side.

One could envision signaling, as in Spence's analysis, but in fact signaling is muted in litigation. [7] A defendant with a weak case can mimic one with a strong case, so a plaintiff is unlikely to accept a defendant's assertion that the suit is without merit. The notable exception is to reveal the material facts, if this can be done credibly without prejudicing the trial. This is the role in practice of discovery, depositions, and exchanges of evidence, which allow relevant information to be disclosed according to rules of procedure. The legal system has long recognized, at least implicitly, that mechanisms for revealing information without prejudicing trials are essential both for fair trials on the merits, and for encouraging pre-trial settlements.

Nevertheless, existing mechanisms are not perfect. Major commercial litigation suffers from prolonged and costly discovery. When large judgments are at stake, finding cracks in the opposing argument can justify expenditures that might have large distributional consequences but need not promote efficiency. Or, the expense can be wasted in a fruitless search for a 'smoking gun' that, were it there, would be the proverbial 'needle in a haystack.' A central issue is whether discovery can be abused to intimidate or harass, simply by imposing substantial expenses on the respondent. Unfortunately, few studies examine the strategic uses of discovery and other procedural privileges, and fewer outline the extent of the potential for abuse. [8] In principle, dissipative expenditures are effective signals only if they cannot be imitated profitably were one's case weaker than it is; in this sense, the possible role of discovery lies mainly in the willingness of the respondent to sustain the disruption and cost without conceding to a settlement.

On the other hand, some disputes cannot be ripe for settlement until substantial communication of material facts occurs; discovery seems essential to settling these cases. In some kinds of litigation it is common for opposing attorneys to disclose information to minimize clients' costs and to facilitate settlement. Clients' interests are served when attorneys perceive rewards from a reputation for cooperation.

Differing Estimates of the Judgment

The second extreme scenario pertains to the last phase of litigation. At the crucial stage before trial, the parties or their attorneys are equally aware of the arguments and evidence on which the trial will depend. In considering choices between settlement and trial, each side relies on an assessment of the likely trial outcomes. This assessment is typically provided by the attorney, whose expert knowledge is essential to determining the acceptable range of settlements. The attorney's work to learn the issues in the suit and to develop the trial arguments produces a joint product: this effort establishes settlement terms *and* prepares for trial. (In English law these tasks are partly separated between a solicitor and a barrister.)

The winner's curse intervenes in this context too. Within the privacy of the attorney-client relationship, each attorney provides the client with an estimate of the expected judgment, and thus the acceptable terms of settlement. Although each side estimates the same quantity, their estimates can differ, and it is the difference that determines whether a settlement is reached. For instance, suppose they settle whenever the plaintiff demands less than the defendant offers; and, to keep matters simple, they split the difference by settling midway between the two compatible offers. [9] This scenario captures the feature that after a settlement each party realizes its offer was more generous than needed; and similarly, after settlement fails each realizes that a more generous offer might have been accepted, and might have been more profitable. This is one sense of *ex post* regret that is inevitable in negotiation — with hindsight, another strategy appears better.

The other sense of regret occurs when a party is blind to the implication that the other's offer is a signal about its estimate. In particular, if the other's attorney is as accurate as one's own, then the average of their two estimates is more accurate than either alone. The winner's curse thus occurs when one ignores the inference that acceptance of the offer reveals that prospects at trial are better than one's attorney estimated. The plaintiff sees acceptance as revealing that the defendant estimated a larger judgment; the defendant sees acceptance as indicating a lower judgment; and for both, the proper inference is that the best estimate is midway between.

These are classic features of adverse selection, in which one's offer tends to be accepted more readily in less favorable circumstances. This 'statistical' description omits the frequent emotional response of reactive devaluation, in which, say, the plaintiff views the settlement as less fair after seeing the defendant's offer, especially if it is large. Or if settlement fails, it is easy to blame the impasse on unrealistic estimates of the judgment or willful attempts to extract a small settlement. Our aim here, however, is to focus on fundamental strategic barriers to settlement, so we suppose the parties' strategies account accurately for the statistical effects of adverse selection.

As one anticipates, the winner's curse discourages generous settlement offers. The magnitude of the effect depends on the parameters of the particular situation. In the simple case where the attorneys provide equally accurate unbiased estimates having the familiar bell-shaped 'normal' distribution, the optimal offer strategies depend only on a relative cost ratio; namely, the ratio of the total trial cost to the standard deviation of estimates. In particular, the plaintiff demands more than her attorney's estimate by a margin that depends on this ratio, and similarly, the defendant offers less. When a settlement occurs, the amount is itself an unbiased estimate of the judgment that would have resulted from a trial. The frequency of settlements is higher if the cost ratio is higher. This reflects the general pattern that settlement rates are higher when trials are more expensive, relative to

the accuracy with which experts can predict the eventual judgment. This is essentially a reformulation of the hypothesis that litigants bargain in the shadow of the law.

The predictions derived from this kind of analysis accord well with data about litigation. The most important requirement of a predictive theory is to explain why most disputes settle without trial, and why settlement rates vary among different kinds of litigation. The chief prediction is that the cost ratio need not be large for the settlement rate to be in the 85% to 95% range typical for several kinds of litigation: ratios of 7 to 15 suffice. Further, the predicted settlement rate is highest for those suits with relatively high trial costs and/or estimating precision. [10] Some are nearly impossible to confirm, such as the prediction that settlements are unbiased estimates of trial judgments (since settlements preclude trials). This prediction, nevertheless, conforms to the general view that settlements are as fair as trials, and better because they are cheaper, so judges impose few impediments to settlements. [11]

Final-Offer Arbitration

These conclusions about the effect of informational disparities on settlement rates apply to other contexts. One is the resolution of disputes by private trials and arbitration. Arbitration of wage disputes is mandatory for some public-sector employees, and others contract in advance to resolve conflicts this way. The arbitrator acts as a private judge to select a settlement binding on both parties. In the version known as final-offer arbitration, each party files an offer in advance and then the arbitrator chooses between them.

A familiar instance is the procedure adopted by the baseball players' union and the owners' association for salaries of players with sufficient experience. If agreement on a player's salary is not reached by a specified date in January then the player and the team owner file proposals from which an arbitrator selects one at a hearing in February. Over the years, the data fit a prediction based on a cost ratio that is low because arbitration hearings are inexpensive compared to litigation, and because less-formal rules of procedure, evidence, and judgment preclude estimates that are as accurate. In particular, the pre-filing settlement rate is about 75%, which is low compared to most kinds of litigation.

This quasi-judicial procedure introduces novel ingredients that, by way of contrast, reveal implicit aspects of litigation. Symmetric models of arbitration predict an additional property of unbiasedness, namely the arbitrator is equally likely to choose each proposal. In fact, players won over 40% of the cases. Another remarkable feature is that the parties' commitment in January to the proposals arbitrated in February provides a window of several weeks in which to bargain further during the interim. In fact, 80% of the cases filed for hearings settled during this interim period, yielding a 95% total settlement rate. This harvest of additional agreements presumably reflects both the additional information

gleaned from the proposals submitted, and the substantial risk of losing at the arbitration hearing.

The key feature distinguishing final-offer arbitration from litigation is the link between the settlement negotiation and the hearing. Because the arbitrator chooses between proposals generated by the negotiation up to the filing date, each party has a substantial incentive to make serious proposals, and more generous proposals are rewarded with greater chances of winning. The link is manifest further in the bargaining during the interim until the hearing, when again the filed proposals generated by prior negotiations determine the risks of a final impasse. My guess is that this linkage accounts for the fact that final-offer arbitration achieves settlement rates comparable to litigation, even though hearing costs are less and the outcome is no easier to predict. To see the potential force of these linkages, imagine that the arbitration were supplemented with a penalty paid from the loser to the winner: increasing the penalty drives the predicted settlement rate to 100%.

Litigation has no linkage between negotiation and trial. References to settlement attempts are excluded from trials, and each case is judged solely on the merits. This separation precludes any possibility that the trial phase induces incentives for settlement earlier in the dispute process. This conforms to the view that the law establishes objective standards for liability and damages, but it also prevents procedural concessions to encourage settlements.

Strikes

The costs of preparing and conducting trials are a kind of inefficiency. They are inefficient because the parties are capable of settling the dispute without a trial, thereby avoiding these costs. The costs of litigation are generally deplored. My impression, however, is that emotional reactions are mild. Trials produce justice; they also pick winners and losers, and attorneys earn a livelihood. Few cases receive vigorous criticism of the costs entailed. [12] Scholars note nevertheless that legal costs are substantial. A telling statistic is that plaintiffs' net receipts from tort judgments are less than the sum of the costs incurred by both parties. [13] Surely this indicates that in many cases one or both parties regret afterwards that they did not make a more generous offer to settle initially.

Emotional reactions to strikes are stronger. By interrupting operations, a strike reduces employment and output. The firm loses profits immediately, and its relationships with customers and suppliers may erode, not to mention embittered relations with its workforce. The effect on union members is usually severe. To sustain the strike they must forego regular income and deplete savings to pay bills for necessities such as groceries and mortgage payments. The waste of their talents and energies to extract a wage settlement from the firm is a tragedy from any viewpoint.

The characteristic reaction after a strike is regret. Regret is universal because after-

wards it is obvious that they could have agreed initially on the same contract without incurring the costs and scars of the strike. [14] This is not to say that the strike was unnecessary, however, and indeed when the contract expires a few years later, a strike is seen again as a useful tactic.

Except for grievance strikes and disputes over work rules, most strikes are about wages or benefits. The union demands a higher wage than the firm offers, in the belief that the firm *might* be able and willing to pay more if pressure is applied. I emphasize 'might' because the union cannot know what wage the firm will accept. There are always hopes that the firm settles quickly for a high wage, and also dismal prospects that a long strike yields little wage improvement. It is this uncertainty, I think, that is the root cause of subsequent regret. After a strike it is evident that initial hopes for higher wages were ephemeral, and that efforts to obtain them were fruitless. Knowing this after the fact, union members wish they had settled earlier for the same wage.

The firm also experiences a sense of tragedy. It claimed initially that it could not afford to pay the wage demanded by the union. Afterwards it laments that a costly strike was necessary to prove its claim — and interprets the final wage settlement as an act of generosity necessary to get the recalcitrant union back to productive work. The lament often takes the form of despair that the only way to convince the union that its claim of penury is credible is to suffer the costs of a prolonged strike.

These viewpoints are two facets of the same underlying phenomenon. The firm has superior information about its valuation of the union's services, and the union knows this. Faced with this uncertainty, the union seeks to extract a wage commensurate with its value to the firm, reflecting the experience and skills of its members. Direct negotiations don't dispel the prospect that the firm is hiding the true value, attempting to garner the larger share for itself. The union sees, therefore, that it must use strong tactics. The advantage of a strike is that it imposes costs on the firm: If you reject this offer then you won't get a better one for a substantial period, and in the meantime production will be disrupted. An alternative message is: We'll accept a low wage only if you demonstrate convincingly that you can't afford a higher wage. Indeed, the firm can foresee that sustaining a strike is the only convincing demonstration.

These implicit messages indicate that a strike is a kind of communication. The strike curtails profits and wages, so the communication is costly. In fact, the costliness of rejecting an offer is necessary, because that is essential to establishing the credibility of the message. It's unfortunate that wage negotiations rely on so cumbersome a means of communication, but talk is cheap, and unconvincing to the union when the matter at stake is the livelihood of its members.

According to this scenario, a strike stems from differences in information between the

firm and the union. Most often the information is asymmetric, in the sense that the firm has superior information. [15] At first glance, this hypothesis seems open to the criticism that the union has access to the firm's accounting reports, 10K filings with the SEC, and often a wealth of internal documents — why then is it unsure what wage the firm can afford? Over 85% of contract negotiations settle without strikes, so indeed the most frequent case may be that the union has enough information to avoid such disputes. To explain the remainder, there are ample sources of uncertainty about the range of wages that a strike could yield. For a craft union, a pervasive problem is that the marginal product of its services is nowhere revealed in accounting statements; in fact, standard accounting practices mask all marginal effects. All unions face the problem that the relevant magnitudes are the firm's opportunity costs in the future. Past data cannot reveal management's predictions about future factor prices, competitive pressures, new products, and new investments. Nor can they reveal the opportunity costs of foregoing plant relocation, capital-intensive technologies, and other means of substitution for the union's services. Even the feasibility and cost of training new employees to replace the union's skilled members may be uncertain.

To understand how a strike accomplishes the communication required to reach a wage settlement, we invoke the processes of signaling and screening. [16]

Strikes as Signaling

Signaling is consistent with the standard story in labor economics, and indeed justifies it. According to this scenario, the firm anticipates it can get a wage that is lower in some proportion to the length of the strike it endures. The relationship between strike duration and the wage settlement is called the union's resistance curve. Signaling implies a simple explanation for this relationship: each wage on the curve is obtainable only after enduring a strike so long that the firm would have preferred to accept a higher wage earlier if its value were higher than it actually is. This is the source of the essential credibility. Bearing the strike long enough enables the firm to argue convincingly that it cannot afford a higher wage. The implicit message is: Look, if I could afford more then I'd have profited from conceding earlier, so believe me when I say that this wage divides the pie fairly. Signaling even incorporates an explicit determination of what 'fairly' means here; namely, it is whatever the wage would result from negotiations without informational differences. [17] This reflects a common practice: often the final phase is a round of serious negotiations to determine the wage after the strike has dispelled the union's more optimistic conjectures about the firm's profitability.

Other modes of signaling are possible depending on circumstances. For instance, if the main consideration is the firm's uncertainty about the union's resolve then it is the union that signals by bearing the costs of the strike.

A peculiar aspect of signaling is that it need not involve serious negotiations until the

strike nears conclusion. A recent instance was the professors' strike in Israel. It was a case in which the government was unsure what salary raises the professors would accept or the duration they would endure the strike. After an initial demand for 100% raises, and a government offer of 20%, the strike persisted for nearly eight weeks without serious negotiations; indeed, the professors' union explicitly avoided a counteroffer. Although other factors were important too, one interpretation of the climax emphasizes the role of a key vote at the Hebrew University, which revealed a 600 to 11 majority in favor of continuing the strike. Serious negotiations followed quickly, and agreement was reached to split the difference: raises of about 60% for senior faculty. I interpret this as an instance of signaling: bearing the strike, and demonstrating willingness to continue, were essential to establishing credibly that the union was unwilling to capitulate and unlikely to dissolve through dissension.

The Israeli case has the typical feature that the strike concludes with a compromise between the initial offers. This excludes some alternative forms of signaling. One possibility, for example, is that strikes are wars of attrition. A war of attrition is the predicted outcome when each party is unsure about the other's daily cost of continuing the battle: the one with the higher cost eventually realizes this and then concedes, leaving all the gains to the winner. This scenario describes well the struggle between firms competing for a market, as in a price war between products that are close substitutes, where the firm with the lower cost eventually ousts the other. But wage negotiations have different ingredients. Because the parties provide complementary inputs to production, fiercely competitive struggles would impair working relationships afterwards. [18] Moreover, strikes usually involve superior information on one side only, and the winner-take-all outcome is apparently rare. For instance, attrition models imply that average wage settlements are unaffected by strike duration, whereas the evidence indicates that wage settlements vary with strike duration. [19]

It is important to realize, however, that some strikes fail dramatically, and embittered relations ensue. [20] After President Reagan quashed the 1981 PATCO strike by hiring permanent replacements, firms applied this tactic successfully (e.g., TWA attendants' strike) or won by threatening (Caterpillar) or partially implementing it (Hormel). In these cases the firms argued that competitive pressures precluded higher wages, and eventually took advantage of the availability of replacements; e.g., TWA hired new employees at lower wages and within eight weeks trained them to replace most of the contingent of 4500 attendants. It seems evident that whenever the union undertakes a strike under the illusion that there is a pie to be divided, when in fact competition in product and labor markets enforces a plain-vanilla market wage, the struggle reverts to something akin to attrition. Either the firm succumbs (Eastern Airlines), the strike collapses (TWA and Caterpillar), or eventually new union leadership adopts a realistic perspective (Hormel). In interpreting these prominent strikes at major firms, nevertheless, it is useful to remember that they are outliers

among the thousands of strikes every year. Statistical analysis of the average run of strikes is better suited to identifying the main factors.

Strikes as Screening

An alternative interpretation of strikes is that they reflect screening. It is surely plausible that a strong, entrenched union wants to exploit its monopoly bargaining position with aggressive price discrimination. According to this strategy, the union lowers its demanded wage in steps — slowly enough and in steps small enough that the firm's best response is simply to accept when waiting for an even lower wage is less profitable than accepting the current demand. In principle, such a strategy enables the union to obtain a wage that is higher in some proportion to the firm's privately known valuation, and to other factors affecting the firm's impatience, such as its discount rate.

One is naturally skeptical that this interpretation assigns the initiative to the union. Unlike signaling, moreover, intertemporal price discrimination relies on an explicit, and somewhat predictable, sequence of serious wage demands. Casual observation suggests the contrary; in fact, the notable absence of serious demands after the first, and the oft-emphasized reluctance of unions to compromise until the strike nears conclusion is one of the hallmarks of journalists' reports. The typical pattern instead is that compromise awaits evidence that the dispute is 'ripe' for settlement, which accords better with the signaling story.

The basic difficulty must lie elsewhere, however, if one is to understand why unions are deterred from exploiting opportunities for price discrimination. I think the problem is in the implicit message from the union to the firm: Think twice before you reject this wage demand, because it will be a long while before you see one much better. This message lacks credibility unless the union can ensure the solidarity necessary to adhere resolutely to the required delay between steps in its sequence of demands. Whenever members' morale might dissolve along the way, screening risks collapse of the union's bargaining strategy. [21]

One can appreciate this risk by imagining the awkward position of the union's leader the day after their offer is rejected by the firm. When they are stressed by the demands of putting food on the kitchen table, members are persuaded easily to make another offer quickly, rather than waiting patiently for weeks. It takes a leader of considerable talent to convince them that a long delay before bettering their offer is absolutely necessary — because they must impose the requisite cost on the firm for rejecting their *previous* offer. Maintaining a reputation for punishing each rejection is essential for screening to work. If the firm finds that a rejected offer is bettered quickly then it anticipates an even better offer soon after: the union's reputation for patience collapses and prospects for wage gains dissolve.

Other Strategic Aspects of Strikes

The exposition above focuses on the interpretation of strikes as signaling and screening. These processes implicitly convey or elicit information credibly, enabling settlement based on mutual understanding of the main economic considerations. Many other factors affect strikes too, so I mention some briefly to indicate the range of possibilities.

A significant consideration is that each party has options outside the existing bargaining relationship. For workers these options represent alternative jobs, but these are not viable options for the union collectively, so the net effect is mainly to constrain its behavior to maintain solidarity. New York and Rhode Island provide unemployment benefits to workers after eight weeks on strike. Data from these states show a pronounced effect from this increase in worker's unemployment income. Settlement rates are tilted significantly around the eight-week mark: they are higher before, and lower after this change in the economic structure.

For the firm, several outside options are important. One is the prospect of ceasing or relocating operations, or displacing a craft union's job with an alternative technology. Another is acquiring a new workforce, as when it opts to hire and train permanent replacements instead of renewing the union's contract. Each of these options puts an upper bound on the range of acceptable wages. When the union knows this bound, the problems of communication are reduced and strikes tend to be shorter. The firm is spared the burden of signaling credibly that its value is not above the bound; or, in a screening context, the union foregoes wage demands above the bound.

This is apparently characteristic of strikes motivated by informational disparities and the need for credible communication. An interesting test of this prediction is a law in Quebec since 1977 prohibiting firms from hiring temporary or permanent replacements. Similar provisions regarding permanent replacements were enacted in Manitoba (1985), British Columbia (1992) and Ontario (1993); these laws also provide reinstatement rights for striking workers. A ban on permanent replacements is proposed in a bill now before the U.S. Congress. Legislators in Quebec evidently believed that the ban on replacements would shorten strikes because it would force firms to the bargaining table (another major motive was to reduce violence). The statistical evidence suggests, however, that strikes have been substantially more frequent and longer, and wage settlements higher, as strategic models predict. [22]

For the union, a further consideration is its option to exert pressure on the firm without foregoing wages. Laws in all states require the firm to pay wages according to the expired contract until an impasse is reached, such as a strike or lockout, and permanent replacements cannot be hired until an impasse occurs. These laws enable the union to receive wages while 'working to rule'. Adhering strictly to the terms of the previous contract impairs productive

efficiency, especially if optimal work rules or job assignments have changed in the years since the previous contract was adopted. The union's choice between a strike and working to rule depends on both the old wage, the production losses imposed on the firm, and the risk of replacement if a strike is chosen.

A significant implication of the work-to-rule tactic is that strikes are only the more obvious part of unions' strategies. [23] Another part includes non-strike delays between the expiration of one contract and agreement on another. To the extent these involve productive inefficiencies, they are a further manifestation of the negotiation process. They can be an important part: even though the firm suffers less disruption of operations, it is bound to pay the old wage, so the pressure to settle can be as great. The union applies this option selectively, using it when the old wage is relatively high and/or the productive inefficiency of adhering to the old work rules is high. Part of the procyclical variation in the incidence of strikes can be attributed to this effect. That is, in boom times with tight labor markets the old wage is more likely to be comparatively low so the union prefers to strike rather than working to rule. [24]

The union's dual options of striking or working-to-rule turn out to be central to the interpretation of the data in terms of strategic models. Attempts to match the predictions to strike data alone fare moderately. Peter Cramton and Joseph Tracy (1992, 1994a) have shown, however, that strategic models fit the data much better when full account is taken of the frequent and sizable delays between expiration of the old contract and agreement on a new one. This is important because non-strike delays amount to a sizable fraction of the total, and more when the old wage is higher. [25] In these studies the best fit is obtained from a model in which, if agreement is not reached quickly, the union chooses optimally between working to rule and striking, depending on how high the old wage is. In either case, the firm endures long enough to signal credibly, and then they settle on the same wage they would have agreed on initially if the union had complete information. [26]

Conclusion and a Caveat

This examination of litigation and wage negotiation focuses on the hypothesis that costly impasses stem from informational disparities. This conjecture stems from the presumption that strategic behavior to exploit or cope with private information is endemic. To the extent this is true, it implies limits on the efficiency of the outcomes that can be attained. Exhortations to resolve disputes by goodwill, mutual understanding, and vigorous effort fall on deaf ears if either party is wary of divulging sensitive information, or skeptical of others' good intentions. To improve efficiency it may be better to concentrate on procedural innovations designed specifically to reduce the hazards of adverse selection or facilitate credible signaling. For instance, final-offer arbitration has had notable success since it was

devised by Carl Stevens (1966).

It is important to realize, however, that this hypothesis cannot be verified definitively from empirical data. Econometricians do not observe participants' private information, and the significance of aggregate data is diminished severely by enormous heterogeneity among the bargaining pairs in large samples. At most, the hypothesis is plausible to the extent it is roughly consistent with the data.

I am encouraged nevertheless by success in interpreting broad patterns, and in examining particular issues. More generally, the progress obtained from detailed studies of negotiation, auctions, and other markets indicates that strategic analysis provides new insights into price formation and sources of inefficiencies. Multilateral trading in other markets introduces additional ingredients, but it is unlikely that the strong effects of informational disparities in bilateral bargaining are eliminated by competitive pressures. Further research will eventually produce a clear view of the strengths and weaknesses of Walrasian models as approximations of how markets work.

References

Admati, Anat, and Motty Perry (1987), Strategic delay in bargaining, *Review of Economic Studies*, 54: 345-64.

Akerlof, George (1970), The market for lemons: quality uncertainty and the market mechanism, *Quarterly Journal of Economics*, 84: 488-500.

Ausubel, Lawrence, and Deneckere, Raymond (1993), Efficient sequential bargaining, *Review of Economic Studies*, 60: 435-61.

Budd, John (1994), Canadian strike replacement legislation and collective bargaining: lessons for the United States, mimeo, University of Minnesota.

Card, David (1990a), Strikes and wages: a test of an asymmetric information model, *Quarterly Journal of Economics*, 105: 625-59.

Card, David (1990b), Strikes and bargaining: a survey of the recent empirical literature, *American Economic Review*, 80: 410-15.

Cho, In-Koo (1990), Delay and uncertainty in bargaining, *Review of Economic Studies*, 57: 575-95.

Cho, In-Koo (1994), 'Stationarity, rationalizability, and bargaining,' *Review of Economic Studies*, 61: 357-374.

Coase, Ronald H. (1972), Durability and monopoly, *Journal of Law and Economics*, 15: 143-9.

Cooter, Robert D., and Daniel L. Rubinfeld (1989), Economic analysis of legal disputes and their resolution, *Journal of Economic Literature*, 27: 1067-97.

Cramton, Peter C. (1992), Strategic delay in bargaining with two-sided uncertainy, *Review of Economic Studies*, 59: 205-25.

Cramton, Peter C., and Joseph S. Tracy (1992), Strikes and holdouts in wage bargaining: theory and data, *American Economic Review*, 82: 100-21.

Cramton, Peter C., and Joseph S. Tracy (1993), The use of replacement workers in union contract negotiations: the U.S. experience 1980-1989, mimeo, University of Maryland.

Cramton, Peter C., and Joseph S. Tracy (1994a), The determinants of U.S. labor disputes, *Journal of Labor Economics*, 12: 180-209.

Cramton, Peter C., and Joseph S. Tracy (1994b), Wage bargaining with time-varying threats, *Journal of Labor Economics*, 12: 594-617.

Cramton, Peter C., Morley Gunderson, and Joseph S. Tracy (1994), The effect of collective bargaining legislation on strikes and wages, mimeo, University of Maryland.

Gertner, Robert, and Geoffrey Miller (1994), Settlement escrows, mimeo, University of Chicago.

Gül, Faruk, and Hugo Sonnenschein (1988), On delay in bargaining with one-sided uncertainty, *Econometrica*, 56: 601-12.

Gunderson, Morley, and Angelo Melino (1990), The effects of public policy on strike duration, *Journal of Labor Economics*, 8: 295-316.

Holt, Charles A., and Roger Sherman (1994), 'The loser's curse,' *American Economic Review*, 84: 643-52.

Institute for Civil Justice (1993), *Annual Report*. Santa Monica CA: Rand Corporation.

Kennan, John, and Robert Wilson (1989), Strategic bargaining models and interpretation of strike data, *Journal of Applied Econometrics*, Supplement, 4: S87-S130.

Kennan, John, and Robert Wilson (1990), Theories of bargaining delays, *Science*, 249: 1124-8.

Kennan, John, and Robert Wilson (1992), Repeated wage bargaining with private information, mimeo, University of Wisconsin.

Kennan, John, and Robert Wilson (1993), Bargaining with private information, *Journal of Economic Literature*, 31: 45-104.

McConnell, Sheena (1989), Strikes, wages, and private information, *American Economic Review*, 79: 801-15.

Mnookin, Robert, and Lewis Kornhauser (1979), Bargaining in the shadow of the law: the case of divorce, *Yale Law Journal*, 88: 950-97.

Mnookin, Robert, and Robert Wilson (1989), Rational bargaining and market efficiency: understanding Texaco v. Pennzoil, *University of Virginia Law Review*, 75: 295-334.

Noldeke, Georg, and Eric van Damme (1990), 'Signalling in a dynamic labour market,' *Review of Economic Studies*, 57: 1-23.

Sobel, Joel (1989), An analysis of discovery rules, *Law and Contemporary Problems*, 52.

Spence, A. Michael (1973), Job market signaling, *Quarterly Journal of Economics*, 87: 355-374.

Spence, A. Michael (1974), *Market Signaling*. Cambridge MA: Harvard University Press.

Stevens, Carl M. (1966), Is compulsory arbitration compatible with bargaining? *Industrial Relations*, 5(2): 38-52.

Rothschild, Michael, and Joseph Stiglitz (1976), Equilibrium in competitive insurance markets: an essay on the economics of imperfect information, *Quarterly Journal of Economics*, 90: 629-649.

Notes

1. The phrase winner's curse is used also to describe bidders' failure to adjust bids appropriately. See Holt and Sherman (1994) for experimental evidence that this is due to inadequate application of the principles of statistical inference, rather than a misplaced desire to win.

2. Spence's exposition actually focused on a competitive market for new employees at firms. In this context, the quality of a worker's labor is some measure of her ability, perhaps enhanced by skills acquired through education. A credential that signals high ability is an educational degree that would be too difficult or costly to obtain if her ability were low. Education might enhance the worker's productivity, but the analysis remains valid even if it does not.

3. In technical terms these differing predictions are called separating and pooling equilibria, depending on whether (a) signaling enables a party to convey credibly its private information; or (b) imitation is sufficiently inexpensive to enable a high-cost incumbent to maintain a reputation that its cost might be low enough to remain profitable in a price war with an entrant.

4. Basic references on signaling models of negotiation are Admati and Perry (1987), Cho (1990, 1994), Cramton (1992), and Noldeke and van Damme (1990); and on screening, Ausubel and Deneckere (1993), Gül and Sonnenschein (1988), and Rothschild and Stiglitz (1976).

5. R. Mnookin and L. Kornhauser (1979).

6. This assumes attorneys have incentives to act in clients' interests, which can be false. In the short run, prolonging a dispute might increase an attorney's income. The long term effect is negative if such actions affect adversely a reputation that attracts clients.

7. This is likely due to the absence of substantial delay costs, because ordinarily litigation does not interrupt productive activities. Credible signaling via rounds of serious offers and counteroffers is not very informative unless each rejection imposes a substantial cost. Other mitigating factors for plaintiffs include accrued interest on judgments, and for defendants, prospects of summary judgment or withdrawal of the suit: one study

of malpractice suits noted that over 40% of claims were dropped. On the other hand, there is a strong incentive to make a serious offer at the final moment 'on the courthouse steps,' or earlier if trial preparation costs are large.

8. Sobel (1989) examines a model in which discovery serves to monitor the signaling content of offers.

9. This can be interpreted as a simple approximation of offers on the courthouse steps. For some models of litigation, this procedure is actually optimal, in the sense that it minimizes the frequency of trials and thereby the expected costs of trials. Offers of this kind can be handled by 'settlement escrows' of the sort proposed by Gertner and Miller (1994). The court clerk accepts sealed offers and declares the split-the-difference settlement when this is feasible — and otherwise reveals nothing.

10. The data surely include cases where a party seeks vindication in court or wants to win at any price, perhaps to sustain a reputation for invulnerability to legal challenges, so at best the fit must be rough.

11. At a more fundamental level, strategic models based on informational disparities are consistent with the evidence that *nonbinding* arbitration "rarely achieves statistically significant reductions in total court costs or average case duration" (Institute for Civil Justice, 1993, p. 31). Similarly, Gunderson and Melino (1990) estimate the effect of nonbinding mediation of wage disputes in those Canadian provinces requiring it as approximately zero. Cramton, Gunderson, and Tracy (1994) find similar effects, including nil effects from mandatory cooling-off periods.

12. One prominent case was *Pennzoil v. Texaco*, in which direct expenses totaled about $200 million, plus an estimated $300 million for the plaintiff's attorney's 10% share of the final settlement. In addition, the firms' share prices fell substantially, and Texaco suffered bankruptcy. See R. Mnookin and R. Wilson (1989).

13. Institute of Civil Justice (1993). Such statistics reflect economies of scale in litigation; larger judgments do not entail proportionately larger costs.

14. Regret also pervades failed strikes, such as TWA's permanent replacement of striking flight attendants, and Eastern's bankruptcy and subsequent dissolution prompted by its machinists' strike. In recent years, reactions included bitterness when a firm forced a return to work by threatening to hire permanent replacements, as in the 1992 and 1994 Caterpillar strikes. In the 1985-6 Hormel strike, the firm hired 550 replacements four months before the strike ended.

15. This common interpretation obtains some independent support from the observation that lockouts are rare compared to strikes, but this interpretation is clouded by the fact that in most states unemployment insurance is payable during lockouts. Some countries adopt measures to diminish informational asymmetries; e.g., German firms include union representatives on key committees. Other institutional factors are likely important too: Gunderson and Melino (1990) note that strikes in the U.S. and especially Canada are longer than elsewhere.

16. Unlike theories of litigation, studies of strikes commonly omit aspects derived from adverse selection and the winner's curse of the sort in Akerlof's model. The reason is that the disputants' valuations are interpreted as independent. There is no common factor to link them, and it seems implausible that one party, say the firm, has superior information about workers' reservation wages or their wages in alternative jobs.

17. For instance, consider simple negotiations based on alternating offers and counteroffers, and suppose each party's impatience for a settlement derives from discounting. Then the wage is the best offer the other cannot afford to refuse; namely, the wage that is as good as the discounted gain from the best counteroffer. This is the implication of the standard bargaining model developed by Rubinstein (1982), who shows that there is a unique subgame-perfect equilibrium, and it has this property.

18. Wars of attrition between animals of the same species (competing for spoils such as food or mates) are moderated by symbolic tests of strength. One might interpret some strikes in that vein; e.g., the Israeli strike was preceded by a two-day warning strike. In general, however, economic disputes seem to offer few inexpensive options for credible communication.

19. The pattern differs between Canada and the U.S. Using Canadian data, Card (1990a) finds no significant duration effect on wages, and with a different data set Cramton, Gunderson, and Melino (1994) find a positive effect that they interpret as consistent with changing strike costs (e.g., inventory depletion) or the firm's uncertainty about the union's resolve. Using U.S. data, McConnell (1989) and Cramton and Tracy (1992, 1994) find a negative effect consistent with a resistance curve. In any case, the resistance curve need not be steep nor easy to detect, since incentive compatibility requires only that the slope corresponds to the interest rate.

20. A friend of mine acquires firms bankrupted by labor-management enmity, and profits by rebuilding cooperation. The bitter Hormel strike is documented in the Oscar-winning film *American Dream* by Barbara Kopple.

21. The formal description of the hazards of screening strategies is called the Coase property. The gist is that screening is ineffective when the interval between offers is short. See Coase (1972) and Gül and Sonnenschein (1988).

22. Gunderson and Melino (1990) and Cramton, Gunderson, and Tracy (1994). The latter estimate that replacement bans increased strike incidence by 50%, duration by 37%, and wages by 12% (or 4.3% when one controls for the previous wage); moreover, unions' wages increased by five times more than they increased their strike costs. Budd (1994) argues that the evidence is not definitive because Quebec implemented other policy changes at the same time; he finds weaker effects with Card's data set and different model specifications. Cramton and Tracy (1993) find from U.S. data that the risk of replacement decreases the incidence of strikes, and a ban on replacements would increase incidence by a third.

23. Cramton and Tracy (1994a) cite the AFL-CIO's 1986 manual for strike leaders that specifically argues the advantages of work-to-rule tactics.

24. See Cramton, Gunderson, and Tracy (1994) and Cramton and Tracy (1992, 1994a) who use Canadian and U.S. data respectively.

25. Cramton and Tracy indicate that about four-fifths of delay days represent non-strike 'holdouts', and they calibrate the cost as about 4% of full production, based on U.S. data. They show that during the 1980s, holdouts increased and strikes decreased substantially, possibly in response (after the PATCO strike) to firms' new tactics of threatening to hire replacements. As mentioned previously, Cramton and Tracy (1993) find from U.S. data in the 1980s that risks of replacement decreased strike incidence, and encouraged holdouts and other in-plant tactics.

26. Cramton and Tracy (1992, 1994ab) develop both the theory and the empirical applications of models in which the union has multiple, and possibly time-varying, threats. The predicted settlement rates from strikes and holdouts are about the same. In the U.S. in the 1980s, the declining incidence of strikes was mainly a compositional effect: the overall incidence of the two tactics remained fairly constant. Card (1990) and Cramton, Gunderson, and Tracy (1994) find another strong effect in Canadian data that provides a test independent of whether it is the firm or the union who signals. Strike incidence is appreciably lower if the *previous* contract followed a long strike, and lower still if that contract's duration was short, even after accounting for the previous wage. Presumably this reflects serial correlation; e.g., if the previous strike demonstrated that the firm's profitability was low then it is again likely to be low. Kennan and Wilson (1992) analyze a model of this phenomenon.