An Economic Approach to Legal Procedure and Judicial Administration

Author(s): Richard A. Posner

Source: *The Journal of Legal Studies*, Jun., 1973, Vol. 2, No. 2 (Jun., 1973), pp. 399-458

Published by: The University of Chicago Press for The University of Chicago Law School

Stable URL: https://www.jstor.org/stable/724058

AN ECONOMIC APPROACH TO LEGAL PROCEDURE AND JUDICIAL ADMINISTRATION

*RICHARD A. POSNER**

INTRODUCTION

IT is frequently alleged that the processes of legal dispute resolution in America are dangerously overloaded, due to delay, congestion, inefficiency, and lack of resources. To appraise this allegation and formulate durable reforms requires an understanding of the operating principles of the system for resolving legal disputes. This article seeks to advance that understanding by means of the powerful tools of economic theory. Although it builds on recent articles by William M. Landes and by the present writer,[1] it is more than an extension of the previous work. That work took for granted the rules of procedure that provide the framework of the legal dispute-resolution system; the emphasis was on how plaintiffs (mainly prosecutors) and defendants maximize utility within its constraints. The present article attempts to explain the procedural rules and practices that give the system its distinctive structure and to predict the effects of changes in one part of the system on the other parts. It thus adds to the literature (as yet small) that is developing a positive economic theory of the institutions of the legal system.[2]

Part I explains the basic analytical framework of the article. The purpose of

* Professor of Law, University of Chicago; Research Associate, National Bureau of Economic Research. Gary S. Becker, Isaac Ehrlich, Richard A. Epstein, Owen M. Fiss, William H. Kruskal, William M. Landes, Melvin W. Reder, George J. Stigler, and participants in workshops at the National Bureau of Economic Research and the University of Chicago commented helpfully on earlier versions of this article. This study has been supported by a grant from the National Science Foundation to the National Bureau for research in law and economics, but it is not an official National Bureau publication since it has not yet undergone the full critical review accorded National Bureau studies, including approval by the Bureau's board of directors.

The appendix at the end of the article presents mathematical treatments of several of the topics discussed.

[1] William M. Landes, An Economic Analysis of the Courts, 14 J. Law & Econ. 61 (1971); Richard A. Posner, The Behavior of Administrative Agencies, 1 J. Leg. Studies 305 (1972).

[2] See Richard A. Posner, Economic Analysis of Law (forthcoming), especially pts. I, VI, for a fuller discussion of this theory. Kenneth C. Scott, Standing in the Supreme Court—A Functional Analysis, 86 Harv. L. Rev. 645, 670-83 (1973), published after this article went to press, also analyzes an aspect of legal procedure in economic terms.

legal procedure is conceived to be the minimization of the sum of two types of costs: "error costs" (the social costs generated when a judicial system fails to carry out the allocative or other social functions assigned to it), and the "direct costs" (such as lawyers', judges', and litigants' time) of operating the legal dispute-resolution machinery. Within this framework the rules and other features of the procedural system can be analyzed as efforts to maximize efficiency.[3] Part II discusses error costs in civil cases, with particular reference to accident cases, the most common type of civil action. Part III considers error costs in criminal and administrative cases. Direct costs are then taken up. Since out-of-court settlements are usually cheaper than trials, the settlement rate affects the overall cost of legal dispute resolution, and Part IV analyzes the factors influencing the decision to settle rather than litigate. Part V analyzes factors influencing litigants' behavior (including their expenditures) in cases that go to trial. Part VI discusses interactions between error and direct costs.

The article touches on a number of topics, including, among others, burden of proof, the right of the defendant before an administrative agency to a trial-type hearing, the constitutional guarantee of counsel to indigent criminal defendants, the English and Continental practice of requiring the losing litigant to reimburse the winner's attorney's and witness fees, delay in court, pretrial discovery, nuisance suits, class actions, and res judicata. Many features of the procedural system are shown to be consistent with our postulated goal of cost minimization; others should be changed if society wants to approach closer to that goal. A number of questions that a comprehensive theory of procedure would include are omitted. In particular, the article does scant justice to the role of courts as makers (as distinct from appliers or enforcers) of law; it emphasizes rather the role of courts in the resolution of factual disputes. However, a brief discussion of the role of precedent—judge-made rules—in judicial decision-making appears in Part VI.

I. A Framework of Analysis

An important purpose of substantive legal rules (such as the rules of tort and criminal law) is to increase economic efficiency.[4] It follows (as demon-

[3] "Efficiency," as used in this article, has its usual economic sense of value-maximizing. Writers on procedure often use the term differently. For example, the Columbia University Project for Effective Justice equated it with "capacity to produce settlements" and "to shorten trials." See its Field Survey of Federal Pretrial Discovery II-7 (Walter E. Meyer Research Inst. of Law, Feb. 1965). As we shall see, such usage can lead to confusion.

[4] See Richard A. Posner, *supra* note 2, at pts. I, VI. I argue there (in ch. 23) that the rules of evidence and other procedural characteristics of the litigation process have interesting parallels to the economic market and encourage the decision of cases on efficiency grounds, but this branch of the economic theory of procedure is not pursued in the present study.

strated in Parts II and III) that mistaken imposition of legal liability, or mistaken failure to impose liability, will reduce efficiency. Judicial error is therefore a source of social costs and the reduction of error is a goal of the procedural system. The reader may challenge the last proposition by citing, for example, the rule excluding from criminal trials evidence obtained by an illegal search. Such evidence is highly probative; its exclusion reduces the accuracy of the fact-finding process in criminal trials. But this type of exclusionary rule is exceptional, and is recognized—and often bitterly criticized—as such.

Even when the legal process works flawlessly, it involves costs—the time of lawyers, litigants, witnesses, jurors, judges, and other people, plus paper and ink, law office and court house maintenance, telephone service, etc. These costs are just as real as the costs resulting from error: in general we would not want to increase the direct costs of the legal process by one dollar in order to reduce error costs by 50 (or 99) cents. The economic goal is thus to minimize the sum of error and direct costs.

Despite its generality, this formulation provides a useful framework in which to analyze the problems and objectives of legal procedure. It is usable even when the purpose of the substantive law is to transfer wealth or to bring about some other noneconomic goal, rather than to improve efficiency. All that is necessary is that it be possible, in principle, to place a price tag on the consequences of failing to apply the substantive law in all cases in which it was intended to apply, so that our two variables, error cost and direct cost, remain commensurable.

To illustrate the utility of the economic approach, consider the question whether the defendant in an administrative action (such as deportation, license revocation, or the withdrawal of a security clearance) should be entitled to a trial-type hearing. The tendency in the legal discussion of this question has been to invoke either a purely visceral sense of fairness or a purely formal distinction between penal and nonpenal sanctions. The economic approach enables the question to be framed in rational and functional terms. We ask first whether error costs would be substantially increased by denial of a trial-type hearing. Error costs (discussed in detail in the next part) may here be regarded as the product of two factors, the probability of error and the cost if an error occurs. If the facts on which the outcome of the administrative proceeding turns are the kind most accurately determined in a trial-type hearing, the probability of error if such a hearing is denied is apt to be great. If, in addition, the cost of an error if one occurs would be substantial because the sanction applied by the agency, whether in formal legal terms penal or not, imposes heavy costs on a defendant, total error costs are likely to be significantly increased by the denial of a trial-type hearing. The increment in error costs must be compared with the direct costs of a hearing; but

these will often be low. The cost inquiries required by the economic approach are not simple and will rarely yield better than crude approximations, but at the very least they serve to place questions of legal policy in a framework of rational inquiry.

II. The Costs of Error in Civil Actions

A. *An Analysis of Error Costs in Accident Cases*

Suppose a company inflicts occasional injuries on people with whom it cannot contract due to very high transaction costs. Victims of these injuries could prevent them only at prohibitive cost (we will initially assume), but the company can purchase various relatively inexpensive safety devices that would reduce the accident rate significantly. In the absence of legal sanctions it has no incentive to purchase such devices since, due to the costs of transacting, it cannot sell anyone the benefits of the devices in increasing safety. If the tort law makes it liable for the costs of these accidents,[5] and is enforced flawlessly, the company will purchase the optimum quantity of safety devices. If the law is not enforced flawlessly, a suboptimum quantity of safety equipment will be procured.

The goal of a system of accident liability is to minimize the total costs of accidents and of accident avoidance. If we assume that the only feasible method of accident avoidance is the purchase of a particular type of safety equipment,[6] then those total costs are minimized by purchasing the quantity of that equipment at which the marginal product of safety equipment in reducing accident costs is equal to the marginal cost of the equipment. This marginal product is the rate at which the number of accidents inflicted by the company declines as the quantity of safety equipment purchased increases, multiplied by the cost per accident. The marginal cost of safety equipment is simply the unit price of such equipment if, as we shall assume, that price does not vary with the amount of equipment that the company purchases.[7] These relationships are depicted in Figure 1. The intersection of the marginal product and marginal cost curves determines the socially optimum quantity of safety equipment for the company to buy and install (q_S).

The company, however, is not interested in minimizing the social costs of

[5] Under either a strict-liability or a negligence standard. See Richard A. Posner, Strict Liability: A Comment, 2 J. Leg. Studies 205 (1973).

[6] The assumption is unrealistic. In particular, the firm's output will also be an important factor in the level of accidents. We exclude it to simplify the analysis; it could be included without affecting our conclusions.

[7] Presumably the firm is too small a purchaser of such equipment to affect the price by varying the quantity that it purchases. The assumption may be valid even if we are considering the behavior of the industry, rather than of a single firm, but is in any event inessential and could be abandoned without affecting our conclusions.

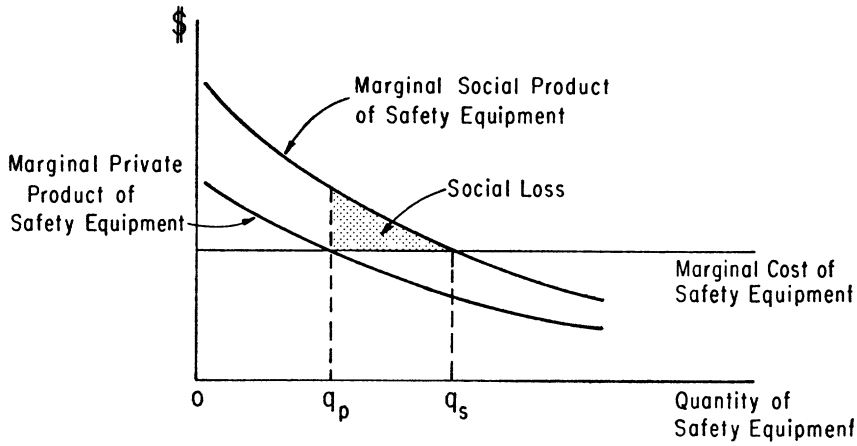

FIGURE 1

accidents and accident avoidance; it is interested only in minimizing its private accident and accident-avoidance costs. The former are the social costs of the firm's accidents multiplied by the probability that the firm will actually be held liable—forced to pay—for those costs. Since legal error presumably causes erroneous impositions as well as erroneous denials of liability, we must add a third term to the firm's cost function: the amount of money that it is forced to pay out in groundless claims. That amount is a function of the legal-error rate and disappears when that rate is zero. We ignore it for the moment.

The company minimizes its private accident and accident-avoidance costs by equating the marginal product of safety equipment in reducing its accident liability to the marginal private cost of that equipment (which we assume is the same as the marginal social cost). This marginal private product is simply the marginal social product weighted by the probability of the firm's being held liable. If that probability is one, the marginal social and private products are the same. But when the probability is less than one—that is, when the legal-error rate is positive—they diverge,[8] leading to a social loss as shown in Figure 1. The higher the error rate, the greater the reduction in the purchase of safety equipment and the greater the social loss.

The analysis is incomplete because we have ignored the possible effect of a positive error rate, operating through the third term in the company's cost function (liability resulting from groundless claims), on the firm's purchase of safety equipment. Suppose that the errors against the company took the form

[8] The effect of error on the purchase of safety equipment is identical to that of a gross-receipts tax on an industry's output. The tax shifts the demand curve (here viewed as the industry's average-revenue curve) downward by the rate of the tax and induces the industry to reduce its output. See appendix for mathematical treatment.

exclusively of accident victims' exaggerating the extent of their injuries. By increasing the company's private accident costs, these errors would increase the marginal private product of safety equipment. Thus, while errors in favor of the company would lower the company's marginal private product curve in Figure 1, errors against the company would shift it back upward. In fact, however, although all errors in favor of the company operate to lower its marginal-product curve, only some errors against the company operate to raise it. The purchase of additional safety equipment will not prevent the erroneous imposition of liability in a case in which no accident would have occurred in any event—the victim fabricated it—or in which the accident was inflicted by someone else and could not have been prevented by the defendant. Such errors do not increase the value of safety equipment to the firm and hence the marginal private product of that equipment. But even here a qualification is necessary. Additional safety equipment might strengthen the company's defense against a suit arising out of an accident actually caused by someone else. The company might be able to argue that, in view of all of the safety precautions it had taken, it could not have caused the accident.[9] Still, it seems a reasonable conclusion that a positive error rate will result in a net reduction in the company's marginal private product of safety, and hence in a net social loss.

A glance back at Figure 1 will confirm that this social loss will be greater, the more serious the accident.[10] This supports our earlier point concerning the relevance of the size of the stakes in an administrative action to the question whether the defendant should have a right to a trial-type hearing. To be sure, the stakes have another effect that is not captured in Figure 1, which treats the error rate as completely exogenous. As shown later in this article, an increase in the stakes in a case will usually induce the parties to spend more money on the litigation. This in turn will reduce the probability of an erroneous result, and so, by our previous analysis, the social loss from error. Aggregate error costs might actually be smaller in a class of big cases than in a class of small ones. But this does not invalidate our analysis of the right to a trial-type hearing, for the denial of such a hearing, in a case turning on disputed factual questions, deprives the defendant of any opportunity to reduce the probability of error by spending heavily on the factfinding process.

Thus far we have assumed that legal error will have no effect on the behavior of accident victims. In fact, by increasing expected accident costs net of

[9] Groundless claims might impose costs in another form: the company might spend heavily on lawyers, etc., to resist them. However, it might economize on these costs in the cases where, due to legal error in its favor, it was not sued at all, or where the claimant was unable to make a strong case.

[10] A higher accident cost per case raises both marginal-product curves by the same proportion. See appendix.

compensation, error encourages prospective victims to engage in self-protection. If adequate compensation were paid in every accident case, the net cost of accidents to the victims would be zero and their incentive to take precautions also zero. But if by reason of error the expected compensation is only (say) 80 per cent of the expected accident cost, the net cost of accidents to victims becomes positive and they have an incentive to adopt precautions that cost less than the uncompensated accident costs that they prevent.

The effect of error is thus to shift safety incentives from injurers to victims. If the victims can prevent the same accidents at lower cost than the injurers, such a shift will produce a net social gain rather than a social loss, but where this is possible the injurers should not be liable in the first place. If, as we assume, the substantive law places liability where it will encourage the most efficient methods of loss avoidance, the shift in safety incentives brought about by error in the legal process will produce a net social loss. But the loss may be slight.

We introduce victims' precautions into our model in Figure 2. The right-hand side of the diagram is identical to Figure 1; the left-hand side is a similar diagram depicting the purchase of safety devices by prospective victims. When legal error is zero, the marginal private product of safety equipment purchased by potential victims is also zero: since they are fully compensated, they have no incentive to invest in safety equipment. A positive error rate increases that marginal private product from zero to a level equal to the marginal social product of victim safety precautions multiplied by the error rate. Thus, if the error rate is 20 per cent, so that a victim has only an 80 per cent chance of being compensated, the value to him of taking precautions rises to 20 per cent of the value of such precautions to him if injurers were never liable. The social benefit depicted in Figure 2 that is generated when victims take precautions because the error rate is positive is larger the higher that rate is. This benefit must be subtracted from the social loss from error in order to determine the net social cost of erroneous failure to impose liability on injurers.

The analysis of legal-error costs would be very different if the purpose of the underlying substantive law were not to improve the allocation of resources but were instead to compensate victims of certain accidents. The amount of undercompensation due to legal error would be equal to the product of the error rate, the cost per accident, and the number of accidents that occur for which the injurer should be held liable. That number will be greater the higher the error rate, for we know from the previous discussion that the number of accidents rises with the error rate, and therefore undercompensation must rise with the error rate. Besides the error rate, the principal factors determining the total amount of undercompensation are the effectiveness of the law (albeit imperfectly enforced) in deterring accidents, and the scope of the law.

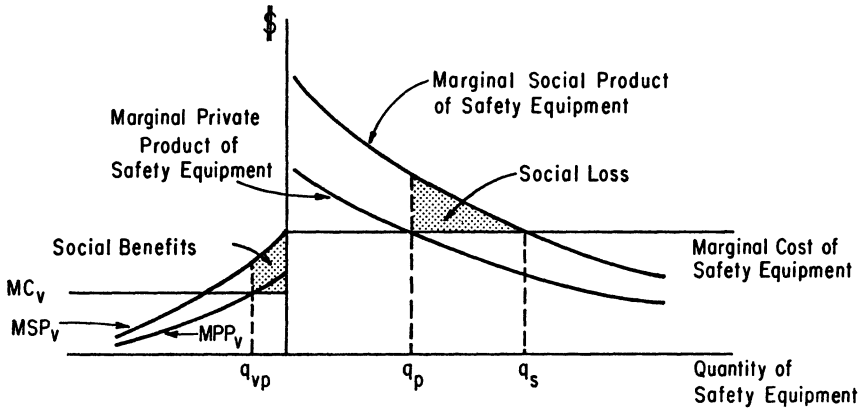

FIGURE 2

For identical error rates, the number of accidents occurring for which the injurer should be held liable will be greater if the legal standard is strict liability than if it is negligence. The costs of error are therefore likely to be higher under a strict-liability than under a negligence standard if failure to compensate where compensation is due is reckoned as a cost, and not merely as a transfer of wealth. Finally, to make undercompensation commensurable with the social loss of Figure 1, we need to know the rate at which a dollar in undercompensation is equated to a dollar in scarce resources consumed; it need not be one to one. In short the cost of legal error may differ dramatically depending on whether the purpose of the underlying substantive law is viewed as allocative or distributive.

B. *Biased and Unbiased Error*

It is useful to distinguish between "biased" and "unbiased" error. Unbiased error in our usage is any error that is as likely to operate against one party to the dispute as it is to operate against the other. Such an error gives judgment to undeserving plaintiffs in about half of the erroneously decided cases and to undeserving defendants in the other half; accepting perjured testimony is an example. A biased error is one more likely to defeat plaintiffs than defendants or vice versa. The previous analysis assumed unbiased error.

Two types of biased error may, in turn, be distinguished. The first arises from a deliberate decision to bias a source of error (as in the rule that the guilt of a criminal defendant must be proved beyond a reasonable doubt). Consider a social-loss function that consists of two terms: the social loss of Figure 1, due to failure to impose liability on injurers in all cases in which

they should be held liable, and the social loss that results when the judicial system awards compensation to a victim who could have averted an accident by appropriate safety precautions at lower cost than the injurer.[11] Assume that if the parties have the same burden of proof,[12] the probability of an accurate determination of liability in both situations will be 90 per cent, meaning that in 10 per cent of the cases in which injurers should be held liable they are not held liable, while in 10 per cent of the cases in which victims should be held liable (denied compensation) for failure to take cost-justified safety precautions they receive compensation. Now let the standard of proof be changed to require that the defendant prove nonliability to a certainty. Victims will win every case. The probability of injurers' being held liable when they should be held liable will rise to one, causing the social loss from legal errors favoring injurers to fall to zero. But the probability of victims' being held liable when they fail to take proper safety precautions will fall to zero, which will cause the social loss from such failures to rise. We cannot be certain whether our total loss function will be higher or lower without knowing the specific values of the relevant parameters, but probably it will be higher. What we have done, in effect, is to impose a standard either of strict injurer liability (with no contributory negligence) or strict victim liability, depending on which probability has gone to zero. Both standards are less efficient than the alternatives.[13] The effects of moderate bias, however, cannot be appraised a priori.

A second type of biased error occurs when a source of error affects the parties' chances unequally. Consider the rule—still followed in some states—that the victim of an accident must prove his freedom from contributory negligence. In a fatal accident to which there were no witnesses, the effect of the rule, if followed to the letter, would often be to prevent recovery even though the victim was in fact free from contributory negligence; the rule would never operate in favor of victims. But courts do not apply the rule in such cases. Instead they presume in the absence of contrary evidence that the victim was exercising due care. The effect is to increase the injurer's in-

[11] There is no presumption that an injurer can avoid a costly accident more cheaply than a victim; in general an efficient allocation of responsibilities for accident prevention will require that prospective victims take some safety precautions. These optimal precautions are to be distinguished from the second-best precautions, discussed previously, that prospective victims take only because of errors in the legal process that permit injurers to escape full liability.

[12] The burdens of proof (in technical legal language, "risk of nonpersuasion") would be approximately equal if plaintiff, to prevail, had only to establish that his version of the facts was more probably true than the defendant's. Whether this is a correct characterization of the civil ("preponderance of the evidence") standard is discussed shortly.

[13] See John P. Brown, Toward an Economic Theory of Liability, 2 J. Leg. Studies 323 (1973); Richard A. Posner, *supra* note 5.

centive to take precautions and reduce the victim's. This is an improvement if we assume that the injurer is more likely to be negligent than the victim; the modified rule is better if the reverse is more likely.

C. *Burden of Proof*

Generally in civil cases, the plaintiff must persuade the factfinder that it is more likely than not that his version of the facts is true. Taken literally, this would mean that an erroneous decision was as likely to be in the plaintiff's favor as in the defendant's, since "more likely than not" implies a percentage barely greater than 50 per cent. It would not follow that half of the judgments in civil cases were erroneous; presumably most are correct.[14] But within the fraction of the erroneous, about as many judgments would be for plaintiffs as for defendants.

In cases where substantial evidence is introduced by the plaintiff, the more-likely-than-not—or "preponderance of the evidence"—standard seems to be adhered to literally, which is consistent with the economic approach. It is interesting to note that the standard implicitly equates a dollar lost by someone erroneously adjudged liable to a dollar lost by one erroneously denied compensation. Some people would say that it is worse to be forced to pay for an injury one did not actually inflict than to be erroneously denied compensation for an injury suffered. But this approach would be inconsistent with the modern economic view that interpersonal comparisons of utility are arbitrary, and it is not followed in ordinary civil cases.[15] Another question raised by the preponderance standard is why the risk of nonpersuasion should be placed on the plaintiff rather than on the defendant. The answer is that since no allocative purpose would be served by shifting a loss in a case where the defendant's liability was indeterminate, the rule economizes on litigation expenditures. This also implies, however, that the burden of persuasion as to

[14] As explained in Part V(A)3 *infra*.

[15] A fuller discussion of the distributive effects of the preponderance standard may be in order. The preponderance standard sets the same value on two types of error, assuming the same dollar amount is involved in the two types of case. One error is to transfer wealth from a deserving defendant to an undeserving plaintiff. The other is to fail to transfer wealth from an undeserving defendant to a deserving plaintiff. Assume that the parties have identical marginal-utility-of-money curves (negatively sloped) and, before the legal dispute in question, the same wealth, W. The plaintiff's claim is that $1000 was wrongfully transferred from him to the defendant. If the claim is false but the plaintiff wins, the plaintiff's wealth rises to W + $1000 and defendant's falls to W — $1000. The loss to the defendant is greater than the gain to the plaintiff because of the identity and negative slope of the parties' marginal-utility-of-money curves. If the claim is true but the plaintiff loses, this means that the dispute itself reduced plaintiff's wealth to W — $1000 and increased defendant's wealth to W + $1000; the erroneous legal decision merely confirms the redistribution. Thus the errors are symmetrical even under assumptions that permit interpersonal comparisons of utility.

defenses (grounds for nonliability that the defendant must plead) should be placed on the plaintiff rather than defendant—as used to be the general rule, in fact, with respect to contributory negligence.

In some cases, modification of the preponderance standard may be in order. Thus, suppose that a pedestrian is struck by a bus in circumstances suggesting that the bus company is liable, but his only evidence of the injurer's identity is that 80 per cent of the buses on this route are owned by the defendant, one of two bus companies that use the route. The evidence would be considered insufficient to satisfy the plaintiff's burden of proof, a result defended on the ground (among others) that otherwise the defendant would be held liable for all bus accidents occurring on the route although responsible in fact for only 80 per cent.[16] This argument is unsatisfactory. It implies that it is better that the bus company escape liability in the 80 per cent of the accidents for which it is in fact responsible than that it pay for some accidents that it did not cause, without explaining *why* this is better. The real issue is different. If the only evidence introduced in the case is that the defendant operates 80 per cent of the buses on the route in question, the probability that a judgment in favor of the plaintiff will be erroneous is at least 20 per cent, which is high. Moreover, were this the only evidence ever introduced in such cases, the other bus company on the route would have no incentive to adopt any safety precautions, so the accident rate would rise.[17] Since the error turns out to be very costly, it makes sense to eliminate the cause of the error if possible to do so at moderate cost. Now if the defendant can adduce additional evidence as cheaply as the plaintiff, no harm is done by permitting a jury to find in favor of the plaintiff when all he has presented is evidence of the bus company's market share; and even if it is more costly for the defendant to present evidence as to which bus company was really liable, so long as it is not *too* costly, the 20 per cent margin of error will be greatly reduced or eliminated. Nonetheless, if the plaintiff can adduce additional evidence more cheaply than the defendant, the pure preponderance standard should yield to a threshold approach in order to induce the plaintiff to present such evidence and thus economize on litigation costs.[18]

[16] Laurence H. Tribe, Trial by Mathematics: Precision and Ritual in the Legal Process, 84 Harv. L. Rev. 1329, 1349-50 (1971).

[17] Not only would the second bus company have a higher accident rate, but its liability-cost advantage might enable it to increase its market share. The result would be a further increase in the accident rate (the weighted average of the companies' individual accident rates) on the route.

[18] A similar approach seems to be implied in Laurence H. Tribe, *supra* note 16, at 1341 n.37, and 1349-50. Consistently with the discussion in the text, the doctrine of *res ipsa loquitur* has been used to induce defendants to produce facts that would be more costly for the plaintiff to produce—such as what when on in defendants' operating room while plaintiff was unconscious. See Ybarra v. Spangard, 25 Cal. 2d 486, 154 P.2d 687 (1944); Louis L. Jaffe, Res Ipsa Loquitur Vindicated, 1 Buff. L. Rev. 1 (1951).

In some cases the preponderance standard may be too high. Suppose that a certain form of radiation for which the emitter is strictly liable increases the risk of developing cancer by one-tenth of one per cent. A person who is exposed to the radiation and later develops cancer will not be able to prove that it was caused by the radiation; there is only a small possibility that the radiation was a causal factor in any particular case. If the emitter is held liable for all cancers developed by people exposed to the radiation his judgment bill will be 1000 times greater than the actual harm he caused, while if the preponderance standard is applied he will escape all liability for the harm he caused. The correct solution would appear to be to permit each cancer sufferer among those exposed to the radiation to recover one tenth of one per cent of the costs of the cancer to him. If damages so computed are too small to justify his incurring the costs of suit, a class action (a device discussed later in this article) should be permitted.

III. Error Costs in Criminal and Administrative Cases

A. *The Optimum Probability of Convicting the Innocent*

The prosecution in a criminal case must prove the defendant's guilt beyond a reasonable doubt. This implies that the probability of convicting an innocent person is very low. Perhaps in no more than five per cent of all cases in which an erroneous judgment is made is an innocent person convicted. Can this biasing of error in favor of criminal defendants be reconciled with our theory of error?

We begin by dividing the total losses from crime into two parts: the social costs of criminal activity and the punishment costs imposed on both guilty and innocent people convicted of crime.[19] We want to determine the optimum probability that an innocent person who is charged with a criminal offense will be convicted, that is, the probability at which the total losses of crime are minimized. This conditional probability of convicting the innocent enters into our loss function in several ways. To begin with, the social cost of criminal activity is a function of the number of crimes committed, which in turn is a function of the rate at which people who commit crimes are apprehended and convicted for them. The rate of conviction of the guilty can be expressed as a decreasing function of the conditional probability of convicting the

[19] Consistently with the emphasis in this part of the article on error costs, we ignore for the moment the costs involved in determining guilt. The preventive (as distinct from deterrent) effects of punishment are also ignored for the moment. For reasons that will become clear in due course, we assume that punishment takes the form of imprisonment, death, flogging, etc., but not of a fine. We ignore the costs of administering the punishment. For other discussions of the optimum probability of convicting the innocent see John R. Harris, On the Economics of Law and Order, 78 J. Pol. Econ. 165 (1970); H. Laurence Tribe, *supra* note 16, at 1378-79, and studies referred to therein.

innocent. This is because a change in the standard of proof that increases the likelihood of an innocent person's being convicted also increases the likelihood that a guilty person will be convicted, and vice versa. If the standard of proof is set at so high a level that the probability of an innocent person's being convicted is zero, the conviction rate for guilty people will also be zero, since only with a zero conviction rate can all possibility of an innocent person's being convicted be eliminated. Conversely, if the standard of proof is set at so low a level that innocent people, if charged with crime, are always convicted, then presumably the rate of conviction of the guilty will also be 100 per cent. In fact conviction rates are moderately high[20] even though the conditional probability of convicting the innocent is very low, which suggests a function of the general shape depicted in Figure 3.

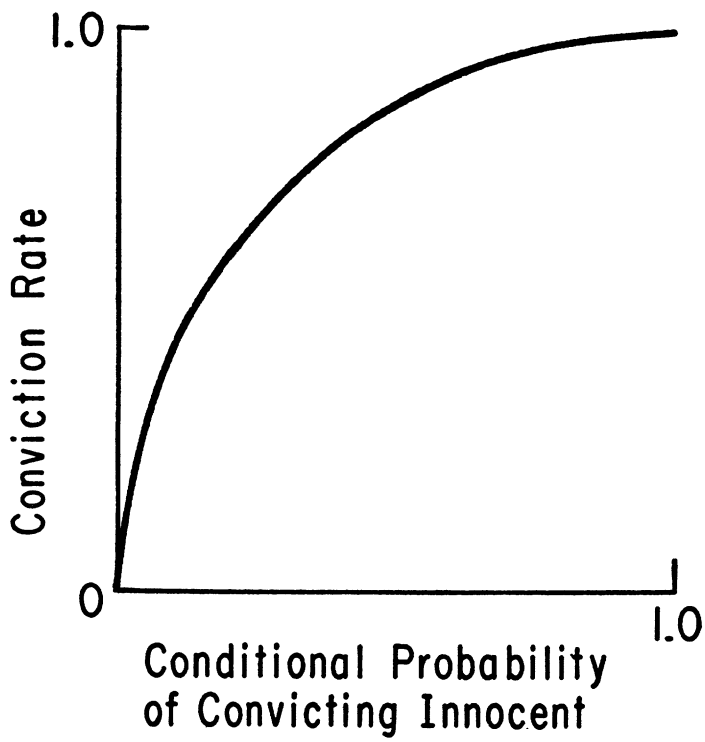

FIGURE 3

[20] See, *e.g.*, figures published annually in the FBI's Uniform Crime Reports. These rates are the rates of conviction of all criminal defendants, whereas we are interested only in the rates of conviction of the guilty. But probably the two kinds of rates do not differ substantially, since the very low probability of convicting an innocent person implies that most acquitted defendants are guilty, assuming—realistically—that only moderate resources are devoted to criminal factfinding (cf. note 22 *infra*).

The conditional probability of convicting the innocent affects the social costs of criminal activity in another way. The number of crimes committed is presumably a decreasing function of the expected punishment costs of crime, and punishing the innocent reduces those expected punishment costs and hence increases the number of crimes and the social costs of crime. To explain, if the expected punishment cost from committing some crime is X, a person who refrains from committing the crime will still have an expected punishment cost, Y, since the conditional probability of punishing the innocent is positive and people consequently face a possibility of being punished for crimes committed by others. The net punishment cost of committing the crime is thus $X - Y$; it is this figure that enters into the prospective criminal's calculations; and the figure is smaller the greater the conditional probability of being punished for someone else's crime.

A reduction in the rate at which the guilty are convicted will affect the punishment costs inflicted on the guilty, which are a social cost too (so long as we abjure interpersonal comparisons of utility). These costs are equal to the product of the number of crimes committed, the probability that someone will be arrested and charged for each crime committed and that he is actually guilty, the probability of his being convicted, and the severity of the sentence imposed. While a reduction in the conviction rate directly reduces one of these factors, it indirectly increases another: the number of crimes committed.

The conditional probability of convicting the innocent also affects, of course, the punishment costs inflicted on the innocent. Those costs are equal to the product of the number of crimes committed, the probability that someone will be arrested and charged for each crime, the probability that this person will actually be innocent, the probability that he will nonetheless be found guilty, and the severity of the sentence imposed. The effect of the conditional probability of punishing the innocent is not, however, so straightforward as this suggests, since that probability also affects the number of crimes committed and the probability of an innocent person's being accused of crime. The latter point is based on the assumption that police and prosecutors prefer to convict guilty rather than innocent people to the extent that it is easier (cheaper) to convict the guilty and therefore that they will attempt to screen out people whom they believe to be innocent so long as this condition holds. The higher the conditional probability of convicting the innocent, the less incentive the police and prosecutor have to screen out the innocent suspect in advance of trial. This suggests another way in which the conditional probability of convicting the innocent affects the social costs of crime. The more advance screening the police and prosecutor do, the less likely a guilty person is to be prosecuted since the screening will exclude some people who appear to be but are not innocent.

Actually to determine the optimum conditional probability of convicting

the innocent would require data that are difficult, perhaps impossible, to obtain, but we can form some tentative impressions by considering the extreme cases. Assume first that the conditional probability of punishing the innocent were much lower than it is today, which would bring it very close to zero. A glance at Figure 3 will show that, at this level, the rate of conviction of the guilty is apt to be negligible. This means that unless criminal punishment has very little effect in deterring crime, the number of crimes committed and hence the total social costs of crime will be very high. And, perhaps surprisingly, the aggregate punishment costs imposed on innocent people may not be trivial, for although only a small fraction of innocent suspects will be charged or convicted, it will be a fraction of a larger whole (all people suspected of crime) since the number of crimes committed will be so much greater. Aggregate punishment costs imposed on the guilty may also, and for the same reason, be high. It is in any event unlikely that a reduction in the conditional probability of punishing the innocent below existing low levels would generate much savings in the punishment costs imposed on innocent people. These punishment costs must already be very small in aggregate, so a further reduction would yield only small savings. In summary, it is unlikely that a substantially lower conditional probability of punishing the innocent than we have today would be optimum, and it seems altogether untenable to suggest that the optimum probability might be zero,[21] since this would imply a zero conviction rate and hence no punishment for crime whatsoever.

Now assume that the conditional probability of convicting the innocent were raised substantially above its existing very low level. One effect would be to increase the rate of conviction of the guilty and thereby reduce the number of crimes and the total social costs of crime. The increase in the conviction rate would be very marked up to a point, but if the conditional probability of convicting the innocent rose to 30 per cent or more, Figure 3 suggests that further increases would have only slight effects on the conviction rate. Moreover, at these high levels of probability of convicting innocent people, the effect of additional convictions in reducing the expected punishment costs of criminals and hence in increasing the social cost of criminal activity would become substantial, while at the same time the punishment costs imposed on both innocent and guilty would be very great. Unfortunately, even if we assume that a conditional probability of convicting the innocent that exceeded 30 per cent could not possibly be optimum, the implications for the proper standard of proof are unclear. The conditional probability of convicting the innocent is a function not only of the standard of proof but

[21] As perhaps implied by Professor Tribe when he states that it is proper to insist "upon as close an approximation to certainty as seems humanly attainable in the circumstances." Laurence H. Tribe, *supra* note 16, at 1374. To the same effect, see his A Further Critique of Mathematical Proof, 84 Harv. L. Rev. 1810, 1818 (1971).

also of the ease of marshaling convincing evidence of innocence, which is generally greater when the defendant is in fact innocent.

Now let us relax the unrealistic assumption in the preceding discussion that the amount of resources devoted to the trial of criminal cases is fixed. Conceivably the only consequence of reducing the conditional probability of convicting the innocent would be an increase in the resources devoted to prosecution. The conviction rate of the guilty might not fall at all,[22] and the additional resources necessary to prevent its falling might be small in relation to the reduction in the punishment costs borne by innocent people. But it is a fair guess that the proportionate increase in prosecutorial resources that would be necessary to maintain existing conviction rates in the face of a conditional probability of convicting the innocent that was at or near zero would be very large; and possibly the proportionate *decrease* in those resources that would be made possible by increasing the conditional probability would be small until that probability reached a level unacceptably high because of its effect on the social loss from convicting innocent people. Moreover, it is unlikely that the effect on the conviction rate of a significant reduction in the conditional probability of convicting the innocent brought about by a change in the standard of proof could be completely offset by an increase in the resources devoted to prosecution. To anticipate later discussion, such an increase might be met by an increase in the resources devoted to defense that would largely nullify the effect of the additional prosecutorial resources in increasing the probability of convicting the guilty. Since a large fraction of criminal defense expenditures are financed by the government, the ability of defendants to match increases in prosecution resources could be curtailed, but the resulting imbalance might move the conditional probability of convicting the innocent back up to where it had been before the attempt to reduce it, for those defendants—the large majority—whose ability to pay for their defense is severely limited.

If our analysis does not establish whether proof beyond a reasonable doubt is the proper standard to apply in criminal cases, still it is helpful in explaining important aspects of criminal procedure, not least the fact that subjective certainty of guilt is ***not*** required. Another aspect is the generally shared impression that the standard of proof is in fact stricter the more serious the offense. The gravity of the offense is positively related to the severity of the

[22] An analogy may be drawn to the statistician's distinction between Type I and Type II errors. A Type I error is the erroneous rejection of the null hypothesis (here, that defendant is not guilty). A Type II error is the erroneous rejection of the alternative hypothesis (defendant is guilty). With a given sample size, a test of statistical significance that reduces the likelihood of committing a Type I error increases the likelihood of committing a Type II error, and vice versa. But both probabilities of error can be reduced by increasing the sample size, which is analogous to increasing the expenditures devoted to finding the truth in a legal proceeding.

penalty, and an increase in severity not only increases the punishment costs imposed on innocent (and guilty) people who are convicted of crime, but also, as we shall see later, induces the parties to spend more money on the litigation. Since greater expenditures should have the effect of reducing the probability of an error whether in favor of or against the defendant, a reduction in the conditional probability of convicting the innocent in grave crimes may reduce the costs imposed on the innocent without reducing substantially the rate of conviction of the guilty. If so, a reduction in the conditional probability may also reduce the social costs of criminal activity through its effect on the expected punishment costs of people who commit crimes.

The analysis also suggests why punitive remedies, when they take the form of fines or other money transfers exclusively, as in the case of antitrust treble damages, are often governed by the laxer civil, rather than the criminal, standard of proof. In such cases, the punishment costs inflicted on the innocent do not represent a net social cost because they show up on the social ledger as equal benefits to the state or other plaintiff. (Indeed, the analysis of error costs in Part II seems fully applicable in such cases.) This distinction may also explain why the burden of proof in a civil fraud case ("clear and convincing evidence") is intermediate between the preponderance and beyond-reasonable-doubt standards. A judgment of liability in such a case imposes a reputational loss on the defendant that is in addition to the money judgment received by the plaintiff but is less costly than most criminal punishments. Most important, the distinction explains why some difference in the standards of proof followed in criminal and civil cases may be entirely consistent with economic theory.

B. *Other Applications*

1. *Administrative Proceedings.* Some administrative proceedings, such as Interstate Commerce Commission reparations cases and National Labor Relations Board back-pay cases, closely resemble ordinary civil cases in that they involve a money transfer. Other administrative proceedings, where restitution is not a feature, resemble criminal prosecutions. A Federal Trade Commission false-advertising order, for example, imposes costs on the defendant that presumably have a deterrent or preventive value, but there is no transfer. However, there is an important difference between such proceedings and criminal cases: prevention, which we ignored in our discussion of criminal cases, often plays a greater role than deterrence in the administrative process. The primary significance of an FTC false-advertising order, for example, is not to impose costs for past violations but to expose the defendant to sanctions if, by repeating his unlawful conduct, he violates the terms of the order.[23]

[23] See Federal Trade Commission Act, 15 U.S.C. § 45(*l*) (1970). However, the FTC's

If we assume both that administrative sanctions are characteristically much less severe than criminal sanctions and that prevention is apt to be a more important objective of the administrative sanction than deterrence, it is easy to see why administrative agencies, even when they impose sanctions that resemble the criminal in that they do not involve a pure money transfer, are not required to follow the criminal standard of proof beyond a reasonable doubt. The less severe the sanction, the higher, by our earlier analysis, the optimum probability of convicting the innocent.[24] And if the major purpose of the sanction is prevention rather than deterrence, we need not worry that by convicting many innocent people the agency will reduce the expected punishment costs of violation and hence the deterrent effect of punishment.

The analysis also suggests, however, that the characteristic combination of prosecution and adjudication in the same agency may be a source of inefficiency. To find the socially optimum probability of punishing an innocent person accused of unlawful conduct, an adjudicator must treat a dollar in cost of punishment of the innocent as equal to a dollar in benefit from punishing the guilty. Having no prosecutorial responsibilities, courts presumably do this. But an agency that is responsible for prosecution may weight a dollar in benefits from successful prosecution more heavily than a dollar in costs of punishing the innocent. The benefits presumably accrue to the agency, however indirectly; it does not bear the costs.

If the agency does not treat the costs imposed on innocent people whom it punishes as a social loss that enters into the determination of the optimum probability of convicting the innocent, it will establish a conditional probability of convicting the innocent that is higher than the social optimum. It will not ignore completely the costs of punishing innocent people, since those costs reduce the deterrent effect of punishment. But since, as mentioned, they do not reduce its preventive effect, and since deterrence may not be an important objective of the administrative sanction, the agency may ignore those costs almost completely with the result that its (private) optimum probability of convicting the innocent accused may be very high. This provides a strong argument for affording defendants in administrative proceedings a right of judicial review of the agency's factfindings (and such other procedural protections as may be necessary to make the right effective[25]), even though the

---

recent practice of compelling firms found guilty of false advertising to print retractions ("corrective advertising") has introduced a significant penal element—unless it is assumed that corrective advertising is as valuable to the consuming public as it is costly to the advertiser.

[24] Where the administrative sanction is very severe, as in deportation, the standard followed, as we would predict, is stricter than the preponderance, and approaches the criminal, standard. See Woodby v. Immigration and Naturalization Service, 385 U.S. 276 (1966).

[25] A recent study found no evidence that one such protection—formal separation of

agency's presumed "expertise" in finding facts within the area of its special competence, and the provision of an internal agency appellate process,[26] might, but for the economic considerations advanced here, argue against any need for judicial review of agency factfindings.

2. *Right to Counsel.* A similar analysis of prosecutors' incentives suggests a possible economic justification for the constitutional guarantee of counsel to indigent criminal defendants. As between two groups of criminal defendants, one guilty but able to afford counsel, the other innocent but unable to afford counsel, we want prosecutors to prosecute only members of the first group, in order to maximize the deterrent (and preventive) effect of criminal punishment. If, however, the prosecutor's maximand is number of convictions (presumbly weighted by seriousness of offense), he will prosecute the guilty only if it is less costly to convict them than it is to convict the innocent. Ordinarily it is; but it may not be if the innocent person is not represented by counsel while the guilty person is. In that event the provision of counsel to indigents may be necessary to avoid an overinvestment in their prosecution. This conclusion would not follow if the prosecutor's maximand were assumed to be deterrence, rather than simply conviction.

IV. Settlement Out of Court

The two preceding parts considered the costs of erroneous judicial determinations. Now we turn to the direct costs of legal dispute resolution. These include the costs of trials (discussed in the next part) and the costs of settling cases without, or before completion of, litigation.

A. *When Are Cases Settled?*

Since settlement costs are normally much lower than litigation costs, the fraction of cases settled is an important determinant of the total direct cost of legal dispute resolution. The necessary condition for settlement is that the plaintiff's minimum offer—the least amount he will take in settlement of his claim—be smaller than the defendant's maximum offer. This is not a sufficient condition: the parties may find it impossible to agree upon a mutually satisfactory settlement price. But we shall assume that settlement negotiations are rarely unsuccessful for this reason[27] and therefore that litigation occurs

---

prosecution and adjudication—reduces the problem of biased adjudication. Richard A. Posner, *supra* note 1, at 323-43. This may be either because other protections are effective (such as judicial review or the internal separation of functions required by the Administrative Procedure Act), or, more probably, because formal separation does not eliminate the agency's felt responsibility for convictions.

[26] Facts in agency proceedings are normally determined by a hearing examiner, comparable to a trial judge, subject to review by the members of the agency sitting as an appellate tribunal.

[27] Typically there are only two parties to a legal dispute. Presumably the problem

only when the plaintiff's minimum offer is greater than the defendant's maximum offer. The plaintiff's minimum offer is the expected value of the litigation to him plus his settlement costs, the expected value of the litigation being the present value of the judgment if he wins, multiplied by the probability (as he estimates it) of his winning, minus the present value of his litigation expenses. The defendant's maximum offer is the expected cost of the litigation to him and consists of his litigation expenses, plus the cost of an adverse judgment multiplied by the probability as he estimates it of the plaintiff's winning (which is equal to one minus the probability of *his* winning), minus his settlement costs.[28] Anything that reduces the plaintiff's minimum offer or increases the defendant's maximum offer, such as an increase in the parties' litigation expenditures relative to their settlement costs, will reduce the likelihood of litigation. Hence measures to reduce litigation costs might actually increase the total costs of legal dispute resolution, by making trials, which are usually costlier than settlements, more attractive than before the measures were introduced.

Anything that increases the plaintiff's minimum settlement offer or reduces the defendant's maximum offer will increase the likelihood of litigation. An increase in the plaintiff's subjective probability of prevailing or in his stakes will do this, but so will an increase in the defendant's subjective probability of prevailing since it will induce him to reduce his maximum settlement offer. An increase in the defendant's stakes in the case will reduce the likelihood of

---

is acute only when there are many parties; then each has a strong incentive to hold out for a large part of the expected value of the transaction. This may be a problem in some multiparty litigation.

[28] If a party is either risk averse or risk preferring rather than risk neutral the expected utility of litigation may be smaller or larger than its expected value. We assume risk neutrality except where otherwise indicated, but the analysis could easily be modified to take account of the existence of nonneutral attitudes toward risk.

We shall also ignore John P. Gould's interesting point, in The Economics of Legal Conflicts, 2 J. Leg. Studies 279, 290 (1973), that when both parties expect to lose the case if it is litigated, each can increase his expected wealth by litigating and at the same time betting with the other party against himself. To illustrate, suppose the stakes in the case are $1000 to each side and each side thinks it has only a 10 per cent chance of winning. Then (if we ignore the costs of settlement and of litigation) plaintiff will accept any offer from the defendant of more than $100 to settle the case. Defendant has an expected loss of $900 so he will settle for anything less than that amount. Assume that the parties decide to litigate and bet each other on the outcome. Plaintiff bets $10,000 that he will lose; defendant bets $10,000 that *he* will lose. The expected value of litigation to the plaintiff is now $9000 plus $100, or $9100, which exceeds any settlement offer the defendant would make, while the expected value of litigation to the defendant is $9000 — $900, or $8100, which exceeds the maximum expected value of a settlement to the defendant (zero). However, Gould's point is probably not empirically important. Such a betting contract could not lawfully be enforced; nor could it be enforced practically: each party would have an incentive to improve the odds by failing to litigate vigorously and by deliberately committing mistakes that made it more likely that he would lose.

litigation by leading him to increase his maximum settlement offer. In the important special case where the stakes to the parties are the same, it can be shown that an increase in those stakes will increase the likelihood of litigation. In that case, litigation cannot possibly occur unless the plaintiff's subjective probability of prevailing is greater than one minus the defendant's subjective probability, for otherwise the plaintiff's minimum settlement offer will be equal to or smaller than the defendant's maximum offer. Assuming that this minimum condition for litigation is satisfied, any increase in the stakes must increase the likelihood of litigation by making the plaintiff's minimum settlement offer grow faster than the defendant's maximum settlement offer.[29]

The approach suggested here assumes that the subjective probabilities, the stakes, and the costs of litigation and of settlement are mutually independent, but they are not. A change in the stakes will affect the amount of money that the parties spend on litigation and this in turn will alter the probabilities of a particular outcome. Settlement costs are probably a function of both litigation costs and stakes. A change in one party's expenditures on litigation, triggered by a change in the stakes or subjective probability of winning of that party, may lead the other party to alter his expenditures on the case, which may induce a further change in the first party's expenditures. As demonstrated in Part V, the interdependence of the parties' expenditures makes it impossible to predict the level of those expenditures unless special, and somewhat arbitrary, assumptions about the parties' reaction patterns are adopted. This in-

[29] A bit of very simple mathematics may help to clarify this point. Let P_p be the plaintiff's subjective probability of prevailing and P_d defendant's, J the stakes in the case, C_p the plaintiff's litigation costs and C_d the defendant's, and S_p and S_d the plaintiff's and defendant's settlement costs, respectively. For the plaintiff's minimum settlement offer to exceed the defendant's maximum offer, the condition for litigation, the following inequality must hold:

$$P_pJ - C_p + S_p > (1 - P_d)J + C_d - S_d. \quad (1)$$

This can be rewritten as

$$J(P_p + P_d - 1) > (C_p + C_d) - (S_p + S_d). \quad (2)$$

A minimum requirement for this condition to hold (assuming that litigation costs exceed settlement costs) is that the left-hand side of the inequality exceed zero, which in turn requires that the sum of the parties' subjective probabilities ($P_p + P_d$) exceed one. If this minimum condition is satisfied, any increase in J, the stakes in the case, will increase the probability of litigation. The first derivative of the left-hand side of inequality (2) with respect to J is simply $P_p + P_d - 1$ and is positive for all cases in which the sum of the parties' subjective probabilities exceeds one. Thus an increase in J must, if what I have termed the minimum condition for litigation is satisfied, always increase the likelihood of litigation.

There is empirical evidence that higher stakes do increase the likelihood of litigation. See H. Laurence Ross, Settled Out of Court 222 (1970); Medical Malpractice—Report of the Secy's Comm. on Medical Malpractice, app., at 13 (U.S. Dep't of Health, Education, and Welfare, Jan. 16, 1973).

determinacy makes the conditions for settlement indeterminate, since not only the plaintiff's and defendant's litigation costs but also their subjective probabilities of prevailing if the case is litigated are functions of their expenditure decisions. We shall propose several possible ways, none altogether satisfactory, of getting around this problem, and in default of a satisfactory alternative will at the same time continue to use the simple approach followed here.

B. *The Effect of Specific Procedural Rules and Conditions on the Settlement Rate*

1. *Court Delay.* Court delay reduces the present value (to the plaintiff) or cost (to the defendant) of a litigated judgment in favor of the plaintiff, assuming that the parties have positive discount rates. If the value of that judgment to the plaintiff when rendered would be the same as the cost of the judgment to the defendant at that time, and if the parties have the same discount rates, the analysis of the effect of delay on the settlement rate is straightforward: delay increases the likelihood of settlement by reducing the stakes in the case. Since most civil cases are personal-injury cases, and since it is widely believed that plaintiffs in such cases generally have higher discount rates than defendants,[30] we are particularly interested in whether our conclusion holds if the plaintiff's discount rate is higher than the defendant's. The effect of such a difference in discount rates is simply to make the expected value of litigating to the plaintiff shrink more rapidly than the expected cost of litigating to the defendant, thus causing the plaintiff's minimum settlement offer to fall more rapidly than the defendant's. This implies that delay will increase the settlement rate in the important class of cases in which plaintiffs' discount rates exceed defendants'. However, delay has other effects that must be considered. Evidence tends to decay with time (witnesses die, forget, etc.), and while there are provisions in most codes of procedure for preserving evidence, they are not completely effective. Since the party with the burden of proof (especially the prosecutor in criminal cases) will be hurt more by decay of evidence than will the other party, we may analogize the effect of delay to an increase in the plaintiff's discount rate, holding defendant's constant.

[30] Presumably borrowing costs are higher for individuals, the usual plaintiffs in personal-injury cases, than for insurance companies, the real defendants in most such cases, although the price at which insurance companies could lend, and hence the opportunity costs to them of money, may be higher than for individuals. Also, an accident may give the victim an acute need for immediate cash, depending on his other resources, insurance, etc. To be sure, it also gives him an asset—the expected value of his legal claim against the injurer—against which he could, in principle, borrow. But in practice a legal claimant can rarely borrow more than his litigation expenses (through a contingent-fee agreement) against this asset. Its value as collateral is impaired by rules limiting the assignability of legal claims.

Such an increase will always increase the likelihood of a settlement by reducing the plaintiff's stakes and hence minimum settlement offer. Another effect of delay, however, is to increase uncertainty about the eventual outcome of the case if litigated, and, as we shall see shortly, an increase in uncertainty is likely to reduce the probability of a settlement.

We could make the analysis more realistic by assuming that a change in the stakes due to delay affected the litigation costs of the parties. If, for example, we assumed that parties increased their litigation expenditures as the stakes in the case increased, but at a diminishing rate, our confidence that an increase in delay would increase the settlement rate would be reduced for we would see that delay also operated to reduce the gap between litigation costs and settlement costs. This effect, however, might disappear if we assumed that settlement costs were also a function of the size of the stakes (and perhaps of litigation costs as well), and in any event would probably be small.[31]

2. *Prejudgment Interest.* A study of judicial administration by Zeisel, Kalven, and Buchholz argues that allowing winning plaintiffs interest on the judgment from the date of the accident or other event giving rise to the plaintiff's claim would not affect the settlement rate, even if the plaintiff had a higher discount rate than the defendant, because the relative value of the judgment to the parties would be unaffected.[32] They use an example where the plaintiff's expected value of litigating, apparently net of litigation costs, is 120 but defendant's expected cost of litigating, again apparently net of litigation costs, is only 100. If prejudgment interest is added, say at a rate of six per cent per annum, and delay is one year, then the plaintiff's expected value of litigation will rise to 127.2 and defendant's expected cost of litigation to 106. The ratio of the two figures is unchanged but that is neither here nor there. The gap between the offers will be larger, so, contrary to Zeisel *et al.*, the likelihood of litigation will be increased. (If, for example, each party's litigation costs were 10, and settlement costs zero, the case would be settled if prejudgment interest were not awarded, and litigated if it were awarded.) The effect of interest is opposite to that of delay: the former increases the stakes in the case, while the latter reduces them. Of course if the defendant's expected cost of litigating is higher than the plaintiff's expected benefit, the addition of interest will cause that expected cost to rise faster than the plaintiff's expected benefit; but this is a case where, as pointed out earlier, litigation is out of the question in any event.

3. *The Federal Rules of Civil Procedure.* When first promulgated in 1938,

[31] A mathematical treatment of this and the following section appears in the appendix at the end of this article. For some empirical evidence that delay leads to a higher settlement rate see William M. Landes, *supra* note 1, at 105.

[32] See Hans Zeisel, Harry Kalven, Jr. & Bernard Buchholz, Delay in the Court 131-36 (1959). Prejudgment interest is generally not awarded in tort cases.

the Federal Rules of Civil Procedure, which govern procedure in civil cases in federal district courts, represented a novel approach to pretrial procedure (one which many states have since imitated). The traditional approach had emphasized the importance of precise and detailed formulation, in the complaint, answer, and other pleadings, of the parties' claims and defenses. The Federal Rules contemplate much more summary pleadings while providing elaborate methods of pretrial "discovery" by which each party can compel the other to disclose pertinent facts in his possession.[33] Liberal pleading under the Federal Rules is related to their emphasis on discovery. The assumption is that a detailed articulation of the parties' legal theories is premature until they have had an opportunity to obtain a better knowledge of the facts.

a. *Pretrial discovery in general.* Which procedural philosophy can be expected to produce a higher settlement rate? We have suggested that a principal cause of litigation is "mutual optimism"—both parties believe they have a good chance of winning. This state will usually come about because one party (or both) thinks his case stronger than it really is or, what amounts to the same thing, think his opponent's case weaker than it really is. It may seem improbable that such mistakes would be an important source of failure to settle. It is common enough in bargaining contexts for each party to enter the negotiation with an inflated idea of the value of his performance relative to that of the other party. If I am negotiating over the sale of my house to you, I am quite likely to think the house a splendid value at my asking price and you are likely to consider it hugely overpriced. But our initial, inaccurate estimates need not prevent a bargain from being struck. I have an incentive to communicate information indicating that the house really is a good value; you have an incentive to impart information (for example, concerning prices at which similar properties are available) that will persuade me that the house is overpriced.

The legal dispute context is different from the usual bargaining context, however. A breakdown in negotiations leads, not to each party's going his separate way as in the usual commercial negotiation, but to a trial in which surprise has important strategic value. I may know something that if disclosed to my opponent would cause him to reduce his expectation of prevailing at trial, but if I disclose the information, and settlement negotiations still break down, I will have lost the value of surprising him with the information at trial.

Under the traditional approach to pretrial procedure the parties to a legal dispute had an incentive not to exchange fully information bearing on the probability of success at trial. The Federal Rules eliminated (or at least

[33] The Federal Rules also facilitate the acquisition of evidence from nonparties, but that aspect is less relevant to the question of their effect on the settlement rate.

greatly reduced) that incentive since under the rules either party can obtain most of the information in the other party's possession before trial. Indeed, an express purpose behind the rules was to put an end to the "sporting" theory of litigation by eliminating the possibility of using surprise, and in so doing they removed a significant obstacle to settlements.

But the analysis is incomplete in two respects. First, the refusal of the parties to exchange information explicitly may not prevent a significant exchange from taking place anyway.[34] The offer that a party makes during the settlement negotiations is itself evidence from which the other party can draw an inference about the information in the possession of the offeror. A series of offers and counteroffers could conceivably enable each party to reconstruct the information in the possession of the other to a high degree of accuracy. But it is doubtful whether this process would completely solve the information problem, if only because each party would have an incentive to make offers that misled the other party. If offers were influenced by this factor, not only would there be an incomplete exchange of information, but the exchange of offers—the heart of the bargaining process—would be distorted.

Second, if discovery may reduce mutual optimism by inducing the communication of information that causes a party to reduce his estimate of his chances of prevailing, so may it reduce mutual pessimism, and hence increase the likelihood of litigation, by generating information about the opponent's case that causes a party to become more optimistic. Nonetheless it can be shown that a reduction in the variance between the parties' estimates of the probability of prevailing and the true probability will, under special but seemingly plausible assumptions, increase the settlement rate.

Assume that there is a class of cases in which the stakes are 1000. Plaintiffs will have a subjective probability of prevailing ranging from zero to one and an expected value (net of litigation costs) from litigation ranging from zero to 1000. We assume that the true probability is the mean of the plaintiffs' probability distribution, as shown in Figure 4, and that there is an identical distribution of subjective probabilities for defendants. Assume further that the excess of litigation costs over settlement costs is 100 in each case and that litigation occurs only in cases where the plaintiff's minimum offer (here defined as the stakes multiplied by his subjective probability of winning) exceeds the defendant's maximum offer (the stakes multiplied by one minus his subjective probability of winning) plus the excess of the parties' litigation costs over their settlement costs.

Table 1 presents the array of plaintiffs' and defendants' offers, together with the percentage (based on Figure 4) of the parties making each offer. To determine the probability that a particular plaintiff's offer will be accepted

[34] The argument that follows was made by Fischer Black in conversation.

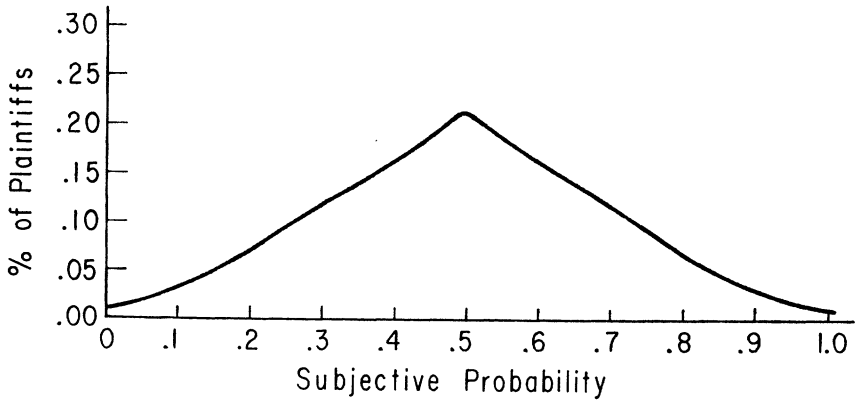

FIGURE 4

we ascertain the percentage of defendants' offers that, when added to the excess of litigation costs over settlement costs, are at least as large as the plaintiff's offer. (Thus, the probability that a plaintiff whose minimum offer is 100 will settle is one since his offer does not exceed any combination of a defendant's offer and the net litigation cost.) By multiplying each row in the second column of Table 1 by the probability that the plaintiffs in that row will settle, and summing the percentages thus obtained, we derive the average probability of settlement—69 per cent.

TABLE 1

Plaintiffs' Minimum Offers	% Making Offer	Defendants' Maximum Offers	% Making Offer	Defendants' Maximum Offers Plus Excess of Litigation Costs Over Settlement Costs
0	.01	0	.01	100
100	.03	100	.03	200
200	.07	200	.07	300
300	.12	300	.12	400
400	.16	400	.16	500
500	.21	500	.21	600
600	.16	600	.16	700
700	.12	700	.12	800
800	.07	800	.07	900
900	.03	900	.03	1000
1000	.01	1000	.01	1100
	1.00		1.00	

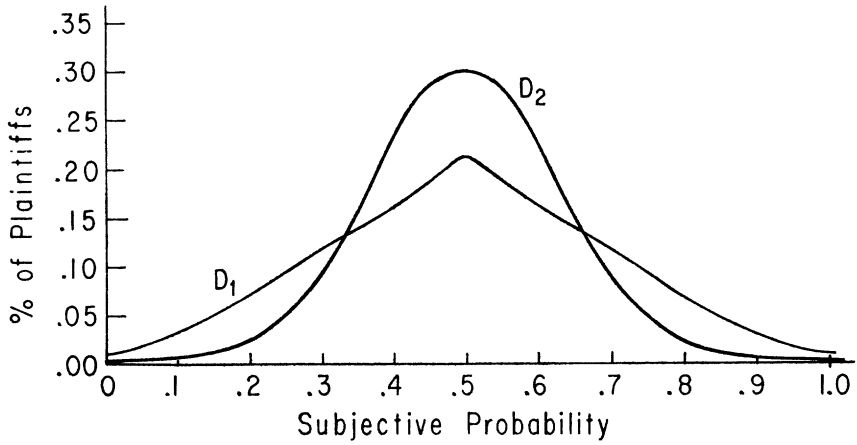

FIGURE 5

Assume now that as a result of the availability of pretrial discovery, the dispersion of subjective probabilities around the true probability of .5 is reduced, as shown in Figure 5, where D_1 is the old distribution from Figure 4 and D_2 is the new, less dispersed distribution. Transferring the coordinates of D_2 to a new table, Table 2, and solving as before, we find that the average probability of settlement rises to 80 per cent. This is so even though our distributions are symmetrical, implying that there are as many optimists as

TABLE 2

Plaintiffs' Minimum Offers	% Making Offer	Defendants' Maximum Offers	% Making Offer	Defendants' Maximum Offers Plus Excess of Litigation Costs Over Settlement Costs
0	.0004	0	.0004	100
100	.004	100	.004	200
200	.02	200	.02	300
300	.09	300	.09	400
400	.24	400	.24	500
500	.30	500	.30	600
600	.24	600	.24	700
700	.09	700	.09	800
800	.02	800	.02	900
900	.004	900	.004	1000
1000	.0004	1000	.0004	1100
	1.00		1.00	

pessimists and that the reduction in uncertainty dispels as much false pessimism as false optimism.

A discovery provision that enables both parties to improve their estimates of the outcome of the case is thus likely to facilitate settlement. But not all discovery provisions are of this sort. Two examples are the recently promulgated rule that permits the plaintiff to discover the defendant's liability insurance coverage[35] and the rule that permits the defendant, in a case in which the plaintiff's health or fitness is in issue, to have him examined by a physician designated by the defendant.[36]

b. *The insurance-coverage and medical-examination rules.* The first of these provisions cannot affect the parties' estimates of their chances of prevailing: its premise is that the information discovered would neither be admissible at trial nor lead to any admissible evidence. It can, however, affect the plaintiff's estimate of the value to him of any judgment that might be entered in his favor at trial. And anything that increases his stakes will increase the likelihood of litigation while anything that decreases them will increase the likelihood of settlement.[37]

If the defendant's insurance coverage is adequate, before discovery of this fact the plaintiff will have discounted the value of a judgment by the probability, less than one, that the judgment would be collectible because the defendant was adequately insured. The effect of discovery is to eliminate the uncertainty that led to the discounting, and thus to increase the plaintiff's stakes. Litigation is more likely. Suppose the defendant's insurance coverage is inadequate. Before discovery plaintiff will have discounted the judgment by his estimate of the probability that it would be collectible. Discovery may induce him either to raise or to lower his estimate of the expected value of the judgment. If the former, then as before the chances of litigation will be increased. If the latter, they will be reduced. However, if before discovery the defendant (more precisely, his insurer, who controls the defense) thinks that the plaintiff has overestimated the defendant's insurance coverage and that this has led the plaintiff to establish a higher minimum settlement price than if he knew the truth, the defendant will disclose the terms of the insurance policy voluntarily, in order to induce the plaintiff to reduce his minimum offer.[38] So there are few, if any, cases where the new discovery provision will

[35] Fed. R. Civ. P., Rule 26(b)(2), effective July 1, 1970. Commenting on the rule, the Advisory Committee stated without further explanation: "It will conduce to settlement and avoid protracted litigation in some cases, though in others it may have an opposite effect." Federal Rules of Civil Procedure—As Amended Through July 1, 1971, at 267 (Fed. Rules Civ. Proc. Pamph., Foundation Press, 1971).

[36] Fed. R. Civ. P., Rule 35.

[37] See Part IV(A) *supra*.

[38] Disclosure will be more advantageous from the defendant's standpoint than suffering entry of a default judgment in any case where the defendant is not completely judgment-

increase the likelihood of settlement but many cases where it will reduce that likelihood.

Let us turn to the case where the defendant is enabled by Rule 35 to form a more accurate impression of the extent of the plaintiff's injuries. By assuming that such discovery does not affect liability but only the size of the damage award, we can again focus entirely on the effect on the stakes. If upon examination by the defendant's physician the plaintiff's injuries turn out to be as serious as the plaintiff claims, the defendant will be led to increase his estimate of the expected cost to him of an adverse judgment.[39] This will make a settlement more likely. If the injuries turn out to be less serious than plaintiff claims, the defendant will be led to reduce his estimate of the cost to him of a judgment in favor of the plaintiff, which will increase the likelihood of litigation. The overall effect of Rule 35 on the settlement rate is therefore difficult to predict.

c. *Liberal pleading requirements.* We consider finally the argument that the looseness of pleading under the Federal Rules may postpone or even prevent settlement if the vagueness of the pleadings conceals a strength in the other party's case. This is unlikely to be a significant effect. Neither party has a strong incentive to attempt such concealment, since he can hardly hope to preserve surprise until trial (and so profit from the tactic). Moreover a party might use the looseness of federal pleading requirements to disguise the weakness of his case and thereby make the other party more pessimistic about *his* chances than if the weakness had been disclosed in the pleadings. This would increase the likelihood of settlement but another consequence would be to increase the costs of settlements. Parties have an incentive to conduct discovery proceedings to probe for weaknesses in the opponent's case that the looseness of the federal pleading requirements may enable the opponent to conceal in his pleadings.

On balance the Federal Rules should increase the settlement rate, although particular provisions may not.[40] But they may also increase the average costs of settlement.

---

proof. Suppose that plaintiff's claim is for $1000 and defendant has only $100 in assets to satisfy the judgment. If a default judgment is entered, plaintiff will collect the full $100. If before entry of the judgment the defendant convinces the plaintiff that he cannot pay more than $100, the plaintiff may agree to accept a settlement for less than $100 in order to avoid the cost of obtaining and collecting the judgment.

[39] See Part IV(A) *supra*. To be sure, the plaintiff would be happy in this case to communicate the true extent of his injuries. But it is more difficult for him to do this than it was for defendant in our previous example to prove to the plaintiff the terms of the defendant's insurance policy. The plaintiff may be reluctant to submit to an examination by the defendant's doctor since, unlike the insurance case, the results will be admissible at the trial if settlement negotiations fail.

[40] The only empirical study of discovery of which I am aware, Columbia University Project for Effective Justice, *supra* note 3, found no effect on the settlement

4. *Reimbursement of Winning Party's Attorney's and Witness Fees.* An interesting question is the probable effect on the settlement rate of adopting the English (and Continental) practice of requiring the losing party in a litigation to reimburse the winning party's attorney's and witness fees. The condition for litigation would not on average be altered,[41] but it does not follow that adoption of the English rule would have no effect on the settlement rate. The rule increases the variance of the expected value of litigation to each party. Under the American rule a plaintiff who wins receives a net benefit equal to the difference between his stakes and his litigation expense; one who loses sustains a net loss equal to his litigation expense. Under the English rule he receives a larger net benefit if he wins and sustains a larger net loss (equal to twice his litigation expense, assuming purely for simplicity that the parties' litigation expenses are the same) if he loses. The analysis for defendant is symmetrical. The greater variance of returns under the English rule makes the expected value of litigation less for risk-averse litigants, which will encourage settlements if risk aversion is more common than risk preference.[42]

Another effect of the English rule is to reward the party who correctly estimates that he will prevail in litigation and to penalize the party who incorrectly estimates that *he* will prevail. With the costs of guessing wrong on the outcome thus higher, the dispersion of subjective probabilities about the true probability of prevailing should be reduced, leading, as we have seen, to a higher settlement rate.

Two observations will complete the discussion of settlements in this part. The first is a reminder of the limitations of the approach to the settlement

rate. See *id.* at I-14 to I-15. However, the study contains no data from which an inference about the effect of the availability of discovery on inducing settlement prior to the institution of a formal action can be drawn. All the study found was that cases in which there is discovery are no more likely to be settled than cases in which there is no discovery. Excluded from the study were disputes that never became "cases" because they were settled before a lawsuit is commenced. Perhaps as significant is the study's finding that 95 per cent of the lawyers interviewed said that discovery increases the likelihood of settlement.

[41] The condition for litigation—inequality (1) in note 29 *supra*—would become

$$P_p(J_p + C_p) - C_p - (1 - P_p)C_d + S_p > (1 - P_d)(J_d + C_p) + C_d - P_dC_d - S_d. \tag{3}$$

If we assume that, on average, $J_p = J_d$ and $P_p = 1 - P_d$, the condition becomes

$$C_p + C_d - (S_p + S_d) < 0, \tag{4}$$

which is identical to inequality (1) if simplified in accordance with the same assumptions.

[42] John P. Gould, *supra* note 28, argues that risk preference is unlikely to increase the litigation rate since parties should be able to find cheaper methods of indulging a taste for gambling. However, the unenforceability of gambling contracts may make the costs of alternative methods of gambling quite high.

process that we have been using. After considering the determinants of litigation expenditures in the next part, we consider some other approaches. The second observation is that while measures that increase the settlement rate may in general reduce the direct costs of legal dispute resolution, they need not always have this effect. Suppose the settlement rate reached 100 per cent. With no trials at all to generate information about the relative efficiency of different substantive legal rules, the error costs of the legal system would be very high. Less obviously, the costs of settlements might also be very high, since trials are a source of information about likely outcomes of litigating that parties use in calculating expected values of litigation and hence minimum and maximum settlement offers. With the settlement rate today about 98 per cent in automobile accident cases,[43] the most numerous civil cases in our courts, we may be in, or approaching, the region where further increases in the settlement rate would reduce the efficiency of the legal system.

V. Expenditures on Litigation

A. *Determinants of Parties' Expenditures on Litigation*

1. *Competition and Collusion.* At the point where settlement attempts fail and the parties decide to litigate, it is tempting to switch from the cooperative model of legal dispute resolution (in which the parties are viewed as attempting to work out a mutually advantageous contract) of the previous part to a competitive model in which each party is viewed as expending resources on litigation in much the same way as a seller expends resources on advertising—in order to persuade the "customer" (the tribunal) of the superior merits of his "product" (case). So abrupt a shift of emphasis would be difficult to justify, however. One reason why a competitive model is appropriate in the case of advertising expenditures is that competing sellers are forbidden to agree to limit those expenditures. Sellers would often be better off if they were permitted to negotiate mutual reductions in advertising and such agreements might be commonplace were it not for antitrust policy. The case of litigation expenditures is similar but in this case agreements to limit competition are condoned, indeed encouraged, by public policy. They are in fact common: it is the rare case where there is no cooperation between the litigants' attorneys to reduce the expense of litigation. To be sure, agreements to limit litigation expenditures as such would be costly to enforce, and are

[43] See references in Richard A. Posner, A Theory of Negligence, 1 J. Leg. Studies 29, 35 n.14 (1972). A recent survey found a somewhat higher rate of litigation in medical malpractice suits (about six per cent), an increasingly important area of personal-injury law. Medical Malpractice, *supra* note 29, at app. p. 14. The higher rate may reflect the fact that medical malpractice is a relatively new and untested area of law compared to automobile accident cases, and so may lend support to the hypothesis that uncertainty encourages litigation.

rare. But the purpose of such agreements can be accomplished indirectly (and much more cheaply) by agreements to dispense with proof of particular facts, to limit the number of witnesses, etc., and such agreements are common.

The difference in the law's treatment of advertising and litigation has a possible economic explanation. Agreements limiting advertising are suspect because they impose heavy costs on consumers, in the form of reduced product information, that are unlikely to be taken into account by the parties to the agreement. An agreement limiting expenditures on litigation will ordinarily not impose significant costs on nonparties to the agreement. To be sure, there is a point beyond which a reduction in litigation expenditures would greatly increase the costs of error in the adjudicative process, which are borne primarily by nonparties to the particular dispute. And these costs are similar to the costs to consumers in our advertising example—they are also information costs. But it is a fair guess that the value of the additional information that would be produced if settlements were forbidden would be less than the cost of the additional resources that the resulting volume of litigation would consume: litigation is a more costly method of producing information than advertising. This is not to say that, were there no public subsidy of litigation (judges' salaries, etc.), the equilibrium settlement rate would be 100 per cent. As the settlement rate approached that level the costs of settlement, which depend in part on the information that is available about likely outcomes if the case is tried, would rise. Uncertainty about probable outcomes would also increase, leading, as we have seen, to a reduction in the likelihood of settlement. But the equilibrium might be higher than the social optimum since the parties to a settlement do not take account of the costs to nonparties. Thus the public subsidy of the courts *may* be justifiable—although it is always necessary to remind that a failure of the market to produce optimum results is not a sufficient condition for governmental intervention.

Despite the prevalence of "collusion" in the process by which parties to a lawsuit decide how much to spend, negotiation will sometimes fail—we have indicated a possible source of high transaction costs in the difficulty of policing an agreement to limit expenditures—so the addition of a competitive model seems indicated.

2. *A Cournot Approach.* Presumably each party chooses the level of investment in the litigation that maximizes the expected value of the litigation to him, which is equal to the stakes in the case multiplied by the probability of prevailing, minus the costs of litigation. The probability of prevailing is a function of what the party spends, what his opponent spends, and various exogenous factors (such as the state of the precedents and the availability of evidence) that weight the effect of expenditures by either party on the probability of a particular outcome. To determine each party's optimum expenditure would require that we specify the precise relationships among the relevant

variables, which we will not attempt to do (except in the appendix at the end of the article). For our purposes it is sufficient to note that each party will seek to equate the marginal product of the resources he invests in the litigation in enhancing the expected value of the litigation to their marginal cost (which we can assume to be constant) and that this marginal product will be greater, and hence the party's expenditures on litigation greater, the larger the party's stakes in the case and the more favorable the law or the evidence is to him; either circumstance will tend to make a dollar of additional expenditures on the litigation more productive. The effect on his optimum expenditure of the level of expenditures chosen by his opponent is less clear-cut. An increase in the opponent's expenditures may induce him to increase his own to overcome their effect or it may so reduce the value of his own expenditures as to induce him to reduce them. Which effect dominates depends on the precise form of the model and the specific values of its parameters.

Because each party's optimum expenditure is in part a function of what the other party spends, there is no equilibrium level of expenditure. If, however, we assume that each party does not consider the impact of his expenditures on the other party's expenditures, an equilibrium level is reached, corresponding to Cournot's solution to the problem of the equilibrium price and output under conditions of noncollusive duopoly. Although the Cournot solution has been strongly criticized in its original context, the major criticisms—that the parties could do better by colluding and that its decision rules are arbitrary—are less applicable to litigants' competition. We have seen why agreements limiting the level of expenditures are difficult to make; and a procedure by which each litigant, in determining his optimum expenditure level, adjusts to successively better estimates of what the opponent is likely to spend (the procedure implied by the adoption of the Cournot approach) seems about as natural as any alternative decision rules they might employ, failing agreement.

3. *Some Implications of the Cournot Approach.* The Cournot approach implies, quite reasonably, that an increase either in the plaintiff's stakes or in the effectiveness of his litigation expenditures, or a decrease either in defendant's stakes or in the effectiveness of *his* litigation expenditures, will induce the plaintiff to spend at a higher rate than the defendant, and vice versa. When the stakes to the parties are the same, the ratio of their litigation expenditures will be positively correlated with the ratio of their subjective probabilities of prevailing if they made the same expenditures. This helps to explain why, as mentioned earlier, we can expect most cases to be decided correctly even if the plaintiff need establish his case only by a bare preponderance of the evidence. If the allegations essential to one party's claim are in fact true, ordinarily it will be easier for him to prove them than for his opponent to disprove them, assuming they spend the same amount of money on

the trial. Stated another way, the effectiveness of the expenditures of the party with the meritorious claim will be high relative to the effectiveness of his opponent's expenditures. This will induce the first party to spend more heavily on litigation than the second, which in turn will increase the first party's subjective probability still further and reduce the second party's still further. If we assume that the parties are equally good estimators, so that the objective probability of prevailing is equal to the mean of the plaintiff's subjective probability of prevailing plus one minus the defendant's subjective probability of prevailing, the objective probability that the party with the meritorious claim will prevail will be even higher than if the parties' expenditures on litigation were fixed at the same level. This reinforces an earlier point that even if the prosecutor did not bear so heavy a burden of proof, the probability of convicting an innocent defendant would be small.

Our adoption of a Cournot approach implicitly rejects the importance of bluff or gaming in decisions on litigation spending. We assume that the parties do not attempt to influence each other's behavior by threatening to use overpowering resources. In principle the use of such threats might be attractive to one or both parties, since, as mentioned earlier, if the defendant (say) could persuade the plaintiff that he was going to spend very heavily, the plaintiff might be led to reduce his expenditures. This would increase the likelihood of the defendant's winning at the pre-threat level of expenditures. Thus, if the bluff worked, the defendant would not have to carry out the threat to overspend). He might actually spend *less* than he had originally planned, since a reduction in one party's expenditures (here induced by the other party's threat to overspend) may, as suggested earlier, lead the other party to reduce his expenditures.

One would be surprised, however, if bluff were often successful in this context. To be effective a threat must be credible: the victim of the threat must be persuaded that it is likely to be carried out if he does not yield. To persuade one's opponent that one will sacrifice more than an optimum amount of resources if the bluff fails requires persuading him of one's irrationality, which will usually be difficult to do. However, a rational threatener *involved in a sequence of similar legal disputes with different people* might carry out a threat to overspend in one dispute in order to establish the credibility of similar threats in the other cases. An analogy may be drawn to the use of predatory pricing by a firm that competes in many different markets against single-market firms. For the large firm to reduce price below cost in all of its markets at once in order to carry out a threat to break its competitors unless they sell out to it on its terms would be extremely costly. The threat to do so would therefore lack credibility. But suppose it threatens the firms in just one of its markets with below-cost prices if they refuse to come to terms. This threat is more credible. The victims know that if the predator carried out his

threat it might enhance the credibility of similar threats against firms in other markets and that the cost to the predator of carrying out the threat would be small relative to his total costs since below-cost prices would be charged in only one market. It was this more plausible version of predatory pricing that the Standard Oil Trust was alleged to have practiced, yet little evidence that it actually did so has been found;[44] indeed the paucity of cases in which predatory pricing has been proved is striking.

Pretrial discovery provides a somewhat different example of the possibility of predatory conduct in litigation. In principle, a litigant could impose heavy costs on an opponent, at little cost to himself, by demanding information in an amount and form very costly for the other party to supply. This practice appears, however, to be rare.[45] We may speculate that this is due partly to the ease with which the other party could retaliate and partly to protective provisions in the Federal Rules of Civil Procedure designed to prevent such abuses of the discovery process.

An interesting form of predation in the litigation context is the "nuisance" suit, a groundless action brought on the theory that the defendant will pay something to avoid the expense of trial. It is said that, knowing this, recurrent defendants, notably insurance companies, refuse to settle nuisance suits—will defend them even if it is costly to do so—in order to deter them. The phenomenon is worth investigating, but the theory is not coherent. In the case of the truly groundless claim the defendant knows that if he calls the plaintiff's bluff the plaintiff will not throw away good money litigating the case, and the plaintiff should know that the defendant knows this. Since the usual nuisance claimant is conceived to be an obscure individual who brings at most a few such claims in his lifetime, he cannot make his threat to litigate credible by pointing to the gains he could obtain if he established the credibility of future threats by carrying out the present one. One is led to predict, therefore, that pure nuisance claims are infrequent, that when made they are usually turned down, and that when turned down the plaintiff does not pursue the matter in court.[46]

A particularly important implication of the Cournot approach to the question of how much money the parties spend on litigation is that an increase in the sum of the parties' subjective probabilities of prevailing may reduce,

[44] See John S. McGee, Predatory Price Cutting: The Standard Oil (N.J.) Case, 1 J. Law & Econ. 137 (1958).

[45] See Columbia University Project for Effective Justice, *supra* note 3, at VII-21 to VII-22.

[46] This is not to say that there are never fraudulent claims having a sufficiently large expected value to support a credible threat to litigate if the defendant refuses to settle on terms favorable to the claimant; or claims that, while unlikely to prevail, are not so weak that they would not justify a nongaming claimant in expending some money on a lawsuit. More on this shortly. For some empirical evidence relevant to the question of nuisance suits see H. Laurence Ross, *supra* note 29, at 204-10.

rather than as heretofore argued increase, the likelihood of litigation. Suppose that, just before trial, the plaintiff discovers a precedent that he thinks significantly enhances his chances of winning at any level of expenditures. The defendant knows about the precedent but thinks it inapplicable to the case at hand. Since the marginal product of the plaintiff's litigation expenditures is now higher he will invest more money in the litigation. The gap between the parties' litigation costs and their settlement costs will increase by the amount of this additional investment, and if the defendant responds by increasing his projected outlays it will increase by the sum of these additional investments. If the gap increases faster than the difference between the expected value of the litigation to the plaintiff and the expected cost to the defendant (excluding litigation expenses), the likelihood of litigation will be reduced. It may well increase faster, since the effect of higher expenditures on the parties' subjective probabilities of prevailing (and hence on the difference between the expected value and expected cost of a litigated judgment) may be largely or entirely offset by the increased expenditures of the opposing party.

To illustrate, suppose that the stakes in a case are $10,000 and that before the discovery of the precedent the plaintiff's subjective probability of prevailing was 60 per cent, the defendant's was also 60 per cent, and each party expected to spend $900 on the litigation (we ignore settlement costs). The plaintiff's minimum offer was then $6000 — $900 or $5100, and the defendant's maximum offer was $4000 + $900 or $4900, so a settlement was impossible. Suppose that the discovery of the precedent increases the plaintiff's subjective probability of winning to 65 per cent, thereby increasing the expected value of the litigation to him by $500, and that this induces him to project an additional outlay of $400.[47] Defendant learns of the additional expenditure planned by the plaintiff, and, believing that the expenditure, unless matched, will reduce the expected value of the litigation to him by $500, also decides to spend $400 more. Each party must reassess the probability of his winning in light of the other party's change in projected expenditures. Suppose that they decide to stick by their previous estimates of 65 and 60 per cent respectively. The plaintiff's minimum offer will now be $6500 — $1300, or $5200, and defendant's maximum offer will be $4000 + $1300 or $5300. The parties will decide to settle. They could in principle decide simply to rescind their decisions to increase expenditures, but a settlement may be cheaper given the difficulty of policing an agreement to limit litigation expenditures.

[47] The precedent has the same effect as would an increase in the stakes: it increases the value to the party of an additional expenditure on litigation. If he thinks an extra dollar spent on the case would increase the probability of his winning by some fraction, anything that increases either the fraction or the stakes makes the extra expenditure more productive for him.

In sum, our conclusions in Part IV are undermined when the parties' expenditures on litigation are allowed to influence their subjective probabilities of winning. But it does not follow that those conclusions must be discarded. It is possible that, above a certain threshold, the outcome of litigation is relatively insensitive to variations in the expenditures of the parties. If so this would argue for retaining the simpler approach of Part IV. Furthermore, the present analysis depends crucially on the assumption that a party does not consider the impact of his expenditures on the other party's expenditures. If he does, then in our last example the discovery of the precedent may not induce the plaintiff to project an additional outlay; he may expect the defendant to match it, leaving both worse off than before. If no additional outlays are projected, the increase in the plaintiff's subjective probability of winning due to discovery of the precedent, will, as in Part IV, reduce the likelihood of a settlement.

B. *The Role of Procedure in Optimizing Litigation Expenditures*

1. *Economizing Procedures and Their Effects.* Many familiar procedural devices appear to be designed, in part at least, to reduce the expense of litigation. Some examples are summary judgment, judicial notice, presumptions, collateral estoppel, requests for admissions by the adverse party, allocation of the burden of pleading and of production of evidence, exclusion of evidence that is merely cumulative, and perhaps the hearsay rule. A particularly clear example is provided by the rules governing venue which are designed to place the trial in the cheapest location for the parties. But whether such devices actually reduce the amount of money spent on litigation is not obvious. If the judge, by taking judicial notice that January 11, 1973, was a Thursday, saves a party the expense of hiring a witness to testify to the fact, it does not follow that the party's litigation expenditures will be lower. The effect of judicial notice is to enable the party to develop the same amount of evidence favorable to his contentions at lower cost. Figure 6 explores the consequences. The party's demand for evidence is equal to the marginal product of evidence in enhancing the expected value of litigation to him. The intersection of the demand curve with the curve representing the supply price of evidence determines the price and quantity of evidence purchased. A reduction in the supply price induces the party to purchase a greater quantity of evidence, q_1, for which he pays a lower price, p_1. Whether pq is larger than p_1q_1 depends on the elasticity of demand between p and p_1. If it is greater than one, p_1q_1 (the party's litigation expenditures after the increase in productivity) will be larger than before; if it is one they will be the same; if it is less than one they will be smaller.

Whether demand is likely to be inelastic in the relevant region depends once

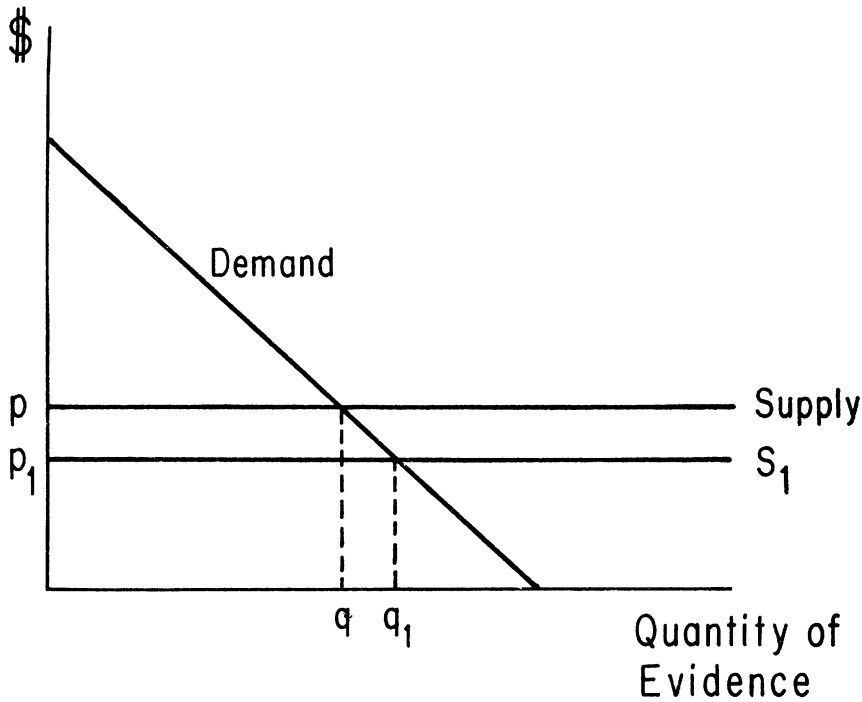

FIGURE 6

again on our assumptions about the parties' reaction patterns. If a measure that makes it cheaper for one party to establish facts favorable to his position also makes it cheaper for the other party to establish facts favorable to *his* position, the perceived marginal product of evidence may be slight, since the effect of additional evidence in enhancing the expected value of litigation will be expected to be offset by the opponent's matching purchase of additional evidence. This assumes, however, that parties are sensitive to the impact of their decisions with respect to litigation expenditures on their opponents' decisions. If the assumption is correct, the procedural devices mentioned at the beginning of this section may well reduce aggregate expenditures on litigation. But regardless of their impact on expenditures, such devices can probably be justified as reducing the error costs discussed in Parts II-III. The use of witness time to establish a fact that is clear beyond doubt does not advance the search for truth. The elimination of such a method of presenting evidence encourages the parties to increase the purchase of evidence that does dispel genuine factual uncertainties.

Similarly, the principal significance of the liberal pleading and discovery provisions that are the most distinctive features of the Federal Rules may

be that they reduce error costs, not that they reduce aggregate expenditures on litigation or increase the settlement rate. The abrogation of the traditional strict pleading requirements has probably reduced the number of meritorious cases dismissed because of a lawyer's oversight. Discovery enables each party to obtain the facts bearing on the merits of his contentions and the deficiencies of the adversary's. Both the insurance-coverage and medical-examination rules discussed earlier in terms of their effect on the settlement rate reduce error costs, the first by enabling plaintiffs to bargain to a figure closer to the true cost of the defendant's conduct and the second by enabling more accurate assessment of the plaintiff's actual damages.

2. *Reimbursement of Attorney's Fees.* The English rule (reimbursement by losing party of winning party's attorney's and witness fees) has been criticized on the ground that it encourages the parties to a lawsuit to spend more than they would under the American rule.[48] Were this criticism sound it would still be unclear whether aggregate expenditures on legal dispute resolution were larger or smaller under the English than under the American rule; but it is unsound. Since the English rule, as we have seen, does not alter the expected value of litigation to the parties, it should not induce the parties to increase their litigation expenditures. What is likely is that the English rule deters, because it penalizes, expenditures generated by unjustified optimism as to the probability of prevailing in a lawsuit. It thus reduces the incidence of the marginal claim.

To explain, in order for a plaintiff to be able to make a credible threat to sue, he must convince the defendant that the expected value of the judgment, if the plaintiff wins, is greater than the plaintiff's expected litigation expenses; that is, $PJ > C$.[49] Since under the English rule (assuming for simplicity that the parties' litigation expenses are the same) the expected value of the judgment to the plaintiff is $P(J + C) - (1 - P)C$, the condition for a credible threat to litigate is $P(J + C) - (1 - P)C > C$, or, simplified, $PJ > 2C - 2PC$. If the probability of winning is very small, the last term of this inequality approaches zero and the expected value of the judgment to the plaintiff must be almost twice as great as under the American rule in order to justify a claim. Conversely, under the English rule fewer meritorious claims will be abandoned because the cost of litigation exceeds the expected value of the claim. When the probability of winning is one, the condition for a credible

[48] See Arthur Goodhart, Current Judicial Reform in England, 27 N.Y.U.L. Rev. 395, 405-06 (1952).

[49] The assumption here, a refinement of our previous discussion of the determinants of expenditures on litigation, is that there is a threshold of expenditures below which a party cannot prevail even if the merits of his claim are clear and the defendant is prepared to spend very little to contest it.

threat to sue is $J > C$ under the American rule but $J > 0$ under the English rule, implying that highly meritorious claims will rarely be frustrated merely because the absolute size of the claim is small. This has important implications for the efficiency with which the judicial system performs its allocative function. In cases where litigation expenses are substantial in relation to the value of an individual claim, the probability of error under the American rule approaches one: the defendant is never brought to bar. Yet if the defendant inflicts injury on many different people, aggregate error costs may be very high even though the individual injury is small.

Would reimbursement be equally sensible in criminal cases? One can argue that the state should not be required to reimburse the expenses of the acquitted defendant. Since as we saw in Part III the probability of convicting an innocent person charged with crime is very low, a substantial fraction of the acquitted probably are guilty, so the principal result of reimbursing the legal expenses of acquitted defendants would be to increase the gains from crime. And providing such reimbursement would supply an argument for relaxing the prosecutor's burden of proof.

C. *The General Problem of Access to Legal Remedies*

The preceding section may have created the impression that if the English practice of reimbursing the winning party's attorney and witness fees were adopted in this country, the problem of the meritorious small claim defeated by the litigation-expense threshold would be solved. It would not be solved. To begin with, the English rule does not fully reimburse the expenses of suit of the winning party. The most important exclusion is the opportunity cost of the party's time, which explains why very small claims would rarely be prosecuted even if the probability of prevailing were one and the plaintiff's outlays were reimbursable with interest at his personal discount rate. The rule could be changed but unfortunately the measurement of the opportunity costs of a party's time would be difficult or impracticable in many cases.

In any case where the plaintiff's probability of winning is less than one, the expected utility of litigation may be negative if the party is risk averse, even if the expected value of the litigation is positive. As mentioned earlier, the English rule actually increases the riskiness of litigation. The rule does not attempt to compensate the risk averse for their aversion, nor would it be practical to do so.

All else aside, the English rule does not create an incentive to sue on every meritorious claim, however small, unless meritorious is defined to mean that the probability of winning is one. When it is less than one, even if only slightly less, the expected value of litigation may well be negative. If the condition for willingness to sue under the English rule ($PJ > 2C - 2PC$) is rewritten

as $J/2C > (1 - P)/P$, it becomes easy to see that even if the probability of winning is .99, a plaintiff will sue only if the stakes are more than one-fiftieth as large as his litigation expense. If the stakes are $1 and his litigation expense is $60, he will not sue; yet the aggregate social costs from inability to enforce claims of this size may be very great.

In some cases the English rule may induce socially unjustified litigation—litigation that costs more than the errors that would result from failure to enforce a class of meritorious claims. Let J in the last inequality be $100, C $150, and P 80 per cent. The expected value of litigation to the plaintiff is now positive, and he will sue if the defendant refuses to settle on acceptable terms; but the cost of litigation to the plaintiff alone is greater than the social cost of not enforcing the claim, which we assume to be equal to the stakes. The example is somewhat misleading, however. The plaintiff's threat to litigate, which the English rule makes credible, should succeed in forcing settlement in most cases. The proper comparison is between the stakes in the individual case multiplied by all the cases, litigated and settled, in which plaintiff prevails and the litigation expense in the individual case multiplied by the number of cases actually litigated.

The English rule is sometimes criticized on the ground that it discourages poor people from suing[50] (or defending against suits?). The criticism is a little hard to fathom. Since the rule reduces the expected cost of litigation for any litigant who has a better than 50 per cent chance of prevailing, one would expect it to help poor people who have meritorious claims or defenses. The criticism would have some basis, however, if litigants were required to furnish, as a precondition either to maintaining or defending against a legal claim, evidence of ability to reimburse the opponent's attorney's and witness fees should the opponent win. Such a requirement would present difficulties for poor people, especially when they were defendants; even when they were plaintiffs and had meritorious claims, they might have difficulty establishing their ability to reimburse the opponent's legal expenses unless the claim itself was considered appropriate collateral. If the requirement is waived, the poor are helped but the deterrent effects of the rule are weakened.

While the English rule has many attractive properties, it does not provide a complete solution to the problem of the small claim. The modern class action may be a more promising solution. The Federal Rules of Civil Procedure now explicitly sanction the use of the class action as a means of achieving economies of scale in litigation by pooling a large number of similar claims.[51] The effect of such pooling is to lower the litigation-expense threshold. Suppose that there are 100 identical claims, each having an expected value if

[50] See Arthur Goodhart, *supra* note 48.

[51] See Fed. R. Civ. P., Rule 23(b)(3), effective July 1, 1966; Advisory Committee's Note, in Federal Rules of Civil Procedure, *supra* note 35, at 240-42.

litigated of $10, and the minimum cost of each plaintiff's litigating his claim would be $50. The claims will not be brought separately. Since they are identical, presumably much of the work of the lawyers in the 100 cases, had they been brought, would have been duplicative. This implies that if the cases could be merged together into one case, the cost of litigation would be less than 100 times the threshold cost for each case, probably much less.

Rule 23 and the Advisory Committee's Note correctly emphasize the number of claimants and the commonality of the claims as factors bearing on the desirability of permitting a class action to be maintained. The more numerous the claimants, the greater the amount of duplication that the class action would eliminate; the more homogeneous the claims, the more likely that if the claims were litigated separately the efforts of the lawyers in the cases would involve substantial duplication.

Two sources of weakness in Rule 23 are the explicit requirement that notice be given to all members of the class and the implicit requirement that the judgment in the case (if it is a damages judgment) be distributed to the members. In cases where the members are very numerous and the damages per member very small, both requirements can be immensely costly. A common response has been to fictionalize the requirements, as by permitting notice via newspaper advertisements that most members of the class are unlikely to read or by depositing the proceeds of the judgment in a trust fund and placing the burden on each member of the class to come forward and claim his (often negligible) share. This is not a bad response when the purpose of the substantive law being enforced by the class action is to promote efficiency. From the standpoint of efficiency the important point is that the defendant be made to pay the social costs of his unlawful (inefficient) conduct, rather than that the judgment be paid to a particular group of people—although failure to compensate might lead victims to adopt inefficient methods of loss prevention which, to the extent that they were effective in preventing loss, would eliminate the incentive of injurers to take what are by hypothesis more efficient preventive measures (if they were less efficient, presumably liability would have been imposed on the victims rather than on the injurers in the first place). But when the purpose of the underlying law is to transfer income to a particular group rather than to prevent inefficient behavior, the costs of a class action may represent a pure waste of resources unless a significant distribution to the intended beneficiaries is made. This is another example of the importance, in measuring legal error, of properly characterizing the substantive law being enforced as either allocative or distributive in basic purpose. If the latter, the use of the class action as a device for the pooling of very small claims should be avoided.

Class actions have another weakness. There is a potential conflict of interest between the lawyer for the class and the members of the class. Since the lawyer is presumably interested in his fee rather than in the judgment as

such, he will be tempted to settle with the defendant, on behalf of the class, for a small judgment plus a large attorney's fee. Such an arrangement will often be attractive to the defendant too, since he is interested in the total payout rather than in how it is divided between the lawyer and the class. The losers are the class members, but if their claims are small, none of them will have an incentive to exercise sufficient supervision over the lawyer to assure that the settlement will be in the interest of the class. Society also loses insofar as the conflict of interest leads to a systematic underestimation of the damages in class actions. The requirement of Rule 23 that a class action may not be dismissed or compromised without the approval of the court is evidence of recognition of the conflict of interest problem. But whether it is an effective control may be doubted. A court is normally dependent for its information on the parties' lawyers, who in this instance have an incentive to collude to withhold full information from the court. However, the conflict of interest problem may be solved by competition among lawyers. If a lawyer can persuade a member of the class to hire him to object to the settlement agreed to by the lawyer for the class, the judge will have an independent source of information on the merits of the proposed settlement.

VI. The Interaction Between Error Costs and Direct Costs

The relationship between error costs and direct costs can be summarized in a loss function having three terms. The first term is error cost. This is a function of the probability of error, which in turn is a function of the fraction of cases litigated, the amount of private expenditures on litigation, and the amount of public expenditures.[52] The second term is the sum of the private and public expenditures in cases that are litigated, and is equal to the total of those expenditures in all cases multiplied by the fraction of cases litigated. The third term is the total expenditures (all private) on cases that are settled, and is equal to the total private expenditures in all cases multiplied by the fraction of cases settled multiplied by the fractional cost of settling rather than litigating. An increase in the fraction of cases litigated, or in the public or private expenditures on litigation, will reduce the probability and hence cost of an erroneous judicial determination. An increase in public expenditures on litigation will reduce the relative cost advantage of settling rather than litigating (the government's subsidy of litigation has increased), and an increase in the relative cost advantages of settling will reduce the fraction of cases tried.

These relationships make clear why it is difficult to predict a priori the

[52] On judges' and clerks' salaries, court house buildings, juries, etc. Although normally assumed to be a small part of the overall costs of the legal dispute resolution system, they are by no means trivial. In fiscal year 1971, the budget of the federal court system alone was $174 million. Judic. Conf. of the U.S., Ann. Reports of Proceedings, 1971, and Ann. Rep. of the Dir. of the Adm. Office of the U.S. Courts, 1971, at 209 (1972).

effect on overall efficiency of changes in the relevant variables. For example, an increase in the fraction of cases litigated will increase the social costs of legal dispute resolution only if the difference between the total costs of litigating cases and the total costs of settling them is greater than the reduction in error costs brought about by increasing the fraction of litigated cases. Otherwise it will reduce the total costs of legal dispute resolution. An increase in public expenditures will reduce error costs both directly and by inducing a larger fraction of cases to be tried, but it will increase the total direct costs of legal dispute resolution both directly and by making litigation relatively more attractive than settlement. Thus there can be no presumption that increasing the public expenditures on the court system will increase social welfare. An expenditure of another $1 million on the court system might cost society several millions—or benefit society by several millions. Finally, an increase in the fractional cost of settlement versus litigation, by lowering the cost of settlement relative to that of litigation, will reduce the direct costs of legal dispute resolution but indirectly increase the error costs. Thus, as argued earlier, measures that increase the attractiveness of settlement in comparison to litigation cannot be regarded as unequivocally desirable.

B. *Some Examples of the Interaction*

1. *Discovery Again.* Pretrial discovery provides many interesting examples of the complex interplay between error and direct costs. We discuss two. The first involves the question whether surveillance evidence (the defendant in a personal-injury case may be keeping watch on the plaintiff to see whether he really was crippled by the accident, as he contends) in the possession of one party may be discovered by the other. Discovery of such evidence would facilitate settlement—but only in those cases where the surveillance confirmed the plaintiff's claim. It might conceivably reduce litigation expenditures by making it unnecessary for the plaintiff to ascertain whether he was under surveillance. In some cases it might reduce the likelihood of an erroneous determination by enabling the plaintiff to prepare a truthful and convincing explanation of an apparent discrepancy revealed by the surveillance. Nonetheless the courts deny discovery of surveillance evidence, on the ground that surprise in its use, by enhancing its effectiveness, is a valuable penalty for fraudulent claims.[53]

The second example involves discovery of the facts and opinions of the other party's experts. A plausible case for such discovery can be made in terms of reducing error costs, avoiding duplication of costly expert preparations, and facilitating settlement by reducing mutual uncertainty. But there

[53] See, *e.g.*, Stone v. Marine Transport Lines, Inc., 23 F.R.D. 222 (D. Md. 1959); but cf. Fed. R. Civ. P., Rule 26(b)(3), effective July 1, 1970 (party may discover own prior statement).

is a problem of freeloading. A party might decide not to hire expert witnesses in the hope that his opponent would do so, so that he could obtain the benefit of their expertise for nothing. If both parties thought this way, no experts would be hired and the likelihood of a correct determination at trial would be reduced. The appropriate solution would be to permit discovery but require the party requesting it to pay the expert a reasonable fee. The Federal Rules are moving in this direction.[54]

2. *Substantive Law Reform.* Often, a major part of the case for a proposed change in substantive law is that it will reduce litigation costs. A good example is the proposal to replace negligence in many areas of tort law by strict liability. By eliminating a major issue—the defendant's liability—this change might be thought to reduce the costs of tort trials and also, by reducing a source of uncertainty, to increase the settlement rate. The impact of the change on both error costs and direct costs must be considered. Replacement of negligence by strict liability would probably not affect the allocative efficiency of the tort system (and therefore increase error costs), so long as a defense of contributory negligence continued (under whatever name) to be recognized.[55] One effect on direct costs would be to increase the number of claims since the liability of injurers would be more extensive. The total costs of the legal dispute resolution process would therefore rise unless the average cost of disposing of a claim fell far enough to offset the effect of the larger number of claims. The average cost might fall. The elimination of a significant issue (negligence) should reduce mutual uncertainty; this might—or might not—increase the settlement rate. The probable effect on the parties' expenditures in litigated cases is highly uncertain. The expected value of litigation to the plaintiff would be higher at any level of expenditure than under a negligence standard and the expected value to the defendant lower, leading, probably, to an increase in the plaintiff's expenditures on litigation and to a decrease in the defendant's. The sum of their expenditures might not change. And even if it declined, the result might be an increase in the rate of litigation (because the gap between litigation and settlement costs was now smaller), assuming that settlement costs did not decline as much. Finally, even if we were completely confident that the average cost of disposing of a claim under strict liability would decline, we could only speculate as to whether the decline would be great enough to offset the increase in total costs due to the larger number of claims.

The effect of "no fault" schemes of automobile accident compensation on the costs of legal dispute resolution is more straightforward. By extinguishing liability altogether in a large class of cases, such schemes should reduce the

[54] See Fed. R. Civ. P., Rule 26(b)(4), effective July 1, 1970.

[55] See Richard A. Posner, *supra* note 5.

number of legal disputes and hence the costs both of settlement and of litigation. (There will be some offsetting effect, however, by virtue of higher negotiation costs and more disputes stemming from the more extensive liability of insurance companies to their policy holders under no-fault plans.) At the same time, by reducing the deterrent effect (if any) of tort liability on careless driving, no-fault schemes may increase the accident rate. The additional accident costs may be viewed as error costs resulting from the abandonment of a system of liability. Whether the savings in direct costs are likely to be as great as the additional error costs is conjectural.

3. *The Jury.* A controversial possibility for economizing on the expenses of legal dispute resolution would be the abolition of the jury, at least in civil cases. Its costs are apparently not trivial, a major part (and one largely concealed from public scrutiny) consisting of the opportunity costs of jurors whose time is more valuable than the compensation they receive as jurors.[56] The abolition of the jury system would not require the hiring of additional judges since a judge presides at every jury trial anyway, and hence would reduce the direct costs of dispute resolution. The case for abolition has been (unwittingly) strengthened by the findings of the University of Chicago Jury Project of substantial agreement between judge and jury,[57] which implies that outcomes would be largely unaffected by the jury's disappearance.

The jury has a special function in a criminal case: to enforce the reasonable-doubt standard. A judge might in some cases be reluctant to acquit a defendant who he was convinced was guilty, although not 95 per cent convinced. If 12 jurors are required to be unanimous for conviction—the usual rule—the likelihood of a conviction if the prosecution fails to carry its burden of proof is greatly reduced.[58]

4. *Res Judicata.* A firm launches an advertising campaign for one of its products. The campaign is a failure; it makes no inroads on the sales of competing products. The firm can if it wishes launch a new (and identical) campaign. Its decision will be affected by the knowledge of the first failure but there is no law against trying as often as it wants. However, a plaintiff who loses a case is barred by the principle of res judicata from bringing a second case against the defendant based on the same claim. In one sense

[56] See Donald L. Martin, The Economics of Jury Conscription, 80 J. Pol. Econ. 680 (1972). Martin estimates that the total opportunity costs of jurors in the American legal system were $232 million in 1962. *Id.* at 692 (tab. 3). This may be an exaggeration. The ease of avoiding jury duty suggests that many jurors may derive nonpecuniary income from jury duty. This would have to be subtracted from the opportunity costs of their time, in calculating the total costs of the jury system.

[57] See Harry Kalven, Jr. & Hans Zeisel, The American Jury (1966). The study was limited to criminal cases.

[58] See Herbert Friedman, Trial by Jury: Criteria for Convictions, Jury Size and Type I and Type II Errors, 26 Am. Statistician, April 1972, at 21.

there is less justification for such a principle than there would be for a direct limitation on the amount that parties might spend on a lawsuit. The first loss should discourage the plaintiff from expending additional resources in trying to vindicate his claim; legal prohibition seems largely superfluous. However, while the savings in direct costs from the rule may not be very great, the error costs imposed by the rule are probably zero, so the rule is an appropriate economizing measure.

The social purpose of product competition is to maximize the welfare of consumers in accordance with their preferences. Since those preferences change from time to time, there is no anomaly if an advertising campaign that once failed later succeeds. But the social purpose of litigation (so far as relevant to this discussion) is to ascertain whether certain alleged events occurred. They either did or did not occur. It would be a contradiction to say that the first time a case was litigated, and plaintiff won, the decision was correct and the second time the case was litigated—when defendant won—the decision was also correct. Moreover, unless there is some reason for thinking that a second decision is more likely to be correct than the first, it cannot be shown that having two trials rather than one increases the accuracy of the dispute resolution process. Consider: *A* brings a suit against *B* and loses. *A* brings a second suit, and wins. *B* now sues to recover the judgment obtained by *A*, contending that the second decision was erroneous, and wins. *A* thereupon commences a new suit against *B* . . . . Wherever we decide to break the chain, we will have no idea whether the last decision was correct, or some previous one. Thus the expected value of relitigation in enhancing the accuracy of the adjudicative process is (in general) zero.

5. *Delay in Court.* To most experts in judicial administration, delay between the filing and final disposition of a legal claim is an unmitigated evil and the proper focus of judicial reform.[59] This is an odd way to look at the matter. Delay is an omnipresent feature of social and economic life. It is only excessive delay that is undesirable, and what is excessive can be determined only by comparing the costs and benefits of different amounts of delay.[60]

A major cost associated with queuing as a method of rationing goods is the opportunity cost of the time people spend in the queue. Where the parties' time is their own while they wait (as when a theatergoer is forced to "wait"

[59] See, *e.g.*, Hans Zeisel, Harry Kalven, Jr. & Bernard Buchholz, *supra* note 32, at xxii; A. Leo Levin & Edward A. Woolley, Dispatch and Delay 115, 119 (1961).

[60] The relevance of costs is rarely considered. For example, the Administrative Office of the U.S. Courts is proud that delay in the federal courts has grown little in the last ten years (see Judicial Conf., etc., *supra* note 52, at 136-38), but does not advertise the fact that during this period the budget of the federal courts grew threefold. Compare Judic. Conf. of the U.S., Ann. Reports of Proceedings, 1962, and Ann. Rep. of the Dir. of the Adm. Office of the U.S. Courts, 1962, at 151, with Judicial Conf., etc., *supra* note 52, at 209.

for six months to see a popular musical), the queue is merely a "figurative" queue.[61] The court queue is a literal queue for defendants incarcerated awaiting trial and for some owners of property "tied up" in litigation. Otherwise it is a figurative queue, but this does not mean that it is costless. Court delay increases error costs because the adaptation of legal rules to altered circumstances is retarded and because evidence decays over time, increasing the probability of an erroneous decision. Clearly, at some level of delay error costs would become prohibitive. Delay also increases error costs by widening the gap between damages and judgments that is created by the fact that the legal interest rate is lower than the market rate and interest is usually allowed not from the date of the event giving rise to the suit but only from the date of judgment. This particular source of error cost from delay could be eliminated simply by increasing the interest rate and computing interest from the date of violation.

Delay is also a source of benefits. Presumably it enables a reduction in the number of judges and other court personnel, court houses, etc. It may increase the settlement rate.[62] A hidden benefit of delay in administrative proceedings is the additional incentive it gives regulated firms to increase technical efficiency between rate proceedings.[63] The analysis of the costs and benefits of delay is especially complicated in criminal cases. One effect of delay is to increase both the punishment costs and litigation costs of defendants not admitted to bail; another effect is to reduce the punishment costs of those defendants who are admitted to bail.[64]

Whether existing levels of delay are optimum is very difficult to judge, in part because the usual statistics of delay do not measure the court queue—the waiting period—accurately. Delay is generally measured from the filing of the defendant's answer to the complaint to the final disposition of the case. This interval is too long because it includes time during which the parties are not waiting at all, but litigating or preparing to litigate or attempting to negotiate a settlement. It is too short because it excludes the period between the event giving rise to the legal dispute (or the earliest time when a settlement might have been made) and the filing of the answer.

In 1972, the average interval between answer and final disposition, in personal-injury cases tried before juries in state courts, was 21.7 months,[65]

[61] See Gary S. Becker, A Theory of the Allocation of Time, 75 Econ. J. 493, 515 (1965).

[62] See Part IV(B)(1) *supra*.

[63] As emphasized in William J. Baumol, Reasonable Rules for Rate Regulation: Plausible Policies for an Imperfect World, in Prices: Issues in Theory, Practice, and Public Policy 108 (Almarin Phillips & Oliver E. Williamson eds. 1967). But in a period of inflation delay may operate to reduce regulated firms' rates below cost, creating inefficiency.

[64] See William M. Landes, The Bail System: An Economic Approach, 2 J. Leg. Studies 79 (1973).

[65] Institute of Judicial Administration, Calendar Status Study—1972, at vi (Aug. 1,

an increase of only three months since 1963.[66] Delays in other kinds of cases, and in the federal courts, appear to be substantially shorter.[67] The situation in a few major cities, however, is a good deal worse.[68] Statistics that actually measure either the court queue or the costs and benefits of court queues of different length are unavailable.

The marked difference in waiting times between jury and nonjury trials is interesting because it suggests that courts are encouraging the choice of the cheaper method of trial by subjecting the more expensive to a much longer queue. A more straightforward method of accomplishing this end would be to charge a substantial fee for plaintiffs demanding a jury trial. The use of price as a method of rationing access to the courts would have the additional advantage, compared to queuing, of providing the court system with information on whether there is in fact a strong demand for prompt trials. But the use of the price system[69] is not among the commonly proposed methods of reducing court delay, and the methods commonly proposed—such as procedural reform to simplify the trial of cases and thereby increase the effective litigation capacity of the courts, and the appointment of additional judges—have, in comparison to the use of price, some serious drawbacks.

The effect of the usual procedural reforms that are suggested (greater use of summary judgment, admissions, judicial notice, and the like) is to increase the productivity of litigation expenditures. The relationship to delay is obscure. An increase in the productivity of evidence will, as we have seen, induce litigants to purchase more of it. Thus, while it is repeatedly suggested that delay in Interstate Commerce Commission proceedings (for example) could be reduced if only the ICC would permit evidence without the cumbersome procedures, such as the best-evidence rule and the right of cross-examination, of common law proceedings, such simplification would induce the parties to increase the quantity of their litigation inputs—expert witnesses and the like—and this might result in even more protracted proceedings, albeit ones of higher quality.

The proposal to reduce delay by adding judges—usually considered the

---

1972). This figure is based on a sample of trial courts which consists mostly of courts in major population centers, where delay is traditionally greater than in less populous areas.

[66] Compare *id.* with Institute of Judicial Administration, Calendar Status Study—1963, at ii (July 15, 1963).

[67] Institute of Judicial Administration, Calendar Status Study: 1957, at i (Sept. 15, 1957); Hans Zeisel, Harry Kalven, Jr. & Bernard Buchholz, *supra* note 32, at 281; Judic. Conf., etc., *supra* note 52, at 136-39, 163-64, 334-36, 353-58.

[68] For example, in 1972 the average delay between answer and final disposition in personal-injury cases tried to juries was 58 months in Chicago, 50.2 months in Manhattan, and 47.9 months in Westchester County, New York. Institute of Judicial Administration, *supra* note 65, at vii.

[69] Proposed in William M. Landes, *supra* note 1.

sovereign remedy[70]—ignores several realistic possibilities that might undermine the effectiveness of the measure. The reduction in delay brought about by the addition of judges might be offset by the lower settlement rate in the personal-injury area, and perhaps in other areas, that can be foreseen if delay is reduced; the additional litigation would create a new source of delay. Moreover, with litigation a speedier method of dispute resolution, disputants who under existing conditions of delay substitute other methods of dispute resolution (such as arbitration) because they value prompt resolution would be attracted back to the courts, and again a new source of delay would be created. An analogy may be drawn to building a new freeway: by improving road transportation the freeway induces some people who previously used other modes of transportation to switch to driving, and this leads to new congestion.

The essential point is that minimization of delay is not an appropriate formulation of the goal of judicial reform. The goal, it has been argued in this article, is to minimize the sum of the error costs and of the direct costs of legal dispute resolution. The problem of delay must be placed within that larger framework of inquiry. Indeed, unless that is done, delay cannot even be defined in a meaningful fashion.

C. *A Note on Precedents*

An important problem in minimizing the sum of error costs and direct costs in legal dispute resolution is the relative weight to be assigned to the benefits from achieving a result in a particular case that is optimal in terms of the facts of that case and the benefits of decision according to precedent (judge-made rule). We will illustrate with reference to the question of assigning liability for accidents discussed in Part II.

From the standpoint of assuring the formal correctness of a liability rule, the best "rule" is simply to direct the court to assign liability in each case in such a way that efficiency will be maximized in light of the particular circumstances found in the case. However, this is not a rule at all, at least if the concept of rule is to have a useful meaning. The essence of a rule is that it abstracts one or a few facts from the totality of relevant circumstances and attaches controlling weight to them. Thus efficiency maximization is a standard but affixing liability to the driver of the following car in a rear-end collision is a rule. By abstracting and attaching controlling weight to the single fact that is probably most important to a determination of which driver could have minimized the sum of the accident and accident-avoidance costs, such a rule probably results in correct application of the standard (efficiency) in most cases. But since the rule suppresses facts that in some cases will dic-

[70] See, *e.g.*, Hans Zeisel, Harry Kalven, Jr. & Bernard Buchholz, *supra* note 32, at 207.

tate that liability be placed on the driver of the preceding car, it will not lead to correct results in all cases, and thus error costs will be generated.

Our hypothetical rear-end rule is imperfect because it violates the principle that the optimum assignment of accident-avoidance responsibilities to the participants in an accident is unique in every case. In one case, optimum avoidance may require that the first driver look in his rear-view mirror more frequently and the second increase the distance between his car and the car ahead from 15 feet to 30 feet; in another, that the first driver clean his rear window and the second install four-wheel disc brakes and an eight-inch rubber bumper; and so on. In contrast, the negligence standard, at least in some versions,[71] is perfectly efficient formally because it simply directs the court to identify in every case the particular allocation of liabilities that assures optimal safety precautions in that case. The main alternatives to negligence as the standard of accident liability are genuine rules: no liability, and strict liability without contributory negligence. Far from considering all of the relevant facts to be grist for the judicial decision-making mill, the no-liability rule considers none relevant, while the strict-liability rule singles out one fact as determinative: whether the defendant caused the plaintiff's injury. The costs that either rule would impose in the form of reduced efficiency would probably exceed the cost savings from substituting a rule for a standard. But it is equally probable that the efficiency gains from deciding every case by application of a general negligence standard would be outweighed by the costs. The costs of doing without liability rules must be very great, as a moment's reflection on the feasibility of managing a large organization without any rules, but only standards, should suggest. The costs of legal rules versus standards is an area that has not been studied extensively, so the brief analysis that follows should be considered highly tentative.

A legal system without rules would be a system without precedents. No decision in one case would control the decision of any future case unless the precise circumstances of a previous case happened to recur in a later case—an unlikely, and strictly perhaps an impossible, eventuality. Such a system would create a great deal of uncertainty concerning legal rights and obligations. The uncertainty would be a source of costs, including costly errors. Suppose that the purpose of a law is to forbid conduct X but the law is sufficiently unclear as to create a question whether Y, a socially desirable activity that resembles X, is not also forbidden and whether X_1, a subset of X, is not in fact permitted. The result will be to impose expected punishment costs on people engaged in Y which will reduce the level of Y below the social optimum, and to reduce the expected punishment costs on people engaged in X_1, which may increase the level of X_1 above the social optimum.

[71] See references in notes 5 and 13 *supra*.

Further, by extending the scope of liability from X to $X + Y$, the uncertain law will probably generate a greater number of claims, and every legal claim is costly to dispose of. The average costs of disposing of a legal claim may also be higher if the claim arises under an uncertain law. The absence of clear guidelines as to legal liability will make prediction of the outcome of a case difficult. This will cause the costs of settling cases to rise (minimum and maximum offers, based on the expected value of litigation to each party, will be harder to calculate), and therefore the settlement rate to fall since litigation is more likely the smaller the gap between trial costs and settlement costs.[72] Uncertainty might also operate directly to reduce the settlement rate. A fall in the settlement rate would be costly since settlements are in general much costlier than trials.[73]

A "rule-less" approach might also increase the probability of erroneous decisions—an ironical result since it is the promise of achieving efficiency in the individual case that is the principal attraction of the approach. A rule conveys information that may be lost when decision by rules is abandoned. Suppose that courts in a series of decisions in railroad spark cases, made over a period of years, evolve the rule that a railroad that causes engine-spark damage to crops will be deemed negligent if it fails to install the "latest and best" spark arrester. The rule summarizes the experience of these courts for a court or judge with less experience in spark cases. The rule may—in fact must—yield incorrect results in some cases but the absence of the rule might lead to incorrect results in more cases by forcing judges to confront the question of the optimum assignment of liability in each spark case anew. Of course records of prior decisions may be preserved even if the courts are assiduous in avoiding the formulation of rules. But without the compact and pointed statement communicated by the rule, judges may find it difficult to extract the prior judicial experience from records in a usable form. In like manner the issuance of rules to guide the decisions of subordinates in a business firm or other large organization may produce a higher percentage of correct decisions than if the subordinate is asked to apply a standard with such help as he can gain from consultation with coworkers who may have had experience with application of the standard in similar circumstances.

The proliferation of rules to impart certainty to the negligence standard

[72] This assumes that uncertainty will not cause an offsetting increase in litigation costs. I am inclined to think that it would not, cf. Part VI(B)(2) *supra*, although one's intuitive reaction is apt to be to the contrary. If uncertainty led to an increase in the costs of litigation that in turn enlarged the gap between litigation and settlement costs, the total costs of legal dispute resolution might actually fall: the increase in the settlement rate brought about by the enlargement of the gap might generate cost savings larger than the additional costs of those cases that went to trial.

[73] For some empirical evidence on this point see H. Laurence Ross, *supra* note 29, at 139-40.

of liability is a fact, although a rather neglected one, of the tort system. In the heyday of the common law of torts the creation of rules was carried on mainly by judges;[74] more recently it has been the work mainly of legislatures. Through the doctrine of "negligence *per se*," the multifarious legislative rules governing traffic safety and other areas have been translated into rules of liability in tort cases, with the result that the dispositive issue in many automobile accident cases, the most common form of tort action today, is whether either party violated a traffic law.[75] This is a far cry from applying a standard of efficiency maximizing on a case-by-case basis. Yet the courts' behavior may be quite consistent with a primary emphasis on achieving efficient allocation. There is a middle ground between the application of a standard of efficiency maximizing on a case-by-case basis and decision according to rules that make no attempt to achieve efficient allocation. The middle ground is occupied by rules, whether formulated by courts or legislatures, that seek to maximize the difference between the gains from efficient allocation and the loss from uncertainty and other undesirable properties of "rule-less" systems. Stated otherwise, the proper role of the negligence standard may not be as a criterion for the decision of individual cases, but as a criterion for the formulation of specific rules of liability that determine the outcome of the individual cases.

Consider once again the rule requiring railroads to install the "latest and best" spark-arresting equipment. As a formula for achieving optimum results in every case, the rule is clearly inadequate since there are bound to be cases where some other solution (such as a reduction in the train's speed, or a shift of the road bed, or a reduction in the number of trains) would be more efficient. It does not follow that the rule is inefficient. Suppose that there are only two liability rules alternative to it that are remotely feasible. One of these is no liability, and it would almost certainly be even less efficient than the "latest and best" rule. The other "rule" is to have no rule but to direct the court to decide in every case what combination of measures by railroad and neighboring farmers would optimize spark damage. This alternative may also be less efficient than the "latest and best" rule. The costs to the legal system imposed by the uncertainty of decision according to a standard may exceed the benefits of a formally more efficient criterion of liability.

[74] Many of these rules are described in Richard A. Posner, *supra* note 43; and see *id.* at 73 n. 38 for references to other compendia of these rules.

[75] The doctrine of negligence *per se* provides that failure to comply with a statute prescribing a standard of care is conclusive evidence of the violator's negligence. In other words, the court makes no independent judgment whether the violator's conduct was negligent. In a sample of trials in automobile accident cases, it was found that 39.2 per cent of the cases involved the question whether either party had violated a statute relating to road safety (computed from files of the Jury Project of the University of Chicago Law School).

APPENDIX

This appendix contains simple mathematical treatments of three topics discussed in the body of the article. They are (1) error costs in (civil) accident cases; (2) the effects of delay and of prejudgment interest on the likelihood that a case will be settled rather than tried; and (3) the implications of a Cournot-type model of the determinants of the amount spent by the parties in a lawsuit that is tried. The assistance of William M. Landes with respect to the first of these mathematical treatments is gratefully acknowledged.

1. *Error costs in accident cases.* To simplify analysis, we assume that accidents can be prevented only by the injurer's purchase of safety equipment; in particular, the victim cannot prevent the accident. Let S be the social costs of accidents and of accident avoidance; q the amount of safety equipment purchased; A(q) the number of accidents inflicted, as a function of q; J the cost of each accident; and C(q) the cost of the safety equipment, again as a function of q. Then the goal of a system of accident liability is to minimize

$$S = A(q) \cdot J + C(q). \tag{A1}$$

Since

$$\frac{dS}{dq} = A'(q) \cdot J + C'(q) \tag{A2}$$

the optimum q is derived by solving the equation

$$A'(q) \cdot J + C'(q) = 0. \tag{A3}$$

By assuming that the injurer's safety-equipment costs are constant, and by relabeling A'(q) M(q) (signifying the marginal product of safety equipment in reducing the accident costs inflicted by the injurer), we can rewrite equation (A3) as

$$M(q) \cdot J + c = 0. \tag{A4}$$

The second-order condition for equations (A3)-(A4) to yield a minimum is that

$$M'(q) \cdot J + C''(q) > 0. \tag{A5}$$

Since $C''(q) = 0$, this condition requires only, and reasonably, that $M'(q) > 0$ (*i.e.*, that the purchase of safety equipment reduce the number of accidents at a diminishing rate).

The injurer is interested in minimizing his private accident and accident-avoidance costs, rather than the social costs (S). Let Pr be his private accident and accident-avoidance costs, P the probability that he will in fact be held liable for those costs, and G(P) the amount of money he is forced to pay out in groundless claims. $G'(P) < 0$ and $G(P) = 0$ when $P = 1$. To derive the injurer's (private) optimum purchase of safety, we differentiate Pr with respect to q and set the result equal to zero:

$$Pr = P \cdot A(q) \cdot J + C(q) + G(P); \tag{A6}$$

$$\frac{dPr}{dq} = P \cdot M(q) \cdot J + c; \tag{A7}$$

$$P \cdot M(q) \cdot J + c = 0. \tag{A8}$$

G(P) drops out in the differentiation because it is assumed to be independent of q; this assumption is discussed in the body of the article.

To explore the effect on q of $P < 1$, we differentiate equation (A8) totally with respect to P, yielding

$$M(q) \cdot J + \frac{M(q)}{dq} PJ \frac{dq}{dP} = 0; \quad \text{hence} \tag{A9}$$

$$\frac{dq}{dP} = \frac{-M(q)}{PM'(q)}. \tag{A10}$$

Now we define two elasticities:

E_M = elasticity of the firm's marginal product function $= \dfrac{\dfrac{M'(q)}{M(q)}}{q}$;

E_P = elasticity of the quantity of safety purchased by the firm with respect to $P = \dfrac{dq}{dP} \cdot \dfrac{P}{q}$.

Hence equation (A10) can be rewritten as

$$E_P = -\frac{1}{E_M}. \tag{A11}$$

Note that $E_P > 0$ since $E_M < 0$ (M(q) being negative and M'(q), from the second-order condition, positive). Thus, reduction in P leads to a reduction in q. Also, equation (A11) demonstrates that this effect will be greater, the more elastic is the marginal social product of safety.

Let L be the social loss generated when $P < 1$ (see Figure 1 in the body of the article for a graphical representation of L). Since the area of a right-angle triangle is one-half the product of the two sides that form the right angle, the formula for L is approximately (unless M(q) is linear, in which case it is exact)

$$L = \tfrac{1}{2}[M(q_p) \cdot J - PM(q_p) \cdot J][q_s - q_p] \tag{A12}$$

where q_s and q_p are the social and private optimum quantities of safety equipment, respectively. Since $q_s - q_p$ is simply the change in q when P goes from one to below one (*i.e.*, dP), by multiplying both sides of equation (A10) by dP, substituting into equation (A12), and simplifying a bit, we can rewrite equation (A12) as

$$L = \tfrac{1}{2}[-PM(q_p) \cdot J(1-P)]\left[\frac{-M(q)}{M'(q)} \cdot \frac{dP}{P}\right]. \tag{A13}$$

Now let E_A = elasticity of number of accidents with respect to quantity of safety $= A'(q)\frac{q}{A}$. Substituting E_M and E_A into (A13), and simplifying, we can rewrite equation (A13) as

$$L = \tfrac{1}{2}\left[\frac{E_A}{E_M} AJ(1-P)\frac{dP}{P}\right], \tag{A14}$$

which as expected is positive, since $1 - P > 0$ for all $P < 1$, $dP = 1 - P$, and both elasticities are positive.

Equation (A14) implies that L is reduced by an increase in E_M and increased by an increase in E_A. L is also larger the smaller P is, since

$$\frac{\partial L}{\partial P} = \tfrac{1}{2}\frac{E_A}{E_M} AJ\frac{P^2 - 1}{P} \tag{A15}$$

and is negative since $P^2 < 1$ for all $P < 1$. L is also larger the larger J is, but, as discussed in the body of the article, this result is misleading. Assume $P = P(J)$ and $dP(J)/dJ > 0$ (*i.e.*, the parties spend more heavily on litigation in large cases and this reduces the probability of error). Then

$$L = \tfrac{1}{2}\frac{E_A}{E_M} A\frac{J(1-P(J))^2}{P(J)}; \quad \text{hence} \tag{A16}$$

$$\frac{\partial L}{\partial J} = \tfrac{1}{2}\frac{E_A}{E_M} A\left[\frac{P(J) - \frac{dP(J)}{dJ}J}{(P(J))^2}\right] - 2 + P(J) + \frac{dP(J)}{dJ}J, \tag{A17}$$

and is positive only if

$$P(J)\left[\frac{1-P(J)}{1+P(J)}\right] > \frac{dP(J)}{dJ}J. \tag{A18}$$

Finally, what are the costs of error if the purpose of accident law is assumed to be to compensate certain accident victims rather than to deter uneconomical accidents? Let A(P) be the number of accidents actually occurring for which injurers should be liable, expressed as a function of P, and Un the amount of undercompensation due to error.

$$Un = (1-P) \cdot A(P) \cdot J \tag{A19}$$

and is larger, the larger J is (since $\partial Un/\partial J$ is simply $(1 - P) \cdot A(P)$ and is positive) and the smaller is P:

$$\frac{\partial Un}{\partial P} = J\left[\frac{dA(P)}{dP} - A(P) - P\frac{dA(P)}{dP}\right]. \tag{A20}$$

Since $dA(P)/dP < 0$, the first and second bracketed terms are negative and the third positive; but the first is larger in absolute size than the third since $P < 1$, so $\partial Un/\partial P < 0$.

2. *Effects of delay and of prejudgment interest on the likelihood of litigation.* The condition for litigation is

$$P_pJ_p - C_p + S_p > (1 - P_d)J_d + C_d - S_d \tag{A21}$$

where P_p is the plaintiff's subjective probability of prevailing, P_d is the defendant's subjective probability of prevailing, J_p is the present value of a judgment to the plaintiff if he wins, J_d is the present cost of the judgment to the defendant if he loses, C_p and C_d are the plaintiff's and the defendant's litigation costs, respectively, and S_p and S_d are the plaintiff's and the defendant's settlement costs, respectively. Since delay and interest will be assumed not to affect litigation or settlement costs, we can rewrite inequality (A21) as

$$P_pJ_p > (1 - P_d)J_d + k, \quad \text{or} \tag{A22}$$

$$P_p\frac{J_p}{J_d} > 1 - P_d + \frac{k}{J_d}. \tag{A23}$$

Now let r_p be the plaintiff's discount rate and r_d the defendant's; J the value of the judgment to plaintiff if he wins the lawsuit, at the time that the final judgment is rendered; bJ the cost of the judgment to the defendant if he loses the lawsuit, also at the time that the final judgment is rendered; and t the interval in years between settlement and judgment. Then by the standard formula for discounting a future receipt or cost to present value $J_p = Je^{-r_pt}$ and $J_d = Je^{-r_dt}$. Substituting into inequality (A23), we obtain the following condition for litigation:

$$\frac{P_p}{b}e^{r_dt - r_pt} > 1 - P_d + \frac{k}{bJe^{-r_dt}}. \tag{A24}$$

Now let

$$y = \frac{P_p}{b}e^{r_dt - r_pt} - 1 + P_d - \frac{ke^{r_dt}}{bJ}. \tag{A25}$$

The larger y is, the more likely is litigation, and the smaller y, the more likely settlement.

$$\frac{\partial y}{\partial t} = (r_d - r_p)\frac{P_p}{b}e^{r_dt - r_pt} - 1 + P_d - \frac{r_dke^{r_dt}}{bJ}. \tag{A26}$$

Clearly, if $r_p \geqslant r_d$, $\partial y/\partial t < 0$ since $P_d < 1$. In these cases delay increases the likelihood of settlement. But if $r_d > r_p$, delay may increase the likelihood of litigation, depending on the specific values of the parameters in equation (A26).

To determine the effect of prejudgment interest, we first rewrite inequality (A22) as

$$P_pJe^{-r_pt} > (1 - P_d)bJe^{-r_dt} + k. \tag{A27}$$

Let

$$y = P_p J e^{-r_p t} - (1 - P_d) b J e^{-r_d t} - k. \tag{A28}$$

$$\frac{\partial y}{\partial J} = P_p e^{-r_p t} - (1 - P_d) b e^{-r_d t}. \tag{A29}$$

$$\frac{\partial y}{\partial J} > 0 \text{ iff } P_p e^{-r_p t} > (1 - P_d) b e^{-r_d t}; \quad \text{hence} \tag{A30}$$

$$\frac{\partial y}{\partial J} > 0 \text{ iff } P_p J e^{-r_p t} > (1 - P_d) b J e^{-r_d t}. \tag{A31}$$

i.e., if the present expected value of the judgment to the plaintiff is greater than the present expected cost of the judgment to the defendant; if not, the case would be settled in any event. Thus the addition of prejudgment interest (which has the effect of increasing J) can only increase the likelihood of litigation.

3. *Implications of a Cournot approach to the determinants of the parties' expenditures on litigation.* Each party may be assumed to select the level of expenditures on litigation at which the expected value of litigation to him is maximized. For the plaintiff,

$$E(V_p) = P_p J_p - C_p. \tag{A32}$$

We shall assume, rather arbitrarily, that

$$P_p = \frac{E_p C_p}{C_p + C_d}, \tag{A33}$$

where E_p is the plaintiff's estimate of the effectiveness of his expenditures relative to the defendant's in procuring a favorable outcome to the litigation. Since, when $C_p = C_d$, equation (A33) reduces to $P_p = \frac{1}{2}E_p$, E_p may be viewed as simply twice the plaintiff's subjective probability of prevailing when each party spends the same amount on the litigation. Similar formulas for $E(V_d)$ and P_d are assumed.

Let C_p^* be the plaintiff's optimum expenditure on the lawsuit and C_d^* the defendant's. By substituting equation (A33) into equation (A32) and differentiating with respect to C_p and C_d, we derive

$$C_p^* = \sqrt{E_p J_p C_d} - C_d \text{ and } C_d^* = \sqrt{E_d J_d C_p} - C_p. \tag{A34}$$

Since

$$\frac{\partial C_p^*}{\partial C_d} = \tfrac{1}{2}\sqrt{\frac{E_p J_p}{C_d}} - 1, \tag{A35}$$

an increase in C_d will induce the plaintiff to increase his own expenditures if $\sqrt{C_d} < 2\sqrt{E_p J_p}$ and to reduce them if $\sqrt{C_d} > 2\sqrt{E_p J_p}$.

If we adopt the Cournot assumption that the parties do not consider the impact of their expenditures on the other party's, we can treat equations (A34) as simultaneous equations and solve for C_p^* and C_d^*:

$$C_p^* = \frac{E_p^2 J_p^2 E_d J_d}{(E_p J_p + E_d J_d)^2};$$
$$C_d^* = \frac{E_d^2 J_d^2 E_p J_p}{(E_p J_p + E_d J_d)^2}. \tag{A36}$$

The sum of the parties' expenditures is

$$C_p^* + C_d^* = \frac{E_p J_p E_d J_d}{E_p J_p + E_d J_d} \tag{A37}$$

and the ratio is

$$\frac{C_p^*}{C_d^*} = \frac{E_p J_p}{E_d J_d}. \tag{A38}$$

Let P_p^* be the objective probability of the plaintiff's prevailing. We assume that

$$P_p^* = \tfrac{1}{2}(P_p + 1 - P_d). \tag{A39}$$

Substituting from equation (A36) we can rewrite this as

$$P_p^* = \frac{E_p(E_p + 1) + E_d(1 - E_d)}{2(E_p + E_d)}. \tag{A40}$$

Thus if E_p is only .4 (meaning that $P_p = .2$ if $C_p = C_d$) and E_d is 1.6 ($P_d = .8$ if $C_p = C_d$), it turns out that P_p^* is zero, not 20 per cent.

By substituting the P's and C's derived from our Cournot solution into inequality (A21), we derive the following condition for litigation to occur (for simplicity we assume $J_p = J_d = J$ and that $S_p = S_d = 0$):

$$E_p^2 + E_d^2 > E_p + E_d + E_p E_d. \tag{A41}$$

It is evident that, for most plausible values of E_p and E_d, the condition is not satisfied—the case will be settled—even when both parties are highly optimistic. For example, if $E_p = E_d$, the condition becomes $2E_p > 2E_p^2 + E_p^2$ or $E_p > 2$, and is never satisfied since an E_p greater than two would imply that the plaintiff's probability of winning exceeded 100 per cent when both parties spent the same amount on the litigation. Inequality (A41) is satisfied only if E_p is very large and E_d very small (*e.g.*, if $E_p = 1.6$ and $E_d = .4$), or vice versa; and the results in these cases are spurious. Since a party will never spend to increase his probability of winning above 100 per cent, it is necessary to impose two additional conditions (derived from equation (A33)):

$$C_p < \frac{C_d}{E_p - 1};\ C_d < \frac{C_p}{E_d - 1}. \tag{A41}$$

A large disparity between E_p and E_d brings these conditions into play. For example, if $E_p = 1.6$ and $E_d = .4$, then equations (A36) yield $C_p = 25.6$ (we assume $J =$

100) and $C_d = 6.4$, which makes P_p in equation (A33) 1.28. After adjusting C_p and C_d in accordance with the conditions in (A41) and substituting the adjusted values into inequality (A21), we find that this case, although involving a large difference between E_p and E_d, is settled too. C_p* becomes 15.625 and C_d* 9.375. Substituting these values into equation (A33) and a similar equation for P_d, and thence into inequality (A21) (the condition for litigation), that condition becomes $84.375 > 94.375$ and is not satisfied.