

ECONOMIC ANALYSIS OF LAW

LOUIS KAPLOW and STEVEN SHAVELL

Harvard Law School and National Bureau of Economic Research

Contents

Abstract	1665
Keywords	1665
1. Introduction	1666
2. Liability for accidents	1667
2.1. Incentives	1667
2.1.1. Unilateral accidents and the level of care	1668
2.1.2. Bilateral accidents and levels of care	1669
2.1.3. Unilateral accidents, level of care, and level of activity	1670
2.1.4. Bilateral accidents, levels of care, and levels of activity	1671
2.1.5. Empirical evidence on the effect of liability on safety	1671
2.2. Risk-bearing and insurance	1671
2.3. Administrative costs	1673
2.3.1. Administrative costs of the liability system	1673
2.3.2. Administrative costs and the social desirability of the liability system	1674
2.4. Magnitude of liability: damages	1674
2.4.1. Basic theory	1674
2.4.2. Nonpecuniary elements of loss	1675
2.4.3. Punitive damages	1675
2.4.4. Accuracy of damages	1676
2.4.5. Components of loss that are difficult to estimate	1677
2.5. Causation	1677
2.5.1. Basic requirement of causation	1677
2.5.2. Uncertainty over causation	1678
2.5.3. Proximate causation	1678
2.6. Judgment-proof problem	1679
2.7. Product liability	1680
2.8. Liability versus other means of controlling accidents	1682
2.9. Intentional torts	1682
3. Property law	1682
3.1. Justifications for property rights	1683
3.2. Emergence of property rights	1684

3.3. Division and form of property rights	1685
3.3.1. Division of property rights	1685
3.3.2. Consolidated form of property rights and the theory of the firm	1686
3.4. Public property	1688
3.4.1. Justifications for public property	1688
3.4.2. Acquisition of public property: purchase versus compensated takings	1688
3.4.3. Compensation for takings	1689
3.5. Acquisition and transfer of property	1689
3.5.1. Acquisition of unowned property	1689
3.5.2. Loss and recovery of property	1690
3.5.3. Acquisition of stolen property and problems of establishing valid title	1690
3.5.4. Involuntary transfer of property: adverse possession	1691
3.5.5. Constraints on sale of property	1692
3.5.6. Gifts	1692
3.6. Conflicts in the use of property: externalities	1693
3.6.1. Socially optimal resolution of externalities	1693
3.6.2. Resolution of externalities through state intervention	1693
3.6.3. Resolution of externalities through bargaining by affected parties	1696
3.7. Property rights in information	1698
3.7.1. Inventions, compositions, and other intellectual works of repeat value	1698
3.7.2. Other types of information	1700
3.7.3. Information valuable as labels	1701
4. Contracts	1702
4.1. Basic theory	1702
4.1.1. Definitions and framework of analysis	1702
4.1.2. Contract formation	1703
4.1.3. Why contracts and their enforcement are valuable to parties	1705
4.1.4. Incomplete nature of contracts and their less-than-rigorous enforcement	1706
4.1.5. Interpretation of contracts	1707
4.1.6. Damage measures for breach of contract	1708
4.1.7. Specific performance as a remedy for breach	1710
4.1.8. Renegotiation of contracts	1711
4.1.9. Legal overriding of contracts	1712
4.2. Production contracts	1713
4.2.1. Value of performance and production cost	1713
4.2.2. Reliance investment during the contract period	1715
4.2.3. Further considerations	1719
4.3. Other types of contract	1720
4.3.1. Contracts for transfer of possession	1720
4.3.2. Donative contracts	1721
4.3.3. Additional types of contract	1722

5. Litigation	1722
5.1. Suit	1722
5.1.1. Private incentive to sue	1722
5.1.2. Socially optimal suit versus the private incentive to sue	1723
5.1.3. Implications of the social and private divergence	1725
5.2. Settlement versus trial	1726
5.2.1. Exogenous beliefs model	1726
5.2.2. Asymmetric information model	1727
5.2.3. Socially optimal versus privately determined settlement	1729
5.3. Litigation expenditures	1730
5.3.1. Private incentives to spend on litigation	1730
5.3.2. Social versus private incentives to make litigation expenditures	1730
5.4. Extensions of the basic theory	1731
5.4.1. Nuisance suits	1731
5.4.2. Shifting of legal fees	1732
5.4.3. Additional elements of trial outcomes	1733
5.4.4. Statistical inference from cases that go to trial	1734
5.4.5. Lawyers as agents of litigants	1734
5.4.6. Insurers as agents of litigants	1735
5.4.7. Voluntary sharing of information	1735
5.4.8. Required disclosure of information – legal discovery	1736
5.4.9. Criminal adjudication	1738
5.4.10. Additional aspects of legal procedure	1738
5.5. Legal advice	1739
5.5.1. Ex ante legal advice: when acts are contemplated	1739
5.5.2. Ex post legal advice: at the stage of litigation	1740
5.5.3. Other aspects of legal advice	1741
5.6. Appeals	1742
5.7. Alternative dispute resolution	1743
5.7.1. Ex ante ADR agreements	1743
5.7.2. Ex post ADR agreements	1744
5.8. Formulation of legal rules	1744
5.9. Relevance to general incentive schemes	1745
6. Law enforcement	1745
6.1. Rationale for public enforcement	1746
6.2. Basic theory of enforcement	1747
6.2.1. Optimal enforcement given the probability of detection	1748
6.2.2. Optimal enforcement including the probability of detection	1750
6.3. Extensions of the basic theory	1752
6.3.1. Accidental harms	1752
6.3.2. Level of activity	1753
6.3.3. Enforcement error	1753
6.3.4. General enforcement	1754

6.3.5. Marginal deterrence	1755
6.3.6. Repeat offenders	1756
6.3.7. Self-reporting	1757
6.3.8. Plea bargaining	1757
6.3.9. Corruption of law enforcement agents	1758
6.3.10. Principal-agent relationship	1758
6.3.11. Incapacitation	1759
6.3.12. Empirical evidence on law enforcement	1760
6.4. Criminal law	1760
7. Criticism of economic analysis of law	1761
7.1. Positive analysis	1761
7.2. Normative analysis	1762
7.2.1. Distribution of income	1762
7.2.2. Victim compensation	1763
7.2.3. Concerns for fairness	1763
7.3. Purported efficiency of judge-made law	1764
8. Conclusion	1765
Acknowledgments	1765
References	1765

Abstract

This is a survey of economic analysis of law, that is, of the emerging field under which the standard tools of microeconomics are employed to identify the effects of legal rules and their social desirability. Five basic subject areas are covered. The first is legal liability for harm. Here we discuss liability rules as incentives to reduce risk, issues of risk-bearing and insurance, and the costs of the liability system. Second, we consider property law, where we address the nature and justification of property rights, public property, the acquisition and transfer of property, externalities surrounding the use of property, and intellectual property. Third, we examine contract law, including the formation of contracts, their interpretation, and remedies for their breach. We focus on production contracts but also discuss other types, including donative contracts. Fourth, we treat the subject of civil litigation, that is, the bringing of lawsuits, and their settlement or disposition at trial. We also mention the appeals process, alternative dispute resolution, the provision of legal advice, and several additional topics relating to litigation. Fifth, we consider public enforcement of law, focusing on the level of law enforcement effort, the magnitude of sanctions, and other issues relevant to criminal law. Finally, we discuss criticisms that are commonly made by legal academics of economic analysis of law and offer concluding remarks.

Keywords

law and economics, liability, accident(s), tort(s), insurance, damage(s), causation, product liability, intentional tort(s), property, property right(s), externalities, regulation, firm(s), public property, takings, intellectual property, patent, copyright, trade secret, trademark, contract(s), disclosure, legal procedure, litigation, suit, settlement, trial, court(s), lawyer(s), judge(s), appeal(s), alternative dispute resolution, law enforcement, crime, plea bargaining, fine(s), imprisonment

JEL classification: K00, K10, K11, K12, K13, K14, K40, K41, K42, H23, L51

1. Introduction

Economic analysis of law seeks to answer two basic questions about legal rules. Namely, what are the effects of legal rules on the behavior of relevant actors? And are these effects of legal rules socially desirable? In answering these positive and normative questions, the approach employed in economic analysis of law is that used in economic analysis generally: the behavior of individuals and firms is described assuming that they are forward looking and rational, and the framework of welfare economics is adopted to assess the social desirability of outcomes.

The field of economic analysis of law may be said to have begun with Bentham (1789, 1827, 1830), who systematically examined how actors would behave in the face of legal incentives and who evaluated outcomes with respect to a clearly stated measure of social welfare (utilitarianism). Bentham's writings contain significant and extended analysis of criminal law and law enforcement, some analysis of property law, and a substantial treatment of the legal process. His work was left essentially undeveloped until the 1960s and early 1970s, when interest in economic analysis of law was stimulated by four important contributions: Coase's (1960) article on externalities and legal liability, Becker's (1968) article on crime and law enforcement, Calabresi's articles and culminating book (1970) on accident law, and R.A. Posner's (1972) general textbook on economic analysis of law and his establishment of the *Journal of Legal Studies*. As this survey will indicate, research in economic analysis of law has been active since the 1970s and is accelerating¹. The field, however, is far from mature; one indication is the lack of empirical work on most topics.

Our focus here will be analytical, and we will cover five basic legal subjects². The first three are the central areas of civil law. We begin with liability for accidents, which can be understood as addressing the problem of probabilistic externalities. Second, we discuss property law, which concerns the nature and justification of property rights, how they are acquired and transferred, how conflicts in the use of property are resolved, and related topics. Third, we examine contract law, including the formation of contracts, their interpretation, and remedies for their breach. The following section

¹ The field of law and economics is presented in a number of books, Cooter and Ulen (1997), Miceli (1997), Polinsky (1989), R.A. Posner (1998) and Shavell (forthcoming), and in two reference works, The New Palgrave Dictionary of Economics and the Law [Newman (1998)] and the Encyclopedia of Law and Economics [Bouckaert and De Geest (2000)]. Journals specializing in law and economics include the *Journal of Legal Studies*, the *Journal of Law and Economics*, the *Journal of Law, Economics, & Organization*, the *American Law and Economics Review*, and the *International Review of Law and Economics*. Also, professional organizations, including the American Law and Economics Association and the European Association of Law and Economics, are now well established.

² The sections on these subjects can be read largely independently of each other. Not treated in our survey are various, more particular areas of law than the five we have mentioned; omitted areas include antitrust law, corporate and securities law, bankruptcy and commercial law, banking law, international trade law, and tax law. Also excluded from this survey are problems addressed by the literatures on public choice and positive political theory.

concerns civil litigation, that is, the bringing of lawsuits by private actors to enforce their rights in the areas of law that we have just discussed. Next, we consider public enforcement of law, focusing on the level of law enforcement effort, the magnitude of sanctions, and other issues relevant to criminal law. Finally, we discuss criticisms that are commonly made by legal academics of economic analysis of law and offer concluding remarks.

2. Liability for accidents

Legal liability for accidents, which is governed by tort law, is a means by which society can reduce the risk of harm by threatening potential injurers with having to pay for the harms they cause. Liability is also frequently viewed as a device for compensating victims of harm, but we will emphasize that insurance can provide compensation more cheaply than the liability system. Thus, we will view the primary social function of the liability system as the provision of incentives to prevent harm.

There are two basic rules of liability. Under *strict liability*, an injurer must always pay for harm due to an accident that he causes. Under the *negligence rule*, an injurer must pay for harm caused only when he is found negligent, that is, only when his level of care was less than a standard of care chosen by the courts, often referred to as due care. (There are various versions of these rules that depend on whether victims' care was insufficient, as we will discuss below.) In fact, the negligence rule is the dominant form of liability; strict liability is reserved mainly for certain especially dangerous activities (such as the use of explosives).

Our discussion of liability begins by examining how liability rules create incentives to reduce risk. The allocation of risk and insurance will then be considered, and following that, the factor of administrative costs. Then we take up a number of important topics bearing on liability: the magnitude of liability (damages), causation, and the judgment-proof problem (assets insufficient to pay for harm). Finally, we consider the subjects of product liability and intentional torts³.

2.1. Incentives

In order to focus on liability and incentives to reduce risk, we assume in this section that parties are risk neutral. Further, we suppose that there are two classes of parties, injurers and victims, and that they are strangers to one another, or at least are not in a contractual relationship. For example, injurers might be drivers and victims pedestrians, or injurers might be polluting firms and victims affected residents.

To begin with, we assume that accidents are unilateral in nature: only injurers can influence risks. Then we consider bilateral accidents, in which victims as well as

³ A comprehensive economic treatment of accident law is contained in Shavell (1987a), which this section largely follows. See also Landes and Posner (1987a) and Calabresi (1970).

injurers affect risks. We also examine two types of action that parties can take that alter risk: first we consider their level of care (such as driving speed) and then their level of activity (number of miles driven).

2.1.1. *Unilateral accidents and the level of care*

Here we suppose that injurers alone can reduce risk by choosing a level of care. Let x be expenditures on care (or the money value of effort devoted to it) and $p(x)$ be the probability of an accident that causes harm h , where p is declining in x . Assume that the social objective is to minimize total expected costs, $x + p(x)h$, and let x^* denote the optimal x .

Under strict liability, injurers pay damages equal to h whenever an accident occurs, and they naturally bear the cost of care x . Thus, they minimize $x + p(x)h$; accordingly, they choose x^* .

Under the negligence rule, suppose that the due care level \hat{x} is set equal to x^* , meaning that an injurer who causes harm will have to pay h if $x < x^*$ but will not have to pay anything if $x \geq x^*$. Then it can be shown that the injurer will choose x^* : clearly, the injurer will not choose x greater than x^* , for that will cost him more and he will escape liability by choosing merely x^* ; and he will not choose $x < x^*$, for then he will be liable (in which case the analysis of strict liability shows that he would not choose $x < x^*$).

Thus, under both forms of liability, injurers are led to take optimal care. But note that under the negligence rule, courts need to be able to calculate optimal care x^* and to be able to observe actual care x , in addition to observing harm. In contrast, under strict liability courts do not need to do the former two; they only need to observe harm⁴.

It should also be noted that, under the negligence rule with due care \hat{x} equal to x^* , negligence would never actually be found, because injurers are induced to choose x^* and thus would be exonerated if they were sued after causing an accident. Findings of negligence may occur, however, under a variety of modifications of our assumptions. Courts might make errors in observing injurers' actual level of care so that an injurer whose true x is at least x^* might mistakenly be found negligent because his observed level of care is below x^* . Similarly, courts might err in calculating x^* and thus might set due care \hat{x} above x^* . If so, an injurer who chooses x^* would be found negligent (even though care is accurately observed) because \hat{x} exceeds x^* . As emphasized by Craswell and Calfee (1986), the chance of errors in the negligence determination leads injurers to choose incorrect levels of care; one possibility is that they would take excessive care in order to reduce the risk of being found negligent by mistake⁵. There

⁴ Compare the discussion of corrective taxes versus regulation in Section 3.6.2.

⁵ This might explain the phenomenon of "defensive medicine", on which see Danzon (1985) and, for empirical evidence, Kessler and McClellan (1996). Whether there is a tendency toward excessive

exist other explanations for findings of negligence as well, including that individuals may not know x^* and thus take too little care, the judgment-proof problem, which may lead individuals to choose to be negligent (see Section 2.6), and the inability of individuals to control their behavior perfectly at every moment or of firms to control their employees.

2.1.2. Bilateral accidents and levels of care

We now assume that victims also choose a level of care y , that the probability of an accident is $p(x,y)$ and is declining in both variables, that the social goal is to minimize $x + y + p(x,y)h$, and that the optimal levels of care x^* and y^* are positive⁶.

Under strict liability, injurers' incentives are optimal conditional on victims' level of care, but victims have no incentive to take care because they are fully compensated for their losses. However, the usual strict liability rule that applies in bilateral situations is strict liability with a defense of *contributory negligence*, meaning that an injurer is liable for harm only if the victim's level of care was not negligent, that is, his level of care was at least his due care level \hat{y} . If victims' due care level is set by the courts to equal y^* , then it is a unique equilibrium for both injurers and victims to act optimally: victims can be shown to choose y^* in order to avoid having to bear their losses, and injurers will choose x^* since they will in fact be liable, as victims will not be negligent⁷.

Under the negligence rule, optimal behavior, x^* and y^* , is also the unique equilibrium. Injurers can be shown to choose x^* to avoid being liable, and since victims will therefore bear their losses, they will choose y^* ⁸. Two other variants of the negligence rule are negligence with the defense of contributory negligence (under which a negligent injurer is liable only if the victim is not negligent) and the

care depends upon the degree of legal error and on whether injurers who are found negligent are held responsible for all harm caused or only the incremental harm attributable to their negligence. On the latter, see Grady (1983) and Kahan (1989).

⁶ In some early, less formal literature on accidents, for example, Calabresi (1970), reference is made to the notion of the "least-cost avoider", the party – injurer or victim – who can avoid an accident at the lower cost. The idea of a least-cost avoider relies on the assumption that each party can undertake a discrete amount of care that is independently sufficient to prevent an accident.

⁷ That this equilibrium is unique follows from three observations: (1) Victims never have an incentive to take care y exceeding y^* (for once they take due care they will be compensated for their losses). (2) Victims will not choose y less than y^* , for if they do so, they will bear their own losses, injurers will take no care, and victims thus will minimize $y + p(0,y)h$. But $y + p(0,y)h = 0 + y + p(0,y)h > x^* + y^* + p(x^*, y^*)h > y^*$, implying that victims must be better off choosing due care y^* than any $y < y^*$. (3) Because in equilibrium victims thus take due care, injurers choose x to minimize $x + p(x, y^*)h$, which is minimized at x^* .

⁸ Uniqueness is demonstrated by the following: (1) Injurers will not take care exceeding x^* . (2) If injurers choose x less than x^* , victims will take no care, so injurers will minimize $x + p(x, 0)h$, which exceeds $x^* + y^* + p(x^*, y^*)h$, which exceeds x^* . Thus, injurers are better off taking care of x^* . (3) Since injurers must choose x^* in equilibrium, victims will choose y^* .

comparative negligence rule (under which a negligent injurer is only partially liable if the victim is also negligent). These rules are also readily shown to induce optimal behavior in equilibrium.

Thus, all of the negligence rules, and strict liability with the defense of contributory negligence, support optimal levels of care x^* and y^* in equilibrium, assuming that due care levels are chosen optimally. Courts need to be able to calculate optimal care levels for at least one party under any of the rules, and in general this requires knowledge of the function $p(x, y)$. The main conclusions of this and the last section were first proved by Brown (1973)⁹.

2.1.3. Unilateral accidents, level of care, and level of activity

Now let us reconsider unilateral accidents, allowing for injurers to choose their level of activity z , which is interpreted as the (continuously variable) number of times they engage in their activity (or if injurers are firms, the scale of their output). Let $b(z)$ be the benefit (or profit) from the activity, and assume the social object is to maximize $b(z) - z(x + p(x)h)$; here $x + p(x)h$ is assumed to be the cost of care and expected harm each time an injurer engages in his activity. Let x^* and z^* be optimal values. Note that x^* minimizes $x + p(x)h$, so x^* is as described above in Section 2.1.1, and that z^* is determined by $b'(z) = x^* + p(x^*)h$, which is to say, the marginal benefit from the activity equals the marginal social cost, comprising the sum of the cost of optimal care and expected accident losses (given optimal care).

Under strict liability, an injurer will choose both the level of care and the level of activity optimally, as his objective will be the same as the social objective, to maximize $b(z) - z(x + p(x)h)$, because damage payments equal h whenever harm occurs.

Under the negligence rule, an injurer will choose optimal care x^* as before, but his level of activity z will be socially excessive. In particular, because an injurer will escape liability by taking care of x^* , he will choose z to maximize $b(z) - zx^*$, so that z will satisfy $b'(z) = x^*$. The injurer's cost of raising his level of activity is only his cost of care x^* , which is less than the social cost, as it also includes $p(x^*)h$. The excessive level of activity under the negligence rule will be more important the larger is expected harm $p(x^*)h$ from the activity.

The failure of the negligence rule to control the level of activity arises because negligence is defined here (and for the most part in reality) in terms of care alone. A justification for this restriction in the definition of appropriate behavior is the difficulty courts would face in determining the optimal z^* and the actual z . Moreover, the problem with the activity level under the negligence rule is applicable to any aspect of behavior that would be difficult to regulate directly (including, for example, research and development activity). If, instead, courts were able to incorporate all aspects of

⁹ Diamond (1974) proved closely related results shortly afterward. See also Green (1976), Emons (1990), and Emons and Sobel (1991), who focus on the case of heterogeneous injurers and victims.

behavior into the definition of negligence, the negligence rule would result in optimal behavior in all respects. (Note that the variable x in the original problem could be interpreted as a vector, with each element corresponding to a dimension of behavior.)

2.1.4. Bilateral accidents, levels of care, and levels of activity

If we consider levels of care and of activity for both injurers and victims, then none of the liability rules that we have considered leads to full optimality (assuming that activity levels are unobservable). As just explained, the negligence rule induces injurers to engage excessively in their activity. Similarly, strict liability with a defense of contributory negligence leads victims to engage excessively in their activity (the number of times they expose themselves to risk), as they do not bear their losses given that they take due care. The reason that full optimality cannot be achieved is in essence that injurers must bear accident losses to induce them to choose the right level of their activity, but this means that victims will not choose the optimal level of their activity, and conversely¹⁰. The distinction between levels of care and levels of activity was first emphasized in Shavell (1980c), where the results of this and the last section were shown.

2.1.5. Empirical evidence on the effect of liability on safety

Only a modest amount of empirical work has been undertaken on the effect of liability on accident risks. See Dewees, Duff and Trebilcock (1996) for a general survey of the literature that exists, and, among others, Devlin (1990), Landes (1982), and Sloan et al. (1994) on liability and auto accidents, Danzon (1985) and Kessler and McClellan (1996) on liability and adverse medical outcomes, and Higgins (1978), Priest (1988), and Viscusi (1991) on liability and product safety.

2.2. Risk-bearing and insurance

We consider next the implications of risk aversion and the role of insurance in the liability system, on which see Shavell (1982a). Several general points may be made.

First, the socially optimal resolution of the accident problem obviously now involves not only the reduction of losses from accidents, but also the protection of risk-averse parties against risk. Note that risk-bearing is relevant for two reasons: not only because potential victims may face the risk of accident losses, but also because potential injurers may face the risk of liability. The former risk can be mitigated through

¹⁰ However, there exist ways to induce fully optimal behavior using tools other than conventional liability rules. For example, if injurers have to pay the state for harm caused and victims bear their own losses, both victims and injurers will choose levels of care and of activity optimally. On the possibility of such decoupling of what injurers pay and what victims receive, see note 107.

insurance that covers losses suffered in accidents, and the latter through liability insurance.

Second, because risk-averse individuals will tend to purchase insurance, the incentives associated with liability do not function in the direct way discussed in the last section, but instead are mediated by the terms of insurance policies. To illustrate, consider strict liability in the unilateral accident model with care alone allowed to vary, and assume that insurance is sold at actuarially fair rates. If injurers are risk averse and liability insurers can observe their levels of care, injurers will purchase full liability insurance coverage and their premiums will depend on their level of care; their premiums will equal $p(x)h$. Thus, injurers will want to minimize their costs of care plus premiums, or $x + p(x)h$, so they will choose the optimal level of care x^* . In this instance, liability insurance eliminates risk for injurers, and the situation reduces to the previously analyzed risk-neutral case.

If, however, liability insurers cannot observe levels of care, ownership of full coverage could create severe moral hazard, so would not be purchased. Instead, as we know from the theory of insurance, the typical amount of coverage purchased will be partial, for that leaves injurers with an incentive to reduce risk. In this case, therefore, the liability rule results in some direct incentive to take care because injurers are left bearing some risk after their purchase of liability insurance. But injurers' level of care will still tend to be less than first-best.

This last situation, in which liability insurance dilutes incentives, leads to our third point, concerning the question whether the sale of liability insurance is socially desirable. (We note that because of fears about incentives, the sale of liability insurance was delayed for decades in many countries and that it was not allowed in the former Soviet Union; further, in this country liability insurance is sometimes forbidden against certain types of liability, such as against punitive damages.) The answer to the question is that sale of liability insurance is socially desirable, at least in basic models of accidents and some variations of them. In the case just considered, the reason is evident. Injurers are made better off by the presence of liability insurance, as they choose to purchase it. Victims are indifferent to its purchase by injurers because victims are fully compensated under strict liability for any losses they sustain. In particular, it does not matter to victims that the likelihood of accident may rise due to the sale of liability insurance. This argument must be modified in other cases, such as when the damages injurers pay are less than harm because injurers are judgment proof. In that circumstance, the sale of liability insurance may not be socially desirable. See Section 2.6.

Fourth, consider how the comparison between strict liability and the negligence rule is affected by considerations of risk-bearing. It is true that the immediate effect of strict liability is to shift the risk of loss from victims to injurers, whereas the immediate effect of the negligence rule is to leave the risk on victims (injurers will tend to act non-negligently). However, the presence of insurance means that victims and injurers can substantially shield themselves from risk. Of course, as was just discussed, insurance coverage may be incomplete due to moral hazard; this makes risk-bearing of some

relevance to the comparison of liability rules, but which rule becomes more favorable is not obvious.

Finally, as we stated at the outset of Section 2, the presence of insurance implies that the liability system cannot be justified primarily as a means of compensating risk-averse victims against loss. Rather, the justification for the liability system must lie in significant part in the incentives that it creates to reduce risk. To amplify, although both the liability system and the insurance system can compensate victims, the liability system is much more expensive than the insurance system (see the next section)¹¹. Accordingly, were there no social need to create incentives to reduce risk, it would be best to dispense with the liability system and to rely on insurance to accomplish compensation¹².

2.3. Administrative costs

2.3.1. Administrative costs of the liability system

The administrative costs of the liability system are the legal and other costs (notably the time of litigants) involved in bringing suit and resolving it through settlement or trial. These costs are substantial; a number of estimates suggest that, on average, administrative costs of a dollar or more are incurred for every dollar that a victim receives through the liability system. In contrast, the administrative cost of receiving a dollar through the insurance system is often below fifteen cents¹³.

The factor of administrative costs affects the comparison between the forms of liability. On one hand, we would expect the volume of cases – and thus administrative costs – to be higher under strict liability than under the negligence rule. This is because, under strict liability, a victim can collect whether or not the injurer was at fault, whereas under the negligence rule fault must be established, so that in many cases of accident there will be no suit or, if there is a suit, it will be dropped after little has been spent¹⁴. On the other hand, given that there is a case, we would anticipate administrative costs to be higher under the negligence rule than under strict liability, because under the

¹¹ Also, victim compensation through liability generally implies that possibly risk-averse injurers bear risk. In some contexts, such as auto accidents, one supposes that injurers are not substantially less risk averse than victims.

¹² Some jurisdictions have implemented “no-fault” regimes (essentially, insurance that covers losses suffered in accidents) for automobile accidents. See Dewees, Duff and Trebilcock (1996). Also, there are intermediate schemes, like workers’ compensation, that provide compensation and charge experience-rated premiums to injurers to instill incentives to reduce risk. See Moore and Viscusi (1990).

¹³ See Danzon (1985, p. 187), Kakalik et al. (1983), and Shavell (1987a, p. 263).

¹⁴ Farber and White (1991) provide evidence that many medical malpractice cases are dropped after discovery, when plaintiffs learn that the defendant probably was not negligent. Relatedly, Ordover (1978) analyzes a model in which victims are uncertain about injurers’ negligence; the result is that some victims of negligence do not sue and others who are not victims of negligence do sue.

negligence rule due care will be at issue. In consequence, it is in theory ambiguous whether strict liability or the negligence rule will be administratively cheaper.

2.3.2. *Administrative costs and the social desirability of the liability system*

The existence of administrative costs and their significant magnitude raises rather sharply the question whether it is worthwhile for society to bear them to gain the benefits of the liability system – the incentives to reduce risk. Unfortunately, it is quite possible for suits to be attractive for private parties to bring even if the social benefits of the liability system are small and make it socially undesirable. For example, victims will have strong incentives to bring suit under a strict liability system however low the risk reduction effect of suit may be. This point about the private versus the social incentive to make use of the legal system will be emphasized in Section 5.1.2.

2.4. *Magnitude of liability: damages*

The magnitude of the payment a liable party must make is known as damages, because it is normally set equal to the harm the victim has sustained. In this section, we discuss various issues relating to damages.

2.4.1. *Basic theory*

As a general matter, damages should equal harm under strict liability for incentives to be optimal in the unilateral model of accidents. Clearly, for injurers to be led to choose optimal levels of care, their expected liability must equal expected harm $p(x)h$, meaning that damages d should equal h . Likewise, for their levels of activity to be optimal, the same must be true¹⁵.

We should add that this point essentially carries over to the situation, not yet considered, where the magnitude of harm is stochastic. In this case, if damages d equal harm, then expected liability will equal expected harm, so incentives will be correct. However, if damages d do not equal actual harm but instead are set equal to $E_c(h)$, expected harm conditional on harm occurring, incentives will also be correct. (For elaboration, see Section 2.4.4.)

Under the negligence rule, analysis of the optimal magnitude of damages is somewhat different. Recall that if damages equal harm h , injurers will be induced to take care of x^* (assuming that due care $\hat{x}=x^*$). It is also the case that damages higher than h would induce injurers to take care of x^* : this will increase the incentive to be non-negligent, to choose x^* , but it will not lead injurers to take excessive care

¹⁵ In the bilateral model, damages equal to harm would also be optimal under a rule of strict liability with a defense of contributory negligence if victims' activity level is not variable. If their activity level is variable, then optimal damages may well be less than harm, for this will induce victims to moderate their level of activity.

because they can escape liability merely by taking care of x^* . Moreover, it can be shown that damages somewhat below h will also induce due care because, by taking due care rather than slightly less care under the negligence rule, injurers do not just reduce liability slightly but avoid liability altogether¹⁶. Thus, optimal damages are not unique but range from a level somewhat below h to any greater level. When, however, one introduces the possibility of uncertainty in the negligence determination (see Section 2.1.1), the situation becomes more complicated. For example, we noted that error in the negligence determination might lead injurers to take excessive care to reduce the risk of being found negligent by mistake. If so, a level of damages exceeding h would only exacerbate this problem, and it might be beneficial for d to be lower than h ¹⁷.

To sum up, we can say that in simple cases damages should equal harm under strict liability and under the negligence rule, although there are complications, such as that concerning uncertainty in the negligence determination. In fact, the law generally does impose damages equal to harm, but subject to some exceptions (which we will note in Sections 2.4.3 and 2.4.5).

2.4.2. *Nonpecuniary elements of loss*

Accidents often involve nonpecuniary losses, such as pain and suffering. To provide injurers with proper incentives to reduce accidents, they should pay for all nonpecuniary harms that they cause. However, it may be better for the state to receive these payments than for victims to receive them. Victims would often not elect to insure against nonpecuniary losses because these losses would not create a need for money, that is, raise their marginal utility of wealth¹⁸. Parents usually would not insure against the death of a child, for example, as this frequently would not generate a need for money, however devastating the loss would be for the parents. Thus, as initially proposed by Spence (1977), liability for pecuniary losses accompanied by an appropriate fine for nonpecuniary losses may be socially desirable¹⁹.

2.4.3. *Punitive damages*

When an injurer's behavior departs substantially from what is appropriate, damages in excess of harm, so-called punitive damages, may be imposed. If imposition of such

¹⁶ This point depends upon the particular formulation of the negligence rule (whether a person who takes less than due care is responsible for all harm caused or only the increment to harm resulting from x falling below \bar{x}). See Kahan (1989).

¹⁷ An additional issue is that erroneous findings of liability tend to remedy the problem of excessive levels of activity under the negligence rule, raising the possibility that setting damages above h would be desirable.

¹⁸ For empirical evidence, see Viscusi and Evans (1990).

¹⁹ Alternatively, victims might enter into contracts under which insurers would receive pain and suffering recoveries in exchange for a reduction in premiums on other coverage.

damages causes expected liability to exceed expected harm, injurers will be induced to take excessive precautions, at least under strict liability, and they will also reduce their levels of activity undesirably²⁰.

Damages exceeding liability are, however, desirable if injurers sometimes escape liability. This possibility arises because injurers may be hard to identify as the sources of harm (the origin of pollution may be difficult to trace) or because victims may not choose to bring suit (litigation costs may discourage legal action). If injurers who ought to be found liable for harm h are in fact only found liable and made to pay damages with probability q , then if damages are raised to $(1/q)h$, injurers' expected liability will be h . Thus, the more likely a party is to escape liability, the higher should be damages when the party is found liable. Accordingly, a firm that dumps toxic wastes at night, or an individual who tries to conceal a bad act, should have to pay punitive damages, but not an injurer who causes harm in a noticeable way. On these points and others, see for example Cooter (1989), Diamond (1997), and Polinsky and Shavell (1998a); and for empirical study, see Eisenberg et al. (1997), Karpoff and Lott (1999), and Polinsky (1997).

2.4.4. *Accuracy of damages*

Much expense is incurred in litigation about the magnitude of a victim's harm, which raises the question of what the social value of greater accuracy is and whether the private value of accuracy is different from the social value. As stressed in Kaplow and Shavell (1996b), the private value of accuracy about harm generally exceeds the social value. To explain, there is social value in establishing harm accurately primarily when injurers know, at the time that they choose their level of care, how much harm they might cause. For example, if an injurer anticipates that the atypically large harm he might cause will be accurately measured, he will exercise an appropriately high degree of care, as is socially desirable²¹. However, injurers often lack (and could not reasonably obtain) considerable information about the harm they might cause when they decide on their precautions. Drivers, for example, know relatively little about how much harm a potential victim would suffer in an accident (the seriousness of injuries, the magnitude of lost earnings). Thus, drivers' incentives to avoid accidents would be largely the same if, instead of using precise measurements of harm, courts employed rough averages (based, perhaps, upon abbreviated litigation over damages

²⁰ Under a perfectly operating negligence rule, punitive damages would not affect injurers' behavior, as explained in Section 2.4.1. But if there is uncertainty in the negligence determination, the problem of excessive precautions may be exacerbated by punitive damages; also, punitive damages may reduce injurers' activity levels (although this effect may be desirable).

²¹ Relatedly, a prospective injurer's incentive to acquire information about the harm he may cause (whether it will be atypically large) will be greater when he knows that harm will be accurately determined.

or upon figures from a table)²². Nevertheless, victims and injurers have very strong incentives to spend to establish damages accurately in court. A victim will always be willing to spend up to a dollar to prove that harm is a dollar higher, and an injurer will always be willing to spend up to a dollar to prove that harm is a dollar lower.

2.4.5. Components of loss that are difficult to estimate

Some components of loss are hard to estimate, for example, the decline in profits caused by a fire at a store (as opposed to the cost of repairing the store) or certain nonpecuniary harms, and the law sometimes excludes such difficult-to-measure elements of loss from damages. This legal policy might be justified when the cost of ascertaining a component of loss outweighs the value of the improvement in incentives that its inclusion would accomplish. However, the cost of estimating a component of loss would be low if rough estimates were used (and the analysis of the last section suggests that this often would not much compromise incentives to reduce risk). Therefore, the policy of excluding components of loss that are hard to evaluate may be unwarranted.

2.5. Causation

2.5.1. Basic requirement of causation

A fundamental principle of liability law is that a party cannot be held liable unless he was the cause of losses. For example, if cancer occurs in an area where a firm has polluted, the firm will in principle be liable only for the cancer that it caused, not for cancer due to other carcinogens.

This principle is clearly necessary to achieve social efficiency under strict liability, because otherwise incentives would be distorted. Socially desirable production might be rendered unprofitable if the firm were held responsible for all cases of cancer.

Under the negligence rule, restricting liability to accidents caused by an actor may be less important than under strict liability: if negligent actors were held liable for harms they did not cause, they would only have greater reason to act non-negligently, but would not take excessive precautions if there were no uncertainty surrounding the negligence determination. In the presence of such uncertainty, however, relaxation of the causation requirement might adversely affect incentives. Further, under both liability rules, absence of the causation requirement might raise the volume of litigation

²² A qualification to this point, emphasized by Spier (1994b), arises where the probability distribution of harm is affected by an injurer's degree of care. If this is so, then accuracy in assessing harm will influence an injurer's incentives to reduce risk even when, at the time he chooses his level of care, he does not have information about the harm that would occur in an accident.

and thus administrative costs. On the basic causation requirement and incentives, see originally Calabresi (1975) and Shavell (1980a)²³.

2.5.2. *Uncertainty over causation*

In many situations there is uncertainty about causation. For example, it may not be known which manufacturer out of many sold the product that resulted in injury, or whether harm was due to the defendant firm or to background factors (was cancer attributable to a firm's pollutant or to unknown environmental carcinogens?). The traditional approach of the law is to hold a defendant liable if and only if the probability that the defendant was the cause of losses exceeds 50%. This approach can lead either to inadequate or to excessive incentives to reduce risk. For example, a firm that supplies only 20% of the market demand will escape liability for any harm caused by its product (assuming that harm cannot be traced to particular firms). Consequently, the firm will have no liability-related incentive to take precautions. If, however, a firm's market share exceeds 50%, the firm will be liable for all harms due to the product that it and other firms sell, for it will always be correctly said to be more likely than not the cause of harm. Thus, the firm's liability burden will be socially excessive (under strict liability). These potential problems of inadequate and of excessive incentives may arise under any liability criterion based on a threshold probability of causation; they are not unique to a 50% threshold. Essentially this point has been made frequently, and it is formally developed in Shavell (1985b).

The legal system has recently adopted (in limited settings) the approach of imposing liability in proportion to the likelihood of causation. Under this approach, a firm supplying 20% of the market would be liable for 20% of harm in every case. Note, therefore, that the firm's liability bill would be the same under this regime as it would be if it paid for all the harm in the 20% of cases it truly caused – implying that its incentives would be socially appropriate. That the proportional liability principle engenders optimal incentives (without there being a need to establish causation in particular cases) is an advantage of the principle relative to the traditional threshold probability criterion. See Rosenberg (1984) and Shavell (1985b).

2.5.3. *Proximate causation*

Even if a party is a cause of losses, he may still escape liability under tort law because he was not the *proximate cause* of losses, where proximately-caused losses are, mainly, those that came about in an ordinary manner and that were not the product

²³ On the causation requirement under the negligence rule, see also Grady (1983) and Kahan (1989), who study restriction of liability to losses that are in excess of the possibly positive losses that the actor would have caused had he not been negligent.

of coincidence. Also, liability often is not found, on causal grounds, where accidents are freak events, such as where a dog imbibes nitroglycerin left at a mining site and then explodes, injuring nearby persons. Allowing parties to escape liability for such unusual accidents is sometimes thought not to undermine incentives, on the ground that no one could have foreseen such accidents. This argument, however, is subject to the criticism that courts may find it difficult to discriminate between accidents that can and cannot be foreseen. Moreover, the argument leads to the *reductio ad absurdum* that there should never be liability: any accident may be viewed as extraordinarily unlikely (of essentially zero probability) if it is described in sufficient detail.

The possibility that a party would not be said to be the proximate cause of losses on account of coincidence (as opposed to the freak character of losses) is illustrated by the following case: a speeding bus happened to be at just the “right” point on its route to be struck by a falling tree; the bus company escaped liability for the injuries to passengers even though they would not have occurred but for the excessive speed of the bus. Allowing parties to escape liability for such coincidental accidents might not affect precautions, however. One presumes that the probability of a bus being struck by a falling tree is independent of its speed, so that imposing liability would not affect the speed at which buses are driven. On proximate causation, see Calabresi (1975) and Shavell (1980a).

2.6. Judgment-proof problem

The possibility that injurers may not be able to pay in full for the harm they cause is known as the judgment-proof problem and is of substantial importance, for individuals and firms often pose risks significantly exceeding their assets (a person of modest means could cause a devastating fire; a small firm’s product could cause many deaths). When injurers are unable to pay fully for the harm they may cause, their incentives to reduce risk will be inadequate and their incentives to engage in risky activities too great. See Shavell (1986).

It should be remarked as well that injurers who may not be able to pay for the entire harm they cause will tend not to purchase full liability insurance, or any at all. This is because purchase of full coverage will involve the purchase of coverage against a loss that a party would not fully bear in the absence of coverage: if a person with assets of \$10 000 buys coverage against liability of \$100 000, he is purchasing coverage against \$90 000 of losses that he would not suffer if he did not have coverage. See Keeton and Kwerel (1984) and Shavell (1986).

Several types of policy response to the dilution of incentives caused by the judgment-proof problem are of interest. First, if there is another party who has some control over the behavior of the party whose assets are limited, then the former party can be held vicariously liable for the losses caused by the latter. Thus, holding a large contractor liable for the accidents caused by a small subcontractor or an employer for

accidents caused by its employees will induce the former to control the risks posed by the latter²⁴. See Kornhauser (1982) and Sykes (1981).

Second, parties with assets less than a specified amount could in some contexts be prevented from engaging in an activity. However, such minimum asset requirements are a somewhat blunt instrument for alleviating the incentive problems under consideration.

A third response to inadequate incentives, one closely related to asset requirements, is regulation of liability insurance. See Shavell (2000). One form of insurance regulation would mandate purchase of (perhaps full) coverage²⁵. This approach would be especially appealing when insurers can observe the precautions taken by injurers. An opposite form of insurance regulation would prohibit purchase of liability insurance. This could improve incentives to take care if insurers cannot observe injurers' precautions, because in that case insurance coverage would dilute incentives to take care when these incentives are inadequate to begin with.

Fourth, the use of Pigouvian taxes equal to expected harm may help to alleviate the judgment-proof problem. When harm will be caused with a low probability, the expected harm will be much less than actual harm; hence, parties with limited assets may be able to pay the appropriate tax on risk-creating behavior even though they could not pay for the harm itself.

A fifth way of correcting for dilution of incentives is for the state to regulate parties' behavior directly, such as with traffic laws or by insisting that food and drugs meet certain safety requirements. Regulation, however, may involve inefficiency because of regulators' limited knowledge of risk and of the cost and ability to reduce it. (We discuss regulation further in Section 3.6.2.)

A final way of mitigating dilution of incentives is resort to criminal liability. A party who would not take care if only his assets were at stake might be induced to do so for fear of imprisonment.

2.7. *Product liability*

We have not yet considered accidents where victims are customers of injurers (or, more generally, where victims are in some contractual relationship with injurers). In this case, the role of liability in providing incentives may be attenuated or even nonexistent. The reason, obviously, is that firms producing risky products may be unable to sell

²⁴ Imposing liability on corporations for behavior of their judgment-proof subsidiaries or requiring that liability of shareholders be unlimited (at least with respect to tort victims) might serve a similar function. See Hansmann and Kraakman (1991). Also, liability might be imposed on parties who supply services to potentially judgment-proof entities and are in a position to monitor them, such as accountants, lawyers, and lenders. See Feess (1999), Kraakman (1986), and Pitchford (1995).

²⁵ Many jurisdictions require liability insurance of those who drive, although the required amount is usually small.

them or may have to accept a reduction in price commensurate with the risk of loss attaching to the products.

If customer knowledge of product risk is perfect, then firms' incentives to reduce risk will be optimal even in the absence of liability. For example, if the expected losses caused by a product risk are \$100, the firm will have to accept a \$100 lower price than otherwise, so it will be willing to spend up to \$100 to eliminate the risk. Therefore, liability is not needed to generate incentives toward safety.

If, however, customer knowledge of risk is imperfect, liability is potentially useful in reducing risk²⁶. In the absence of liability, firms that increase safety generally will be unable to obtain an increase in the price fully reflecting the reduction in risk²⁷. (Indeed, in the extreme case where customers cannot observe anything about the true risk, firms would have no incentive to reduce it.) Therefore, the prospect of liability for product-caused harms will increase incentives to reduce risk. Also, imposing liability will result in prices that reflect the full costs of products, leading to more efficient purchasing decisions.

A question concerning liability is whether court-determined liability or customer-selected liability, namely, warranties, is likely to be better²⁸. The answer depends on the nature of customers' information or lack thereof and on other factors. For example, suppose that customers cannot directly determine the risk associated with a product but realize that firms will minimize production costs plus expected accident losses if they have to bear those losses. Then, consumers may rationally elect to purchase a full warranty – essentially to adopt strict liability – because they know that the product with that warranty will really be cheaper than an apparently similar product sold without the warranty at a lower price. In this case, warranty selection leads to optimality.

Suppose instead that customers misperceive risk. Then their selection of warranties may be skewed, as emphasized by Spence (1977). For example, if customers believe the risk of a product failure causing a loss of \$10,000 to be 1% when it is really 5%, then a warranty would not be purchased: a seller of a full warranty would have to charge \$500 for it, but the perceived expected value of it would be only \$100. In this circumstance, it might be better for the courts to impose liability because that would create incentives to reduce risk. However, in many contexts, customers can significantly reduce product risks by exercising care in the use of products, and producer liability might dilute their incentives to do so (assuming that defenses such as contributory negligence are unsuccessful because of difficulties in observing customers' behavior). Also, the administrative costs of liability are high. Thus, whether imposition of liability will improve social welfare, given customers' ability to purchase warranties, involves a complicated weighing of considerations, and courts' ability to do this is not clear.

²⁶ See, for example, Goldberg (1974). Another potentially useful policy is supplying information about risk to customers, on which see Magat and Viscusi (1992) and Viscusi and Magat (1987).

²⁷ However, as emphasized by Schwartz and Wilde (1979), this point depends on the fraction of customers who are uninformed about the risk.

²⁸ See generally Priest (1981) and Rubin (1993).

2.8. *Liability versus other means of controlling accidents*

Liability is only one means of controlling harm-causing behavior; safety regulation and Pigouvian taxes are among the alternatives, as we indicate in Section 2.6²⁹. For a general comparison of methods of controlling harm, see our discussion of regulating externalities in Section 3.6.2.

2.9. *Intentional torts*

To this point, we have examined liability for accidents, but we have not dealt explicitly with so-called intentional torts, such as assaulting someone or stealing his property (which also are crimes)³⁰. See Landes and Posner (1981). An intentional tort may be defined as a harm that an injurer causes in which either of two things are true: the injurer acted in a manner that caused harm to occur with a very high probability, or the injurer obtained utility from the victim's suffering itself.

It would be possible to apply the foregoing analysis of accidents to intentional torts without modification. The conclusions reached did not depend on the magnitude of the probability of harm or on the source of benefits to injurers. However, both of these aspects of intentional harms suggest changes in assumptions that could alter our analysis and conclusions.

First, in situations where harm would be very likely to occur, bargaining between injurers and victims would often be possible³¹. If so, it may be desirable to forbid injurers from harming victims unless they obtain consent in advance, presumably in exchange for payment. Thus, thieves would be required to buy property they want, rather than simply take it and pay damages under a liability rule. On these issues, see Calabresi and Melamed (1972) and Kaplow and Shavell (1996a).

Second, where injurers derive utility directly from the fact that harm is suffered by victims, some analysts suggest that injurers' utility should not count in assessing social welfare. If so, deterrence becomes more valuable. Also, it may be optimal to deter some harmful acts even when the injurer's benefit exceeds the victim's loss, which calls for damages greater than harm, or for supplemental sanctions, notably imprisonment. (One suspects, however, that with most such intentional torts, injurers' benefits rarely exceed victims' losses.)

3. **Property law**

We begin our discussion by reviewing reasons why property rights should exist and by describing instances of their emergence. Then we consider the major questions

²⁹ For comparisons of liability and safety regulation, see Kolstad, Ulen and Johnson (1990), Schmitz (2000), Shavell (1984a), and Wittman (1977).

³⁰ On crime, see Section 6.4.

³¹ We have implicitly ignored bargaining, except in our product liability discussion, because with most accidents – such as automobile accidents – bargaining between potential victims and potential injurers would be infeasible.

addressed by property law: the division and form of property rights, public property, the acquisition and transfer of property, and conflicts in the use of property (externalities). Last, we examine the subject of intellectual property. Many of the topics in this section have not been formally analyzed.

3.1. Justifications for property rights

A time-honored and fundamental question is why should there be any property rights in things³². That is, in what respects does the protection of property and the ability to transfer property promote social welfare? One justification for the protection of property is that it furnishes incentives to work, a common example being that people would not grow crops unless they could keep the product of their labor. Similarly, property rights provide incentives to maintain and improve durable things: to repair buildings, to fertilize and irrigate land, to conserve renewable resource stocks³³.

Another justification for property rights is that, were they absent, individuals would spend time and effort trying to take things from each other and protecting things in their possession, and they would often find themselves involved in conflict. Enforcement of property rights by the state, while involving its own costs, reduces these serious disadvantages that would be incurred in the absence of property rights. A related benefit of enforcing property rights is that it protects people against risk. In the absence of protection of property rights, individuals would face the possibility that their property would be taken from them (even though they might also enjoy the possibility that they would be able to take property from others).

In addition, it is important that a system of property rights allows for things to be transferred freely. Most obviously, if things can be traded, they will tend to be allocated to those who value them most³⁴. Moreover, the ability to transfer things is indirectly necessary to our enjoyment of economies of mass production and specialization of labor, for when a large quantity of a good is produced by a single entity, the output ultimately will have to be distributed, which is to say, transferred, to many other individuals, and the entity will also often need to obtain inputs from other parties. In addition, transferability of property (particularly of land) allows it to be used effectively as collateral, thus enabling credit markets to function³⁵.

³² A related question concerns how such rights should be protected. See Calabresi and Melamed (1972) and Kaplow and Shavell (1996a).

³³ Problems with conserving renewable resources that arise in the absence of property rights are often referred to as the tragedy of the commons. See Gordon (1954), Hardin (1968), Libecap (1998), and Ostrom (1998).

³⁴ We also note that protecting the security of property rights promotes the transfer of property: without protection of property rights, prospective buyers would not be inclined to buy things that might subsequently be stolen, and prospective sellers would be wary of making their ownership of valuable possessions known to others.

³⁵ Empirically oriented literature on the various benefits of property rights includes Alston, Libecap and Schneider (1996), Atwood (1990), Besley (1995, 1998), and Feder and Feeny (1991).

Early writing about property rights – by Bentham (1830), Blackstone (1765–1769), and Hobbes (1651), among others – stressed the justifications involving incentives to work and avoidance of strife. Today, the virtues of property rights seem to be taken for granted or are only casually asserted. Further, they are often conflated with the case for private property and the market system. This is a mistake, in that the various benefits from property rights that we mentioned could be enjoyed under a centrally planned economy. For example, incentives to work can be provided by paying workers on the basis of effort, even if a state enterprise owns what they produce. (Indeed, employees of profit-maximizing firms in private-enterprise economies are generally motivated by pay rather than by the literal ability to sell what they produce.) And the benefits of avoiding strife and theft might be enjoyed just as much under a centrally planned economy as under a market economy. The arguments for the social value of the market-enterprise system over central planning are different from those justifying the existence of property rights *per se*. (The arguments favoring market systems are based largely on the informational burdens that central planners face, problems of corruption, and the like.)

3.2. Emergence of property rights

We would expect property rights to emerge from a background of no rights or only poorly established rights when the various advantages of their existence substantially outweigh the costs of establishing and maintaining the rights³⁶. Property rights will be likely to arise in these circumstances because, if many individuals recognize that they will probably be better off under a regime with property rights, pressures will be brought to bear to develop them.

Various examples of the emergence of property rights have been studied. Umbeck (1981) examines property rights during the California Gold Rush. When gold was discovered in California in 1848, property rights in land and minerals were largely undetermined and there were virtually no authorities to enforce the law. Almost immediately, however, arrangements were made to protect property rights in gold-bearing land and river beds. This encouraged individuals to pan for gold, to build sluices, and otherwise to invest to extract gold; it also curbed wasteful efforts to grab land and gold from one another.

An additional example of historical interest is the establishment by the Indians of the Labrador Peninsula of rights in land where none had existed. Demsetz (1967) connects this 17th-century event to the increased value of furs. He suggests that without property rights in land, overly intensive hunting of fur-bearing animals (especially beaver) would have taken place and the stock of animals would have been depleted.

³⁶ We note that property rights can be established and enforced by the state or informally, through social norms. On the latter, see Ellickson (1989, 1991) and Sethi and Somanathan (1996). Property rights also might be enforced by private organizations, such as the Sicilian mafia. See Gambetta (1993).

A more recent instance of the emergence of property rights concerns resources of the sea, as described in Biblowit (1991) and Eckert (1979). For most of history, there were no property rights in the ocean's fisheries because the supply of fish was inexhaustible for all practical purposes, but fish populations have come under strain with the use of modern fishing methods. To provide incentives to preserve fisheries, it has come to be accepted that countries have property rights in fish found in their coastal waters. Also, property rights have recently been established in the sea bed to foster exploration and extraction of oil and mineral resources. Another important example of the appearance of property rights concerns rights to the electromagnetic spectrum; assignment of these rights prevents garbling of signals and encourages investment in programming and transmission as well as trade of rights to high-value users. See DeVany et al. (1969) and McMillan (1994)³⁷.

3.3. Division and form of property rights

3.3.1. Division of property rights

From a conceptual viewpoint, what we speak of somewhat loosely as property rights can be divided into more basic rights: particular possessory rights, and rights to transfer these rights. A possessory right in a thing is the right to use it in a specified way at a named time and under a particular contingency. A right to transfer a possessory right is the right to give or sell a possessory right to another person. Thus, what we commonly conceive of as "ownership" of something (say, land) entails both a large swath of possessory rights (rights to build on land, plant on it, and so forth, under most contingencies, and into the infinite future) and associated rights to transfer them.

In fact, property rights in things are generally held in substantially agglomerated bundles, but there is also significant partitioning of rights contemporaneously, according to time and contingency, and according to whether the rights are possessory or are for transfer. For example, an owner of land may not hold complete possessory rights, in that others may possess an easement giving them the right of passage upon his land, or the right to take timber, or the right to extract oil if found (thus a contingent right). A rental agreement constitutes a division of property rights over time; wills provide for future and often contingent division of rights (depending on the survival of beneficiaries). Trust arrangements, such as those under which an adult manages property for a child, divide possessory rights and rights to transfer.

The division of possessory rights may be valuable when different parties derive different benefits from them, because gains can then be achieved if rights are allocated to those who obtain the most from them. There are, however, several disadvantages to the division of possessory rights or too fine a division of the rights. Individuals may

³⁷ Literature on the emergence of property rights is surveyed in a general discussion of property rights and economic activity in Libecap (1986).

wish to exercise the same rights at the same time (a person with a right of passage may wish to use a path that is currently blocked by the owner's use); externalities and related conflicts may arise (a person with a right of passage might trample crops). In addition, logistical problems may impede the division of rights (consider the problem of many individuals trying to share the use of a single automobile).

We also note that possessory rights and rights to transfer are ordinarily combined because this promotes efficiency: possessors will make appropriate investments if they are the ones who will benefit from subsequent sales, and possessors will ordinarily have superior knowledge about which opportunities for sale are most profitable. Sometimes, however, separation of possessory rights and rights to transfer may be beneficial. A child may own property but not have the right to sell it because an adult trustee can make decisions superior to those of the child; a renter of an apartment may not have the right to sublet it because he does not have sufficient reason to consider the character of another tenant (such as whether the tenant would be likely to disturb neighbors).

3.3.2. Consolidated form of property rights and the theory of the firm

Ownership of separate productive assets is often consolidated; namely, it is held by a single entity, the firm. The question of what constitutes the benefits of this form of ownership was initially posed by Coase (1937) and has subsequently been developed by, among others, Williamson (1975, 1985), Klein, Crawford and Alchian (1978), Grossman and Hart (1986), and Hart (1995)³⁸. Here, we review the main factors that bear upon the relative advantages of separate versus consolidated holding of assets by firms³⁹.

First, consolidated ownership of assets reduces transaction costs because internal transfers of goods and services may be accomplished by command, eliminating the need for negotiation and bookkeeping expense⁴⁰. Such reduction of transaction costs, however, often could be obtained as well by separate owners if they entered into long-term supply contracts, honored standing orders, and the like.

Second, consolidated ownership may lead to a dilution of incentives to work, in comparison to the situation where each individual owns the assets he uses in production. Firms can combat this incentive problem in two familiar ways: if they can observe individuals' efforts, they can penalize shirking; if not, they

³⁸ See also Alchian and Demsetz (1972), Hart (1989), Hart and Moore (1990), Holmström and Tirole (1989), and Jensen and Meckling (1976). For a discussion of different forms of consolidated ownership (including employee-owned firms, cooperatives, and nonprofits), see Hansmann (1996).

³⁹ The subject of consolidated versus separate ownership of productive assets could be viewed as falling under the heading of division of property rights (separate ownership being division of consolidated ownership), but we find distinguishing the two subjects helpful.

⁴⁰ This savings may involve some sacrifice. For example, information on the profitability of separate functions may be lost (unless there is internal transfer pricing, which may involve transaction costs similar to those of market exchange).

can tie compensation to measures of output⁴¹. Of course, both methods have costs. (Interestingly, the latter may re-introduce transactions costs, such as if transfer pricing is required to compute a manager's contribution to the firm's profits.)

Third, consolidated ownership enables a firm to avoid breakdowns in bargaining that would occur under separate ownership due to asymmetric information. For example, under separate ownership, the seller of a factor input might overestimate its value to the next-stage producer and demand too much for it, stymieing an efficient transfer. Under consolidated ownership, efficient transfers can be ordered⁴². Alternatively, however, separate owners could contract in advance for transfers to occur at a predetermined price.

Fourth, consolidated ownership may help to alleviate problems of inadequate investment in assets. An asset owner may not have a sufficient incentive to make a relationship-specific investment (upgrading a plant for producing a factor input) because he anticipates that his gains will be partially expropriated by the owner of a complementary asset at the time when he is to put his asset to use. But if both assets are owned by the same party, the problem of expropriation of the gains from the relationship-specific investment in the first asset will be mitigated, and investment in it should be more efficient. However, the other individual's incentive to invest in what otherwise would have been his asset may be dulled if the first party owns both assets; thus, consolidated ownership does not necessarily improve investment incentives overall. Additionally, it may sometimes be possible under separate ownership of assets to guarantee that investments in them be sufficient by making a contract to that effect; but this requires that investments be observable.

We close by noting that the distinction between consolidated ownership of assets by firms and separate ownership is blurred because, as we have mentioned, under separate ownership together with contractual arrangements, it is often possible to replicate the advantages of firms. Indeed, separate ownership combined with sufficiently encompassing contracts may be indistinguishable from the consolidation of ownership of assets by firms. Conversely, firms themselves can be understood to consist of a set of contracts (a corporation is a particular contract among its shareholders).

⁴¹ Ellickson (1993), among others, suggests that most communal farming efforts have failed because individual rewards were not linked to effort or output, which led to widespread shirking. It may be observed as well that modern firms succeed despite their often large size through monitoring of workers' behavior (which may be more feasible with the use of mechanized technology) and use of performance pay.

⁴² The manager might know that the transfer is efficient without knowing the precise cost and/or value of the factor input, for the cost distribution may be below the value distribution. When distributions are, instead, substantially overlapping, a manager will not know whether a transfer is efficient, and in this case bargaining between separate parties may well promote efficiency.

3.4. Public property

Before continuing with our analysis of property rights, we consider briefly an important class of property: that owned by the public. We review the justifications for public property and then two methods of acquisition of such property: by purchase and by unilateral public taking.

3.4.1. Justifications for public property

The main justifications for public property concern problems with private supply. The government builds and maintains roads, for example, because private supply often would not be forthcoming due to difficulties that would be faced in collecting for road use. And even if roads were privately supplied, suppliers would charge tolls, raising problems of monopoly pricing and wasteful expenditures on toll-collecting.

Problems with private supply, however, do not constitute an argument for public ownership of goods, only for public financing of them or for public regulation of private suppliers. A road could be constructed, maintained, and owned by a private party paid by the state. And when private ownership might involve problems of monopoly pricing, government regulation is an alternative to direct ownership. These observations underlie the growing attention to privatization of public property and of government activities. The comparative virtues of public versus private ownership depend on the relative abilities of the government and of the private sector to operate efficiently and maintain quality⁴³.

3.4.2. Acquisition of public property: purchase versus compensated takings

The state may acquire property through purchase or through exercise of the state's power of eminent domain, which is to say, by taking the property. In the latter case, the law typically provides that the state must compensate property owners for the value of what has been taken from them, and it will be assumed that this is the case until the next section.

The difference between purchases and compensated takings is that the amount owners receive is determined by negotiation in the former case but unilaterally by the state in the latter case. Because of possible errors in governmental determinations as well as concerns about the behavior of government officials, purchase would ordinarily be superior to compensated takings. An exception, however, arises where the state needs to assemble many contiguous parcels, such as for a road. Here, acquisition by purchases might be delayed or prevented by hold-out problems, making the power to take socially advantageous.

The actual pattern of governmental acquisition of property largely reflects these simple observations. Most state acquisition of real estate, and virtually all acquisitions

⁴³ See for example Hart et al. (1997), Shleifer (1998), and Viscusi et al. (1995, pp. 468–470).

of moveable property, is through purchase. Governmental takings are restricted mainly to situations where there is a need for roads, dams, and parks, and to establish certain private rights-of-way, such as for railroads or utility lines⁴⁴.

3.4.3. *Compensation for takings*

Assuming that there is a reason for the state to take property, consider the effects and desirability of a requirement that the state pay compensation to property holders. As emphasized by Blume et al. (1984), payment of compensation to property owners creates a potential moral hazard: it leads them to invest excessively in property. For example, a person may build a home on land that might be taken by the state for use for a road because he will be compensated for the home if the land is taken. However, building the home might not be socially justified, given the probability of use of the land for a road, which would require destruction of the home.

A second effect of compensation for takings is that risk-averse property owners will bear less risk⁴⁵. But were takings not compensated, insurance against takings would be likely to emerge. Moreover, private insurance would naturally alleviate the problem of excessive investment in property⁴⁶.

Third, payment of compensation also may alter the incentives of public authorities to take property by reducing possible problems of overzealousness and abuse of authority. However, requiring compensation may also exacerbate potential problems of too little public activity (public authorities do not directly receive the benefits of takings). Therefore, it is not clear whether a compensation requirement improves the incentives of public authorities. For further discussion of these various issues about compensation for takings, see Kaplow (1986a, 1992a)⁴⁷.

3.5. *Acquisition and transfer of property*

We return now to the subject of private property and consider a number of topics relating to its acquisition and its transfer.

3.5.1. *Acquisition of unowned property*

Wild animals and fish, long-lost treasure, certain mineral and oil deposits, and, historically, unclaimed land, constitute primary examples of unowned property that

⁴⁴ See Bouckaert and De Geest (1995) on the related topic of private takings.

⁴⁵ It should be noted, however, that many property owners – namely, firms with diversified ownership – are not very risk averse.

⁴⁶ Notably, insurance premiums would be based on the value of property, so further investments would raise premiums.

⁴⁷ On the topic of compensation for loss in value of property due to regulation (as opposed to the physical taking of property), see Fischel (1995) and Miceli and Segerson (1996).

individuals may acquire. The law has to determine under what conditions a person will become a legal owner of such previously unowned property, and a general legal rule is that anyone who finds, or takes into his possession, unowned property becomes its owner.

Under this finders-keepers rule, incentives to invest in capture (such as to hunt for animals or explore for oil) are optimal if only one person is making the effort. However, if, as is typical, many individuals seek unowned property, they will invest a socially excessive amount of resources in search: one person's investment or effort usually will not simply increase the total probability of success, but rather will come, at least partly, at the expense of other persons' likelihood of finding unowned property⁴⁸.

Various aspects of the law governing the acquisition of property may be regarded as ameliorating this problem of excessive search effort under the finders-keepers rule⁴⁹. Notable examples are that regulations may limit the quantities that can be taken of fish and wild animals, the right to search for oil and minerals on the ocean floor may be auctioned, and oil extraction may be "unitized" (assigned to one party)⁵⁰.

3.5.2. *Loss and recovery of property*

When property is lost by its owner and is found by another person, the question arises whether the original owner should retain property rights or the finders-keepers rule should apply. The general stance of the law is that original owners maintain their property rights in lost things (unless they abandon them). This beneficially discourages original owners from socially excessive investment in preventing losses: a farmer might otherwise invest in an expensive fence to prevent his cattle from straying, which might be inefficient because often his private loss would not constitute a social loss (someone would be likely to find the strays). Moreover, original owners usually can either search themselves or efficiently organize recovery efforts by others (including by offering rewards). If, however, original owners cannot do this, the finders-keepers rule does have the advantage of inducing recovery effort, even though the rule tends to encourage races to find the effectively unowned property. In any event, if original owners retain property rights, finders may simply hide what they find, which reduces the value of what is found without producing the aforementioned benefits to original owners.

3.5.3. *Acquisition of stolen property and problems of establishing valid title*

A basic difficulty associated with sale of property that a legal system must solve is establishing validity of ownership, or "title". How does the buyer know whether the

⁴⁸ This problem is similar to the tragedy of the commons. See Gordon (1954) and Hardin (1968).

⁴⁹ For a survey of relevant literature, see Lueck (1998).

⁵⁰ Other laws limit indirectly how much property can be taken by individuals by giving them title only if they make productive use of the property that they find. This was true of homestead laws that gave land to individuals who worked it and of water rights regimes that gave priority to the extent that water supplies were regularly used. Such rules, however, create excessive incentives to exploit property.

seller has good title, and how does the buyer obtain good title? If these questions are not readily answered, sales transactions are impeded, and theft may be encouraged.

One route that legal systems may take involves the use of registration systems: lists of items and their owners. Important examples are registries of land, ships, motor vehicles, and many financial instruments. Presuming that an item is recorded in a registry, it will be easy for a buyer to check whether the seller holds good title to it, and the buyer will obtain title by having his name recorded in the registry as the new owner. Also, a thief obviously cannot claim that something he has stolen is his if someone else's name is listed as the owner in the registry. Registries are usually publicly established, and listing in registries often is mandatory (or it may be encouraged by making registration a condition to asserting a valid legal claim). Partial explanations for the public role in registries are the coordination problem that may be involved in creating them and the problem of insufficient private incentives to register property to provide a general deterrent against theft. (An individual contemplating registration will not take into account that, as the proportion of registered property rises, thieves anticipate that it will be more difficult to sell stolen property and thus are discouraged from theft.)

For most goods, however, registries do not exist because of the expense of establishing and maintaining them relative to the value of the goods and of the deterrence of theft. Two legal rules for determination of title are available (and both, to some extent, are employed) in the absence of registries. Under the original ownership rule, the buyer does not obtain good title if the seller did not have it; the original owner can always claim title to the item if he can establish his prior ownership. Under the bona fide purchaser rule, a buyer acquires good title as long as he had reason to think that the sale was bona fide (that the seller had good title) – even if the item sold was in fact previously stolen or otherwise wrongfully obtained. These rules have different effects on incentives for theft. Notably, under the bona fide purchaser rule, theft is made attractive because thieves will often be able to sell their property to buyers (who will be motivated to “believe” that the sale is bona fide); the buyers can use the now validly held property or resell it. Another social cost of the bona fide purchaser rule is that original owners will spend more to protect their property against theft because theft will be more frequent and, when it occurs, owners will be less likely to recover their property. (These costs of protection, note, are analogous to those arising under the rule allowing finders of lost property to keep it.) Finally, under the bona fide purchaser rule, buyers will not have an incentive to expend effort determining whether there exists a third-party original owner. This is an advantage in the direct sense that it reduces transaction costs, but it also compromises deterrence of theft.

3.5.4. Involuntary transfer of property: adverse possession

The legal doctrine of *adverse possession* effectively allows involuntary transfer of land (and some other types of property): a person who is not the owner of land is deemed to become its legal owner if he takes possession of it and uses it openly and

continuously for at least a prescribed period, such as ten years. Some have suggested that a rationale for the rule is that it permits the transfer of land from those who would leave it idle to those who will use it productively. But this overlooks the possibility that there may be good reasons for allowing land to remain idle (perhaps it will be built upon later, and thus an investment in it now would be a waste). Furthermore, a prospective adverse possessor could always bargain with the owner to rent or buy the land. Additionally, the rule suffers from the disadvantage that it induces landowners to expend resources policing incursions onto their land and it encourages others to attempt adverse possession. (Observe that these latter arguments are similar to those in the preceding sections that favored rules protecting original owners.)

A historical justification for the rule is that, before reliable land registries existed, it allowed a landowner to establish good title to a buyer relatively easily: the seller need only show that he was on the land for the prescribed period. Another advantage of the rule is that it reduces disputes that would arise where structures turn out to encroach on neighboring parcels⁵¹.

3.5.5. *Constraints on sale of property*

Legal restrictions are often imposed on the sale of goods and services. One standard justification for such policies is externalities. For example, the sale of handguns may be made illegal because of the externality that their ownership creates, namely, crime, and a tax may be imposed on the sale of a fuel because its use pollutes the air. See Section 3.6. The other standard justification for legal restrictions on sale is lack of consumer information. For instance, a drug may not be sold without a prescription because of fear that buyers would not use it appropriately. Here, though, one must compare the alternative of the government supplying relevant information to consumers (say that the drug has dangerous side effects, or that it should only be taken with the advice of a medical expert)⁵².

3.5.6. *Gifts*

The making of gifts, including bequests, is the major way in which property changes hands other than by sale. Gifts are, as one would expect, rather freely permitted because, like sales, they typically make both involved parties better off⁵³. It should be observed that, in the absence of a state subsidy, the level of giving may well fall short of the socially optimal level because a donor's private incentive to make a gift does not take into full account the donee's benefit. See Kaplow (1995b)⁵⁴. In addition, some

⁵¹ On adverse possession, see Netter (1998).

⁵² For further discussion, see Section 4.1.9 on legal overriding of contracts.

⁵³ There are some limits on disinheritance one's immediate family and other rules that prevent individuals from controlling the use of their gifts long into the future (the rationale for which is not entirely clear).

⁵⁴ See also Friedman (1988) on the gift externality.

gifts, particularly to charities, may support public goods or accomplish redistribution, which may provide a further ground for subsidy⁵⁵. In fact, the law does favor certain types of giving by conferring tax advantages on donees (and, in the case of charities, on donors). On the other hand, heavy gift and estate taxes are levied on large donative transfers to individuals.

Another issue concerning gifts is that a person may want to make a transfer in the future, in which case issues concerning contracts to give gifts arise. This subject will be discussed in Section 4.3.2 on donative contracts.

3.6. Conflicts in the use of property: externalities

3.6.1. Socially optimal resolution of externalities

When individuals use property, they may cause externalities, namely, harm or benefit to others. As a general matter, it is socially desirable for individuals to do more than is in their self-interest to reduce detrimental externalities and to act so as to increase beneficial externalities.

It should be noted, as emphasized by Coase (1960), that the socially optimal resolution of harmful externalities often involves the behavior of victims as well as that of injurers (and similarly with regard to generators of positive externalities and beneficiaries). Where victims can do things to reduce the amount of harm (install air filters to avoid pollution) more cheaply than injurers, it is optimal for victims to do so. Moreover, victims can sometimes alter their locations to reduce their exposure to harm. When the latter possibility is not incorporated into the analysis of externalities (suppose that victims are assumed to continue to live adjacent to a hazardous waste site), what is referred to as the optimal resolution of externalities may only be conditionally optimal.

3.6.2. Resolution of externalities through state intervention

We now consider various means of government intervention, along the lines of Shavell (1984a,c, 1993a)⁵⁶. For convenience, we confine our attention to the case of harmful externalities, and we assume (until the next section) that parties affected by externalities cannot bargain with the generators of externalities.

Under direct regulation, the state restricts permissible behavior. It might impose a quantity constraint (a fisherman may be required to limit his catch to alleviate depletion of the fishery) or other behavioral constraints (a factory may be required to use a smoke scrubber). Closely related to state regulation is privately-initiated regulation through use of the legal *injunction*, whereby a potential victim can enlist the power of the state to force a potential injurer to take steps to prevent harm or to cease his activity.

⁵⁵ See, for example, Atkinson (1976) on redistribution and charitable contributions.

⁵⁶ See also Bovenberg and Goulder (2002, Chapter 23 of this Handbook) on environmental taxation.

Society can also make use of financial incentives to induce injurers to reduce harmful externalities. Under the Pigouvian tax, a party pays the state an amount equal to the expected harm he causes, for example, the expected harm due to a discharge of a pollutant into a lake. An additional type of financial incentive is a subsidy, an amount paid by the state to a party equal to the reduction in expected harm from some benchmark level that he accomplishes.

There is also liability – a privately-initiated means of providing financial incentives – as we discuss in Section 2. Under strict liability, a party who causes harm has to pay the victim for his losses. (Such liability differs from the corrective tax because payment is to the victim rather than to the state, and also because injurers pay for actual harm rather than for expected harm.) Under the negligence rule, an injurer must pay the victim only if the injurer failed to take a cost-effective precaution.

In fact, liability and regulation are the preeminent tools that society uses to control externalities; the use of corrective taxes and subsidies is unusual. Since Pigou (1932), who first emphasized the problem of externalities, economists have focused on corrective taxes and regulation, essentially ignoring liability. We will now sketch some factors bearing on the relative desirability of these methods of controlling externalities. The review of factors will show that any of the methods (or a combination) could be the best, depending on the context.

One factor of relevance is the quality of the state's information. If the state has complete information about acts, that is, it knows the injurer's benefit or cost of precautions along with the victim's harm, then all of the approaches allow achievement of optimality. But if the state's information is imperfect, it will not be able to calculate which actions (such as installing a smoke scrubber) are desirable and thus sometimes will err. However, if the state knows the expected harm, it can induce injurers to act optimally under the corrective tax or a rule of strict liability, because the injurer, who is presumed to know the cost of a precaution, will then appropriately balance the cost against the reduction in expected harm that would be brought about⁵⁷.

We emphasize that this basic informational argument favoring Pigouvian taxes or strict liability over regulation or the negligence rule extends to the case where the state is uncertain about the magnitude of harm. The reason, essentially, is that under the former rules, the state only needs to estimate expected harm (as the injurers themselves implicitly supply complete information about the costs of precaution when making their decisions). By contrast, under regulation and the negligence rule the state must estimate *both* expected harm and precaution costs. Because the state's effectively available information is strictly better under the corrective tax or strict liability, it can achieve a superior outcome. (This point holds notwithstanding Weitzman's argument

⁵⁷ This advantage, as it applies to the comparison between strict liability and the injunction, is suggested by Calabresi and Melamed (1972) and is further explored in Kaplow and Shavell (1996a) and Polinsky (1980b).

suggesting that quantity regulation may be superior to corrective taxation⁵⁸.) An implication is that the use of pollution taxes is superior to the use of tradeable pollution permits because, under the latter, the government sets the total quantity of pollution using its own estimate of abatement costs rather than implicitly relying on firms' information⁵⁹.

A second factor is the information available to victims. For many externalities, victims have better information than the state about who is causing harm or about its extent – because they actually suffer the harm – so they are the most appropriate enforcement agents, suggesting the desirability of the liability tool or the injunction. In other instances, however, victims may be unaware of the harm or its cause, making the state a better enforcer. State enforcement, such as by regulation or by corrective taxes based upon statistical evidence of expected harm, avoids the need to identify, say, which pollutants ultimately harmed which victims.

A third factor concerns the level of activity of an injurer (how much a firm produces, how many miles a person drives), as opposed to the precautions an injurer takes given the level of activity (whether a firm uses a smoke scrubber while producing, whether a person exercises care when driving). Regulation and the negligence rule are most often concerned with precautions taken but not with the level of activity: a factory may be required by regulation to install smoke scrubbers but not to reduce its output. Thus injurers may not have incentives to moderate their level of activity although that would be desirable (their activity may result in harm despite the exercise of optimal precautions – even with smoke scrubbers, some pollution will result). By contrast, under the corrective tax and strict liability, injurers pay for harm done, so that they will optimally moderate their level of activity (as well as efficiently choose their level of precautions).

A fourth pertinent factor, noted above, is the ameliorative behavior of victims. Under regulation, corrective taxation, and other approaches that do not compensate victims for their harm, victims have a natural incentive to take optimal precautions (or to relocate) because they bear their residual losses; they will want to take any precaution (install air filters to reduce pollution) whose cost is less than the reduction in harm it accomplishes. Under a strict liability rule, however, a victim might not have such an incentive because he would be compensated for his losses. But under a negligence rule, victims are not compensated if injurers have behaved properly, and, under strict

⁵⁸ Weitzman's (1974) conclusion that regulation could be superior to taxation rests on his assumption that the state must, in advance, set a corrective tax rate that is independent of the quantity of pollution. Yet, when the marginal harm depends on the quantity of pollution, the optimal tax rate depends on the quantity of pollution. See Roberts and Spence (1976). Kaplow and Shavell (2002b) emphasize that taxes that depend on quantity are usually feasible to implement and are superior to quantity regulation.

⁵⁹ To be sure, tradeable permit regimes are themselves superior to quantity constraints imposed at the level of individual firms because trading allows a given total pollution target to be reached at minimum cost.

liability, compensation might be given only to victims who took optimal precautions (if this can be determined)⁶⁰.

Still another factor is administrative costs, the costs borne by the state in applying a legal rule and the legal and related costs borne by the affected parties (aside from direct costs, such as the costs of precautions). Liability rules possess a general administrative cost advantage over regulation in that under liability rules, administrative costs are incurred only if harm is done. This advantage may be significant when the likelihood of harm is small. Nevertheless, administrative costs will sometimes be lower under other approaches. For example, compliance with a regulation may readily be detected in some circumstances (determining whether factory smokestacks are sufficiently high would be easy) and also may be accomplished through random monitoring, saving enforcement resources. Also, imposing corrective taxes might be inexpensive. Notably, suppose that they are levied at the time of the purchase of a product. In contrast, liability rules might be expensive to employ. For example, demonstrating the source of a particular harm and its extent may be difficult. Also, when industrial pollution affects millions of individuals on an ongoing basis, the cost of a continuous flow of individual suits (or even class actions) that measure damages victim-by-victim is likely to be in excess of the cost of alternatives.

Last, the ability of injurers to pay for harm is of relevance. For liability rules to induce potential injurers to behave appropriately, injurers must have assets sufficient to make the required payments; otherwise they will have inadequate incentives to reduce harm, as discussed in Section 2.6. Where inability to pay is a problem, bonding requirements may be helpful, and regulation may become more appealing (although it may need to be enforced through the threat of nonmonetary, criminal sanctions). In addition, corrective taxes have an advantage over liability rules when harm is probabilistic because, under the corrective tax, an injurer would pay only the expected harm (with certainty) rather than the actual harm (if there is a 1% chance of causing \$1 000 000 of harm, the payment would be only \$10 000). Many firms that would be able to pay the tax and thus have correct incentives would not be adequately deterred under a liability rule, on account of their inability to pay for harm when it actually occurs.

3.6.3. Resolution of externalities through bargaining by affected parties

Parties affected by unregulated externalities will sometimes have the opportunity to make mutually beneficial agreements with those who generate the externalities. In the classic example, if a factory's pollution causes harm of \$1000 that can be prevented by installing a smoke scrubber that costs \$100, then, in the absence of any legal obligation on the factory, one might expect a potential victim of pollution to pay the factory to install the scrubber. An agreement for any amount between \$100 and \$1000

⁶⁰ For further discussion of this aspect of liability rules, see Section 2.1.

would be mutually beneficial. Let us first consider this possibility and then evaluate its significance⁶¹.

If it is posited that there are no obstacles to reaching a mutually beneficial agreement concerning externalities, then that will occur. This tautology is one version of the Coase Theorem; Coase (1960) stressed the point that externality problems could be remedied through private bargains. A closely related version of the Coase Theorem asserts that the outcome regarding the externality – whether a smoke scrubber is installed or instead pollution is generated – does not depend on the legal rule that applies. For example, if the scrubber costs \$100 and there is no law that controls pollution, a bargain as we have described it will come about and the scrubber will be installed; and likewise if there is a law that leads to installation of the scrubber, the same will happen⁶². The outcome, however, might be affected by the legal rule because of the level of wealth of parties. Most obviously, the potential victims might not have assets sufficient to pay for the scrubber, in which case the scrubber would not be installed unless a legal rule leads to this; moreover, legal rules may affect the distribution of wealth and thus the demand for goods, including that of being free from pollution⁶³.

There are, however, many obstacles to bargaining. Bargaining may fail to occur when victims are numerous and face collective action problems in coming together. This is often the situation with respect to victims of industrial pollution. Similarly, in important contexts, bargaining will be impractical because victims will not know in advance who will injure them; this is the case for automobile accidents and most other accidents between strangers. Another reason that bargaining may not occur is that victims might not know that they are exposed to a risk (such as from an invisible carcinogen). Also, of course, the cost of bargaining between just one potential victim and one potential injurer who know of each other can discourage them from engaging in the process. If these reasons do not apply and victims and injurers do engage in bargaining, asymmetry of information may lead to bargaining impasses; for example, where a victim thinks that a smoke scrubber would cost a factory only \$50 and it really costs \$100, he may offer too little to the factory to reach an agreement. In all, these problems that reduce the likelihood of bargaining occurring, and also its success if it does take place, make the importance of legal rules to remedy externalities substantial.

⁶¹ For experimental studies on bargaining and entitlements, see, for example, Hoffman and Spitzer (1982) and Croson and Johnston (2000).

⁶² Similarly, if there is a law permitting victims to enjoin factories from polluting but pollution does less harm than it costs to prevent, the factory would pay the victim to forgo the injunction, resulting in the same outcome – pollution – as would occur with no regulation of pollution.

⁶³ The outcome following from a legal rule might also be affected by an “endowment effect”, wherein individuals’ valuations depend on whether or not they originally enjoy legal protection. See Kahneman, Knetsch and Thaler (1990).

3.7. *Property rights in information*

Legal systems accord property rights in information, including inventions, books, movies, television programs, musical compositions, computer software, chip design, created organisms, and trademarks. The generation and use of such information, and therefore the law governing it, is growing increasingly important in modern economies. We divide our review of this subject into three parts: First, we discuss certain information like an invention that can be used repeatedly to produce something; here we discuss patent, copyright, and trade secret law. Second, we examine diverse other types of information and its legal protection. Third, we consider labels of various types and their protection under trademark law.

3.7.1. *Inventions, compositions, and other intellectual works of repeat value*

The classic forms of intellectual works that receive legal property rights protection are inventions and literary, musical, or other artistic compositions. The well-known description of socially optimal creation and use of such intellectual works is as follows. First, it is socially optimal for an intellectual work, if created, to be used by all who place a value on it exceeding the marginal cost of producing or disseminating the good (or service) embodying it; thus a new mechanical device should be used by all who place a value on it exceeding the cost of its manufacture, and a book by all who value it more highly than its printing cost. Second, an intellectual work should be created if the cost of doing so is less than its total value to the public, net of production cost.

Given this description of social optimality, the advantages and disadvantages of property rights in intellectual works are apparent. In the absence of property rights, a creator of an intellectual work will obtain profits from it only for a limited period – until competitors are able to copy the creator's work. Thus, the generation of intellectual works is likely to be suboptimal. But if there exist property rights, whereby a creator of an intellectual work obtains a monopoly in goods embodying the work, incentives to produce the works will be enhanced (although they will still be less than ideal because innovators do not capture all of the surplus that their works create)⁶⁴. The major drawback to intellectual property rights, however, is that monopoly pricing leads to socially inadequate production and dissemination of intellectual works⁶⁵. This problem can be severe where the monopoly price is much higher than the cost of production. A good example is computer software, which may be sold for hundreds of dollars a copy even though its cost of dissemination is essentially zero. Another problem (with

⁶⁴ Kitch (1977) emphasizes a somewhat different view, under which patent rights are often granted at an early stage of invention, and the rights allow their holders to develop the inventions into commercially viable products.

⁶⁵ Relatedly, subsequent innovators whose inventions depend on prior patented works will need to obtain licenses from existing patent-holders, and hold-up problems may arise. See Chang (1995), Green and Scotchmer (1995), and Heller and Eisenberg (1998).

patent rights in particular) is the race to be the first to develop intellectual works. Given that the rights are awarded to whoever is first, a socially wasteful degree of effort may be devoted to winning the race, for the private award of the entire monopoly profits may easily outweigh the social value of creating a work before a competitor does⁶⁶.

Patent law and copyright law are the most familiar forms of legal intellectual property right protection⁶⁷. The extent of protection afforded by each body of law is partial in various dimensions, however, so that they might be considered to represent a compromise between providing incentives to generate intellectual works and mitigating the monopoly problem. Patents and copyrights are limited in time (usually 20 years for patents, and the author's lifetime plus 50 years for copyrights) and also in scope⁶⁸. As an example of the latter, the copyright doctrine of fair use often allows a person to copy short portions of a copyrighted work. This probably does not deny the copyright holder significant revenues (a person would be unlikely to purchase a book just to read a few pages), and the transaction costs of the copier having to secure permission would be a waste and might discourage his use.

A distinct form of legal protection is trade secret law, comprising various doctrines of contract and tort law that serve to protect not only processes, formulas, and the like that might be protected by patent or copyright law, but also other commercially valuable information such as customer lists. An example of trade secret law is the enforcement of employment contracts stipulating that employees not use employer trade secrets for their own purposes. A party can obtain trade secret protection without having to incur the expenses and satisfy the legal tests necessary for patent or copyright protection. Also, trade secret protection is not limited in duration (Coca-Cola's formula has been protected for over a century). However, trade secret protection is in some respects weaker than patent protection; notably, it does not protect against reverse engineering or independent discovery. On the economics of trade secret law, see Friedman, Landes and Posner (1991).

An interesting and basic alternative to property rights in information is for the state to offer rewards to creators of information and for information that is developed to be made available to all who want it⁶⁹. Thus, under the reward system, an author of a book would receive a reward from the state for the writing of the book – possibly based on sales of the book – but anyone who wanted to print it and sell it could do so. Like the property rights system, the reward system encourages creation of information because the creator gains from producing intellectual works. But unlike the property rights system, the reward system results in the optimal dissemination of information

⁶⁶ The economic literature on intellectual property, focusing on patents, is discussed in Scherer and Ross (1990) and Tirole (1988); see also the historical review in Machlup (1958) and Reinganum's (1989) survey on the timing of innovation.

⁶⁷ See Besen and Raskind (1991), Gordon and Bone (2000), Landes and Posner (1989), and Menell (2000).

⁶⁸ See Gilbert and Shapiro (1990), Kaplow (1984), Klemperer (1990), and Scotchmer (1996, 1999).

⁶⁹ See Kremer (1998), Shavell and van Ypersele (2001), and Wright (1983).

because the intellectual works are placed in the public domain; anyone may use them for free. Hence, the reward system may seem to be superior to the property rights system. A major problem with the reward system, however, is that the state needs information about the value of innovations to determine rewards. We note that, to some degree, society does use a system akin to the reward system in that it gives grants and subsidies for basic research and for other intellectual works. But society does this largely when these intellectual works do not have direct commercial value.

3.7.2. Other types of information

There are many types of information different from what we have discussed above. One type of information is that which can be used only a single time, for example, where oil is located under a particular parcel of land. With regard to this type of information, there is sometimes no need for property rights protection. If the party who possesses the information can use it himself (to extract the oil), then once he does so, the issue of others learning it becomes moot – there will be no further value to the information. To the degree, though, that the party is unable to use the information directly (perhaps he cannot conveniently purchase drilling rights), his having property rights in the information might be valuable and beneficially induce the acquisition of information⁷⁰. Moreover, we observe that giving property rights in the information will not undesirably reduce the use of information when the optimal use of it is only once. In fact, the legal system usually does furnish property rights protection in such information as where oil is located through trade secret law and allied doctrines of tort and contract law⁷¹.

Another type of information is that relevant to future market prices. Here, the private and the social value of gaining such information can diverge, as emphasized by Hirshleifer (1971). For example, a person who first learns that a pest has destroyed much of the cocoa crop and that cocoa prices are therefore going to rise can profit by buying cocoa futures. The social value of his information inheres principally in any beneficial changes in non-financial behavior that it brings about. For example, an increase in cocoa futures prices might lead candy producers to reduce wastage of cocoa or to switch from chocolate production to production of another kind of candy. But the profit that a person with advance information about future cocoa prices makes can easily exceed its social value (suppose he obtains his information only an hour before it would otherwise become available, so that it has no social value) or fall short of its social value (suppose that he obtains information early on, but that his profits are low because he has limited funds to invest in futures). Hence, it is not evident whether it is socially desirable to encourage acquisition of such information about

⁷⁰ In addition, firms may need to be able to prevent employees from diverting a firm's benefit to themselves.

⁷¹ See also our discussion of disclosure in Section 4.1.2 on contract law.

price movements by giving individuals property rights in the information. The law does not generally discourage such information acquisition (but an exception is regulation of trading based on insider information), and the law often encourages acquisition through trade secret protection⁷².

Last, consider information of a personal nature about individuals. The cost of acquiring this information is the effort to snoop, although the information is sometimes adventitiously acquired, so costless. The social value of the information involves various complexities. The release of information of a personal nature to the outside world generally causes disutility to those persons exposed and utility for others, the net effect of which is ambiguous. Further, a person's behavior may be affected by the prospect of someone else obtaining information about him: he may be deterred from socially undesirable behavior (such as commission of crimes) or from desirable but embarrassing-if-publicly-revealed behavior, and he may make costly efforts to conceal his behavior. Thus, there are reasons why the acquisition and revelation of personal information are socially undesirable, and reasons as well why they might be socially beneficial. The law penalizes blackmail and in this way attempts to discourage profit from acquisition of personal information⁷³. But otherwise the law does not generally retard the acquisition of personal information, and it also extends limited property rights in such information; notably, an individual who wants to sell to a publisher personal information he has obtained usually can do so.

As this brief discussion has illustrated, the factors bearing on the desirability of protecting property rights in information vary significantly according to the type of information and call for analysis quite different from that concerning information of repeat value that we considered above.

3.7.3. *Information valuable as labels*

Many goods and services are identified by labels. The use of labels has substantial social value because the quality of goods and services may be hard for consumers to determine directly. Labels enable consumers to make purchase decisions on the basis of product quality without going to the expense of independently determining their quality (if this is even possible). A person who wants to stay at a high-quality hotel in another city can choose such a hotel merely by its label, such as "Ritz Hotel"; the consumer need not directly investigate the hotel. In addition, sellers who label their output will have an incentive to produce goods and services of quality because consumers will recognize quality through sellers' labels. The existence of property rights in labels – that is, the power of holders of the rights to prevent other sellers from using holders' labels – is necessary for the benefits of labels to be enjoyed.

In view of the social value of property rights in labels, it is not surprising that the legal system allows such rights, according to trademark law. Also, trademarks are

⁷² On insider trading, see Leland (1992) and Scott (1998).

⁷³ See Ginsburg and Shechtman (1993), R.A. Posner (1993b), and Shavell (1993c).

of potentially unlimited duration (unlike patents or copyrights), which makes sense because the rationale for their use does not wane over time. The guiding principle of trademark protection is prevention of consumer confusion, so that a new trademark that is so similar to another (Liz Clayborne and Liz Claiborne) that it would fool people would be barred, but an identical trademark might be allowed if used in a separate market. Trademarks are required to be distinctive words or signs, for otherwise normal usage would be encumbered. (If a restaurant obtained a trademark on the words “fine food”, other restaurants would be limited in their ability to communicate.) On the economics of trademark law, see Landes and Posner (1987b).

4. Contracts

The private and social functions of contracts and of contract law are examined here. In Section 4.1 the basic theory of contracts is considered, in Section 4.2 production contracts (which have been the focus of a substantial literature) are analyzed, and in Section 4.3 several other types of contract are discussed.

4.1. Basic theory

4.1.1. Definitions and framework of analysis

A contract is a specification of the actions that named parties are supposed to take at various times, as a function of the conditions that then obtain. The actions usually comprise delivery of goods, performance of services, and payments of money, and the conditions include uncertain contingencies, past actions of parties, and messages sent by them.

A contract is said to be complete if the list of conditions on which the actions are based is exhaustive, that is, if the contract provides *explicitly* for all possible conditions. Otherwise, a contract will be referred to as incomplete. Typically, incomplete contracts do not include conditions that, were they easy to include, would allow both parties to be made better off in an expected sense. It should be noted that an incomplete contract may well not have literal gaps in that it will cover all conditions, at least by implication. Consider, for example, a contract stating merely that a specified price will be paid for a bushel of wheat. Although this contract is incomplete because it does not mention many contingencies that might affect the buyer or the seller of wheat, it has no gaps, as it stipulates what the parties are to do (pay a price, deliver a bushel of wheat) in all circumstances.

A contract is Pareto efficient if it is impossible to modify in a manner that raises the expected utility of both of the parties; such a contract will sometimes be referred to simply as efficient or as mutually beneficial.

Contracts are assumed to be enforced by a tribunal, which will usually be interpreted to be a state-authorized court, but it could also be another entity, such as an arbitrator or

the decision-making body of a trade association or a religious group. (Reputation and other non-legal factors may also serve to enforce contracts but will not be examined here⁷⁴.) Enforcement refers to actions taken by the tribunal when parties to the contract decide to come before it. Tribunals may impose money sanctions – so-called damages – for breach of contract or insist on *specific performance* of a contract – require parties to do what a contract specifies (for example, convey land). Tribunals may also fill gaps, settle ambiguities, and override terms in contracts.

4.1.2. Contract formation

The formation of contracts is of interest in several respects.

Search effort. Parties expend effort in finding contracting partners, and it is apparent that their search effort will not generally be socially optimal. On one hand, they might not search enough: because the surplus gained when one party locates a contract partner will generally be divided between them in bargaining, the private return to search may be less than the social return. On the other hand, parties might search more than is socially desirable because of a negative (“common pool”) externality associated with discovery of a contract partner: when one party finds and contracts with a second, other parties are thereby prevented from contracting with that party⁷⁵. Both of these externalities arise in Diamond and Maskin (1979), who examine a specific model of search and contracting. Although policies to promote or to discourage search might be desirable, one wonders whether social authorities could obtain the information needed to determine the nature of problems with search effort.

Mutual assent and legal recognition of contracts. A basic question that a tribunal must answer is at what stage of interactions between parties does a contract become legally recognized, that is, become enforceable. The general legal rule is that contracts are recognized if and only if both parties give a clear indication of assent, such as signing their names on a document. This rule obviously allows parties to make enforceable contracts when they so desire. Moreover, because the rule requires mutual assent, it protects parties against becoming legally obligated against their wishes. Thus, it prevents the formation of what would be undesirable contracts, and it means that search for contracting partners will not be chilled due to the risk of unwanted legal obligations.

However, certain legal doctrines sometimes result in parties becoming contractually bound without having given their assent; there exist cases in which a party became contractually bound when the other party with whom he was negotiating made substantial investments in anticipation of contract formation. This legal policy not only may result in undesirable contracts, it may also induce wasteful early investment as a strategy to achieve contract formation. It is true that early investment is sometimes

⁷⁴ See, for example, Bernstein (1992, 1998), Charny (1990), Greif (1998), and Klein and Leffler (1981).

⁷⁵ Compare our discussion in Section 3.5.1 of excessive incentives to search for unowned property.

efficient, but a party who wants to make early investment could attempt to advance the time of contract formation or make a preliminary contract about the matter. See Bebachuk and Ben-Shahar (2001), Craswell (1996), Katz (1996), and Wils (1993).

Offer and acceptance. Mutual assent sometimes is not simultaneous; one party will make an offer and time will pass before the other agrees. An issue that this raises is how long, and the circumstances under which, the offeror will want to be held to his offer, and whether he should be held to it. If an offeror is held to his terms, offerees will often be led to invest effort in investigating contractual opportunities. Otherwise, offerees might be extorted by offerors if the offerees expressed serious interest after investigation. The anticipation of such offeror advantage-taking would reduce offerees' incentive to engage in investigation and thus diminish mutually beneficial contract formation. Hence, it may be in offerors' and society's interests for offered terms to be enforced for some period of time. Yet offerors' circumstances may change, making it privately and socially advantageous for them to alter contract terms. On this and other issues concerning offer and acceptance, see Craswell (1996) and Katz (1990b, 1993).

Disclosure. The law may impose an obligation to disclose private information at the time of contract formation⁷⁶. Such a legal duty is beneficial in the respect that disclosed information may be desirably employed by the buyer; suppose, for instance, that he learns from the seller that the basement of his new house leaks and thus decides not to store valuables there. However, as initially emphasized by Kronman (1978a), a disclosure obligation discourages parties from investing in acquisition of information. For example, a company might decide against conducting aerial surveys to determine the mineral-bearing potential of land if it would be required to disclose its findings to sellers of land, as sellers would then demand a price reflecting the value of the land. The social welfare consequences of the effect of a disclosure obligation on the motive to acquire information, analyzed in Shavell (1994), depend on whether the information is socially valuable or mere foreknowledge, on whether the party acquiring information is the buyer or the seller, and on inferences that would be made from silence⁷⁷.

Duress and emergency. Even if both parties have given their assent, a contract will not be recognized if it was made when one of the parties was put under undue pressure, as when he is physically or otherwise threatened by another. This legal rule has virtues similar to those of laws against theft; it reduces individuals' incentives to expend effort making threats and to defend against them.

In addition, contracts may not be legally recognized if they are made in emergency situations, such as when the owner of a ship in distress promises to pay an exorbitant amount for rescue. Nonenforcement in such situations beneficially provides victims with implicit insurance against having to pay high prices, but it also reduces incentives

⁷⁶ For discussions of various ways that asymmetry of information affects the contract terms that parties will agree to when there is no compulsory disclosure, see Ayres and Gertner (1989), Bebachuk and Shavell (1991), Spier (1992b), and Stole (1992).

⁷⁷ On inferences from silence in other contexts, see Fishman and Hagerty (1990), Grossman (1981), and Milgrom (1981), and for an empirical study of mandatory disclosure, see Mathios (2000).

for rescue (yet rescue incentives might tend to be excessive, for the same reasons that there is excessive fishing effort)⁷⁸.

4.1.3. Why contracts and their enforcement are valuable to parties

At the most general level, parties make contracts when they have a need to make plans. They want contracts enforced to ensure that promised payments are made and to prevent opportunistic behavior that otherwise might occur over the course of the contractual relationship and stymie fulfillment of their plans. There are two basic contexts in which parties make enforceable contracts.

The first is that concerning virtually any kind of financial arrangement. The necessity of contract enforcement here is transparent. For example, because borrowers would not be forced to repay loans in the absence of contract enforcement, loans would be unworkable without enforcement. In financial arrangements, there is often a party who extends credit to another for some time period, and contract enforcement prevents his credit from being appropriated, which would render the arrangements impossible. In addition, financial contracts that allocate risk would generally be made useless without enforcement because, once the risky outcome became known, one of the parties would not wish to honor the contract.

The second context in which parties make enforceable contracts involves the supply of custom or specialized goods and services – those which cannot simply be purchased on a spot market in a simultaneous exchange for money. The need for enforcement of agreements for supply of custom goods and services inheres mainly in averting what is often described as the holdup problem (discussed further in Sections 4.2.1 and 4.2.2). To illustrate, consider a buyer who wants a custom desk that would be worth \$1000 to him and would cost \$700 for a seller to produce. In the absence of contract enforcement, the buyer will not pay the seller in advance (for the seller could walk away with what he receives). The buyer will pay the seller only after the seller makes the desk. But at that point, the seller's production cost is sunk and he is vulnerable to holdup; the situation is that he has a desk that, being custom-made, has little or no alternative value⁷⁹. The outcome of bargaining between him and the buyer might thus be a price lower than the seller's cost of \$700; say the price is \$500. If so, and the seller anticipates receiving only the \$500 price, he will not produce the desk. This is true even though production and sale at a price between \$700 and \$1000, such as \$800, would be mutually beneficial for the seller and the buyer. Enforcement of the buyer's promise to pay \$800 for the desk on delivery, or of the seller's promise to produce and

⁷⁸ On rescue, see Landes and Posner (1978).

⁷⁹ Similar forms of holdup would arise in the absence of contract enforcement where parties want to convey property that already exists, such as land; for instance, a seller might worry about being held up by the buyer if he waits and forgoes a present opportunity to sell his land to a new party who makes a bid for it.

deliver the desk (if the buyer paid the price of \$800 in advance), is thus desirable for the parties.

More broadly, enforcement of contracts will stimulate all manner of investments that, like the seller's expenditure on production, have specific value in a contractual relationship. Enforcement will lead buyers to train workers to use new contracted-for equipment, sellers to engage in research to reduce production costs, and so forth. In the absence of contract enforcement, there would be too little investment in these things, for, at the final stage of negotiation for performance and for payment, each side would be subject to holdup by the other, so would tend to obtain only a part of the surplus created by its investment.

The foregoing idea of contract enforcement as a cure for holdup-related underinvestment was initially stressed in the economics literature by Klein, Crawford and Alchian (1978), Grout (1984), and Williamson (1975). However, the general notion that contract enforcement is privately and socially desirable because it fosters production and trade is made (usually with little articulation) by most writers on contract law and, one supposes, has always been appreciated. See, for example, Farnsworth (1982, pp. 16–17) and Pound (1959, pp. 133–134).

4.1.4. Incomplete nature of contracts and their less-than-rigorous enforcement

Although enforceable contracts are desirable, they are observed to be substantially imperfect. They are significantly incomplete, leaving out all manner of variables and contingencies that are of potential relevance to contracting parties, and they also often fail to employ included variables in a mutually beneficial manner. Moreover, contracts are not enforced rigorously, despite the seeming strength of the reasons for contract enforcement: penalties for violation of contractual obligations are often modest, and breach is not an uncommon event.

There are three important reasons for the incompleteness of contracts. The first is the cost of writing more complete contracts. Parties may not include variables in a contract, or not in a detailed, efficient way, due to the cost of evaluating, agreeing upon, and writing terms. (In particular, parties will tend not to specify terms for low-probability events, because the expected loss from this exclusion will be minimal, whereas the cost of including the terms is borne with certainty.)

The second reason for incompleteness is that some variables (effort levels, technical production difficulties) cannot be verified by tribunals⁸⁰. Of course, many such variables can be made verifiable (effort could be made verifiable through videotaping), but that would involve expense.

The third reason for the incompleteness of contracts is that the expected consequences of incompleteness may not be very harmful to contracting parties.

⁸⁰ The problem of unverifiability of variables is diminished by the possibility that parties can plan in their contract to use a tribunal of experts in their area, such as individuals in the same business as the contracting partners. In many industries, this practice is common.

Incompleteness may not be harmful simply because a tribunal might interpret an imperfect contract in a desirable manner. In addition, as we shall see, the prospect of having to pay damages for breach of contract may serve as an implicit substitute for more detailed terms. Furthermore, the opportunity to renegotiate a contract often furnishes a way for parties to alter terms in the light of circumstances for which contractual provisions had not been made. Finally, in some settings parties' concern for their reputation may induce them to refrain from opportunistic behavior.

That contracts are less than rigorously enforced is intimately related to their incompleteness. For incomplete contracts not to disadvantage parties, tribunals must be able to reinterpret or override imperfect contractual terms rather than always enforce these terms as written. Also, for damage measures for breach to be employed beneficially by parties, notably for parties to be able to escape from contractual obligations when performance and renegotiation are difficult, damages payments must not be excessive. Additionally, for parties to avoid bearing high risks in the form of payments that they would be induced to make when renegotiating imperfect contractual terms, the damages for breach must again not be severe. These points will be expanded in the discussion below of contract interpretation, remedies for breach, and renegotiation.

4.1.5. Interpretation of contracts

Contractual interpretation, which includes a tribunal's filling gaps, resolving ambiguities, and overriding literal language, can benefit parties by easing their drafting burdens or reducing their need to understand contractual detail⁸¹. For example, if it is efficient to excuse a seller from having to perform if his factory burns down, the parties need not incur the cost of specifying this exception in their contract, assuming that they can trust the tribunal to interpret their contract as if the exception were specified⁸².

It may be worthwhile elaborating somewhat by viewing contract interpretation more formally, as a function that transforms the contract individuals write into the effective contract that the tribunal will enforce. Given a method of interpretation, parties will choose contracts in a constrained-efficient way. Notably, if an aspect of their contract would not be interpreted as they want, the parties would either bear the cost of writing a more explicit term that would be respected by the tribunal, or else they would not

⁸¹ On various aspects of contract interpretation, see, for example, Ayres and Gertner (1989), Hadfield (1994), and Schwartz (1992).

⁸² Another example, where it may be efficient for a tribunal to override particular terms that appear in contracts, is when a seller offers only a detailed, fine-print contract in conjunction with the sale of an inexpensive good or service. Because it would be irrational for consumers to read such contracts, sellers would have incentives to include inefficient, one-sided terms if such terms would be enforced. See Katz (1990c). The extent to which such contracts will be problematic will depend on the fraction of consumers who are informed about contract terms and shop among competing sellers. See Schwartz and Wilde (1979).

bear the cost of writing the more explicit term and accept the expected loss from having a less than efficient term. The best method of contract interpretation will take this reaction of contracting parties into account and can be regarded as implicitly minimizing the sum of the costs the parties bear in writing contracts, the losses resulting from inefficient enforcement, and adjudication costs⁸³.

4.1.6. *Damage measures for breach of contract*

When parties breach a contract, they often have to pay damages in consequence. The damage measure, the formula governing what they should pay, can be determined by the tribunal or it can be stipulated in advance by the parties to the contract⁸⁴. One would expect parties to specify their own damage measure when it would better serve their purposes than the measure the tribunal would employ, and otherwise to allow the tribunal to select the damage measure. In either case, we now examine the functioning and utility of damage measures to contracting parties (assuming here that there is no renegotiation of contracts).

Clearly, the prospect of payment of damages is an incentive to perform contractual obligations, and thus generally promotes enforcement of contracts and the goals of the parties, as discussed in Section 4.1.3. As emphasized in Section 4.1.4, however, damages for breach in fact are not chosen to be so high that they virtually guarantee performance of contracts as written. Under the commonly employed measure of *expectation damages*, damages equal the amount that compensates the victim of breach for his losses.

Why are damages not chosen to be so high as to guarantee performance? An important explanation is that parties do not always want performance of the less-than-complete contracts that they write. For example, suppose that a contract is very incomplete: it merely states, "The seller will produce a custom desk for the buyer and receive full payment of \$800 in advance." The buyer and the seller do not really want the desk always to be produced. It is readily shown that, had they made a Pareto-efficient complete contract, they would have specified that there should be performance if and only if the production cost is less than the \$1000 value of the desk to the buyer. (For instance, in a complete contract, they would have jointly decided against a contractual term specifying performance when the production cost is \$2000, for the

⁸³ The determination of the optimal method of interpretation may involve subtleties. For example, according to the optimal method, a term might not be interpreted in the way that is best in the majority of transactions. Suppose that term *A* is best in the majority of transactions and that the parties to these transactions can include *A* explicitly, at little cost on a per-contract basis, because they are repeat players. Suppose that term *B* is best only in the minority of transactions, but that for the parties to these transactions to include *B* explicitly will not be cheap on a per-contract basis because they are not repeat players. Then the optimal method of interpretation would make *B* the default term even though it is best in only a minority of transactions.

⁸⁴ A contractual provision that states a particular amount of damages is referred to as a *liquidated damages* clause.

seller would have been willing to reduce the contract price sufficiently to induce the buyer to strike the term.) Now if the incomplete contract calling for the desk always to be produced is enforced by the expectation measure of damages of \$1000, the seller will behave exactly as he would have under the Pareto-efficient complete contract, that is, he will perform if and only if the production cost is less than \$1000. Higher damages than the expectation measure might induce performance when it is inefficient, and lower damages might lead to breach when that is inefficient. Indeed, for this reason, the parties would often agree to choose the expectation measure over other measures of damages.

This understanding of damage measures as a device to induce the behavior that the parties would have specified in more complete contracts sheds light on the notion held by some legal commentators and philosophers that contract breach is immoral, that it constitutes the breaking of a promise. That belief is often incorrect, it is submitted, and might fairly be considered to be the opposite of the truth. The view that a contract breach is the breaking of a promise overlooks the point that the contract that is breached is generally an incomplete contract, and that the "breach" constitutes behavior that the parties truly want and would have provided for in a complete contract. In the example of the simple incomplete contract calling for a desk to be produced, the seller who finds that his production cost would be \$2000 will commit breach under the expectation measure. But in so doing, he will be acting precisely as would have been set out in a Pareto-efficient complete contract, and it is that contract which is best regarded as the promise between the parties that ought to be kept.

The point that a moderate damage measure, and in particular the expectation measure, is desirable because it induces performance if and only if the cost of performance is relatively low was apparently first clearly stated (informally) in R.A. Posner (1972), who emphasized the social efficiency of the measure. Shavell (1980b) formally demonstrated this and also stressed the mutual desirability of the expectation measure for contracting parties and its role as a substitute for more complete contracts⁸⁵.

Several more comments should be made about damage measures and incentives. First, damage measures influence the motive of contracting parties to make reliance investments (so called because the investments are made relying on contract performance). Reliance investments are illustrated by the earlier-noted instance of a buyer training workers to use a contracted-for machine or by advertising the contracted-for appearance of an entertainer. Under the expectation measure, there is a tendency for reliance investment to exceed the Pareto efficient level: the buyer will treat an investment like advertising as one with a sure payoff – either he will receive performance or receive expectation damages, a form of insurance – whereas the actual

⁸⁵ Two other writers, Birmingham (1970) and Barton (1972), adumbrate these points, although the meaning of their articles is at times obscure. See also Diamond and Maskin (1979), who consider damage measures in analyzing search behavior.

return to investment is uncertain, due to the possibility of breach (advertising will be a waste if the entertainer does not appear). This tendency toward overreliance due to the receipt of contract damages was initially noted in Shavell (1980b), and stands in contrast to the problem of inadequate reliance investment associated with lack of contract enforcement. The issue of reliance investments has been elaborately analyzed, as will be described in Section 4.2.2.

A second comment is that the value of damage measures as an incentive toward efficient performance would not exist if renegotiation of contracts in problematic contingencies would always result in efficient performance. But, as will be discussed below, it seems plausible that renegotiation would not always result in efficiency.

An important function of damage measures that is quite distinct from their incentive role concerns risk-spreading and compensation. Notably, because the expectation measure compensates the victim of a breach, the measure might be mutually desirable as a form of insurance if the victim is risk averse. However, the prospect of having to pay damages also constitutes a risk for a party who might commit breach (such as a seller whose costs suddenly rise), and he might be risk averse as well. The latter consideration may lead parties to want to lower damages (see Polinsky 1983) or to avoid use of damages as an incentive device, by writing more detailed contracts (for instance, the parties could go to the expense of specifying in the contract that a seller can be excused from performance when his costs are high)⁸⁶. A full consideration of damage measures and efficient risk allocation would also take into account whether the risk that a party bears is detrimental or beneficial⁸⁷, whether the risk is monetary or non-monetary⁸⁸, and whether the parties can obtain insurance.

4.1.7. Specific performance as a remedy for breach

As observed at the outset, an alternative to use of a damage measure for breach of contract is specific performance: requiring a party to satisfy his contractual obligation⁸⁹. Specific performance can be accomplished with a sufficiently high threat or by exercise of the state's police powers, such as by a sheriff removing a person from the land that he promised to convey. (Note that if a monetary penalty can be employed

⁸⁶ When parties do not so specify in advance, certain legal doctrines may serve this function. See Joskow (1977), R.A. Posner and Rosenfield (1977), and Sykes (1990).

⁸⁷ For example, if a party wants to breach because he has a superior opportunity, optimal damages might be higher, although adjusting damages in the case of beneficial risks is not likely to matter as much on risk-bearing grounds.

⁸⁸ For example, if the victim's loss is non-monetary, such as the loss due to failure of musicians to appear at a wedding, financial compensation in the form of damages may not constitute an optimal form of insurance. See Section 2.4.2.

⁸⁹ Some economists have employed the term "specific performance" in an unconventional sense, to refer to enforcement of all provisions in a contract, including any damage measure named in it. Thus, they would say that a contract is specifically performed when the parties name expectation damages in their contract and parties who breach are thus required to pay these damages.

to induce performance, then specific performance is equivalent to a damage measure with a high level of damages.)

It is apparent from what has been said about incomplete contracts and damage measures that parties should not want specific performance of many contracts that they write, for they do not wish their incomplete contracts always to be performed. It is therefore not surprising that, in fact, specific performance is not used as the remedy for breach for most contracts for production of goods and for provision of services. Additionally, it may be observed that specific performance might be peculiarly difficult to enforce in these contexts because of problems in monitoring and controlling parties' effort levels and the quality of production.

However, specific performance does have advantages for parties in certain contexts, such as in contracts for the transfer of things that already exist, like land, and specific performance is the usual legal remedy for sellers' breaches of contracts for the sale of land. This point is discussed briefly below, in Section 4.3.1. On specific performance and its general comparison to damage remedies, see Bishop (1985), Kronman (1978b), Schwartz (1979), Shavell (1984b), and Ulen (1984). (Specific performance also is examined in some of the articles on production contracts cited in Section 4.2.2.)

4.1.8. Renegotiation of contracts

Parties often have the opportunity to renegotiate their contracts when problems arise. Indeed, the assumption that they will do this has appeal because, having made an initial contract, the parties know of each other's existence and of many particulars of the contractual situation. For this reason, much of the economics literature on contracts assumes that renegotiation always occurs when inefficiency would otherwise result; see, for example, Hart (1987), Hart and Holmström (1987), and Rogerson (1984).

Nevertheless, in many circumstances contracts will not be renegotiated because parties are not in contact with each other when difficulties are experienced and one party would benefit from acting quickly. A problem may occur during the course of production and the producer may have to decide on the spot whether to abort the process or proceed at greater cost. Or a new bid may be heard and have to be immediately answered. Furthermore, even if the parties are in contact with one another, asymmetric information may lead to breakdowns in renegotiation.

In any event, let us assume that successful renegotiation tends to occur and consider how it affects the welfare of contracting parties. Plainly, renegotiation often allows parties to avert Pareto-inefficient breach decisions. For example, if damages exceeding the expectation measure or specific performance were the remedy for breach, a seller might be led to perform when his production cost exceeds the value of performance to the buyer. To avoid this inefficient outcome, the seller might pay the buyer to release him from his obligation to perform. That renegotiation may result in performance if and only if it is efficient means, as we noted, that damage measures for breach are not necessary to accomplish this, and also helps to explain why contracts lack detail.

But even if renegotiation tends to occur, it may represent only a partial substitute for explicit contractual terms or for appropriate damage measures for breach. One reason (see Section 4.2.3) is that renegotiation cannot affect actions that are taken before the time of renegotiation, which influence the likelihood of nonperformance; renegotiation can only affect future decisions about breach. Another reason involves the allocation of risk-bearing. Consider, for instance, the substantial risks borne by a producer who may have to purchase a release from an obligation to perform when his production costs would be extremely high. Such risks could be mitigated by use of a clause excusing him from performance or by a damage measure such as expectation.

Additionally, the prospect of renegotiation affects the incentives of parties to invest in the contractual relationship. A party's level of reliance investment will be inefficient if renegotiation results in the extraction of part of the surplus that the party's reliance investment creates. Yet renegotiation is influenced by, among other elements, the damage measure that applies for breach, and if the damage measure is appropriately chosen, the damage measure together with renegotiation may, in principle, spur desirable reliance investment; see Section 4.2.2.

One presumes that the ability to renegotiate is usually desirable for contracting parties, because it allows them to improve their situation when difficulties arise and to write simpler contracts than otherwise. Thus, we would expect that parties will want their renegotiated contracts enforced, and the law generally does enforce renegotiated contracts. However, the ability to renegotiate can also work to the detriment of parties because they might thereby be prevented from committing themselves to particular outcomes in their initial contract. See Jolls (1997) and the literature cited therein, especially Fudenberg and Tirole (1990). Nevertheless, the law usually prevents parties from binding themselves not to renegotiate, even though that could in theory be done⁹⁰.

4.1.9. *Legal overriding of contracts*

A basic rationale for legislative or judicial overriding of contracts is the existence of externalities. Contracts that are likely to harm third parties are often not enforced, for example, agreements to commit crimes, price-fixing compacts, liability insurance policies against fines, and certain simple sales contracts (such as for machine guns)⁹¹.

⁹⁰ It is true that parties will not usually be able to bind themselves against engaging in renegotiation, for they could ordinarily renegotiate in secret. However, as Jolls (1997) observes, one of the parties will usually prefer that the original contract be enforced, so that if the courts stand ready to enforce the original contract, renegotiation cannot result in a new contract. For example, in the standard principal-agent contract, after the agent exerts effort, the principal and the agent will have an incentive to arrange for the agent to be paid a constant amount. But if this were contemplated, then after the output is realized, the agent would have an incentive to assert the original contract if his pay would be higher according to it, and the principal would assert the original contract if he could pay less under it.

⁹¹ See also E.A. Posner (1995), who suggests that such contractual limits as usury laws, which constrain consumers' ability to borrow, might be justified by a type of externality: when high-risk borrowers fail, they may become eligible for social welfare programs, imposing costs on taxpayers.

Another general rationale for nonenforcement of contracts is to prevent a loss in welfare to one or both of the parties to contracts. This concern may motivate nonenforcement when a party is incompetent, lacks relevant information, or is in an emergency situation (see Section 4.1.2). The rationale also applies in the context of contract interpretation by tribunals; as discussed in Section 4.1.5, contract interpretation may amount to overriding terms of contracts, and this may promote the welfare of contracting parties by allowing them to write simpler contracts⁹².

Additionally, contracts sometimes are not enforced because they involve the sale of things said to be inalienable, such as human organs, babies, and voting rights. In many of these cases, the inalienability justification for lack of enforcement can be recognized as involving externalities or the welfare of the contracting parties⁹³.

4.2. *Production contracts*

In this section, the literature on production contracts is discussed. The first case considered is that where symmetrically informed, risk-neutral parties enter into contracts, and the only variables of concern are the value of performance and production cost. Then the case where parties make reliance investments to raise the value of the contract during the contract period is examined. Finally, several other issues, including risk-bearing and asymmetric information, are reviewed. Throughout, when remedies for breach are discussed, one can imagine them to be chosen either by the parties or by the courts.

4.2.1. *Value of performance and production cost*

Assume that a risk-neutral buyer and a risk-neutral seller have met; the seller faces uncertain production cost c , which he will learn before he decides whether to produce; v is the certain value of performance to the buyer; and the parties are symmetrically informed. The Pareto efficient outcome is for the seller to produce if and only if $c < v$. (That is, in a complete contract, with terms for all contingencies, performance would be required if and only if $c < v$; a change in the contract price would compensate a party for agreeing to alter a term from any initially considered contract under which performance does not occur if and only if $c < v$.)

In the absence of contract enforcement, then (amplifying on Section 4.1.3) there would be too little production because the buyer would only pay the seller for actual delivery of the good and cannot guarantee the price. In particular, supposing that the seller would obtain a fraction α of the surplus from a transaction (α reflects bargaining strength), he would obtain a price of αv . (After the seller produces the good, the surplus

⁹² Also, at least in theory, nonenforcement of contracts might also be beneficial to parties where they would be led to include terms constituting wasteful signals of unobservable characteristics. See Aghion and Hermalin (1990).

⁹³ See generally Rose-Ackerman (1985) and Trebilcock (1993).

from the transaction would be v , presuming for simplicity that the custom good has no alternative value for the seller.) Thus, the seller would decide to produce only when $c < \alpha v$, rather than whenever $c < v$.

Suppose now that there is contract enforcement and that the parties are not able to renegotiate before the seller decides whether to produce (an assumption that is relaxed below). If c is verifiable by the tribunal, the parties could write a complete contract specifying performance if and only if $c < v$. The parties would want a damage measure d for breach of this contract to be sufficiently high to induce performance when $c < v$, and thus any d exceeding c would work.

If c is not verifiable, the parties are able to write an incomplete contract specifying "The seller shall deliver the good to the buyer, who will pay price p at the outset", accompanied by damages d for seller breach. Under such a contract, the seller will perform when $c < d$ and will commit breach otherwise⁹⁴. If the expectation measure is employed, that is, $d = v$, the seller will perform if and only if $c < v$, so that performance will be efficient⁹⁵. If damages d exceed v , there will be excessive performance, as there will be if there is specific performance. If d is less than v , there will be too little performance. The points of these paragraphs were, as noted, emphasized in R.A. Posner (1972) and Shavell (1980b).

If, instead, it is assumed that the buyer and the seller can renegotiate their contract after c becomes known but before the seller decides whether to produce, then, given symmetric information, it is natural to suppose that there will always be Pareto efficient performance, regardless of d .

Let us also note that if the buyer's value v is uncertain as well as the seller's production cost c , the major difference in the outcome is that, since v cannot be prescribed as damages in the contract, v must be verifiable for the expectation measure $d = v$ to be applied by the tribunal (c still need not be verifiable)⁹⁶.

⁹⁴ Because we assume that the price p is paid at the outset, the seller faces cost c if he performs and will compare it to damages of d that he would have to pay if he breaches. If the price were to be paid only at the time of performance, then the seller would perform if and only if $c - p < d$. Hence, the performance that is induced under d if the price is paid at the outset will be achieved under $d' = d - p$ if the price is paid only at performance.

⁹⁵ A related issue concerns post-breach mitigation behavior of the buyer: efficiency requires that if there is a breach, the buyer should mitigate the consequences of breach by searching for alternative suppliers and the like. Let z be mitigation expenditure of the buyer to raise his post-breach alternative value, say $w(z)$. Efficiency requires the buyer to choose z to maximize $w(z) - z$; let z^* be the optimal value of z . If y is the gross value of seller performance to the buyer, then we can define v , the net value of performance, as $v = y - (w(z^*) - z^*)$. Thus, expectation damages for breach should equal this v , not the gross value y . And if damages equal v , then the buyer will choose z^* if he is the victim of a breach, and the net value of performance will actually be v . On this issue of mitigation of the consequences of breach, see, for example, Wittman (1981).

⁹⁶ However, if c is verifiable and v is not, Pareto-efficient performance can be achieved by constructing the contract so that the *buyer* will commit breach by refusing to pay for performance when performance would be inefficient. Specifically, let the price p be paid at performance, and let damages for buyer breach be

4.2.2. Reliance investment during the contract period

Now assume that parties can make investments during the period of the contract that affect its value v or the production cost c . Such investments are, as noted, sometimes called reliance investments, since they are made in anticipation of contractual performance. We will begin with the case in which just one party invests before discussing the case where both sides invest.

Buyer makes reliance investment and seller's costs are uncertain. Suppose that one party to the contract invests, for concreteness the buyer, and that the other party faces uncertainty⁹⁷. Specifically, let r be the buyer's reliance investment (training of workers to use a contracted-for machine) and let $v(r)$ be the value of performance given r , where v is increasing in r . The buyer chooses r before the seller learns c and decides about producing. The Pareto-efficient decision of the seller is to produce when $c < v(r)$, and the efficient decision of the buyer is therefore to choose r to maximize

$$\int_0^{v(r)} (v(r) - c) g(c) dc - r,$$

where g is the density of c . Thus, the optimal r , denoted r^* , is determined by $v'(r) G(v(r)) = 1$, where G is the cumulative distribution of c . The point to note here is that the marginal return to reliance investment is only a contingent return, for the investment pays off only with probability $G(v(r))$, when $c < v(r)$ (when production turns out to be efficient).

In the absence of contract enforcement, there will be too little production, as before; it will occur only when $c < av(r)$. But now, in addition, the buyer will choose an incorrect value of r because he will only obtain a fraction $1 - \alpha$ of the value created by investment⁹⁸.

Assume next that there is contract enforcement and that the parties do not renegotiate before the seller's production decision (we relax this assumption below). This is the setting analyzed in Shavell (1980b), who first studied reliance investment. If c and r

$d = p - c$, the seller's profits. Then the buyer will breach and refuse performance whenever $v - p < -(p - c)$, or when $v < c$. (If, as is realistic, it is assumed that $p - c$ cannot be negative, then the parties can choose p high enough that it always exceeds c (assuming c is bounded), with the buyer being compensated for the high p through an up-front rebate.) The parties' ability to determine who will make the breach decision, as described here, is emphasized in Edlin (1996).

⁹⁷ We comment in note 99 below on another case of reliance investment: where the party who chooses the reliance investment is the same party who faces uncertainty, such as where the seller chooses r to lower his production cost and faces uncertainty about his production cost.

⁹⁸ Specifically, he will choose r to maximize $(1 - \alpha)v(r) G(av(r)) - r$, so the first-order condition determining r is $(1 - \alpha)v'(r) G(av(r)) + (1 - \alpha)v(r) av'(r) g(av(r)) = 1$, or $(1 - \alpha)v'(r)[G(av(r)) + av(r)g(av(r))] = 1$. Although one might expect r to be less than r^* , it is apparent from the first-order condition that there is a possibility that the r chosen would exceed r^* . The reason is that increasing r raises the probability that the buyer will obtain performance from the seller.

are verifiable by the tribunal, the parties can write a contract specifying efficient performance (when $c < v$) and also specifying r^* ; again, they would want the contract enforced by a damage measure high enough to ensure performance, and any such measure of damages would serve their purposes.

Now assume that c and r are not verifiable, that the parties write a simple contract specifying "The buyer will pay price p at the outset and the seller will deliver the good to him", and consider what occurs under different damage measures. If the expectation measure is employed, that is, $d = v(r)$, the seller will perform when $c < v(r)$, so that performance will be efficient. However, as the buyer will always receive $v(r)$ (either he obtains performance, worth $v(r)$, or damages of that amount), he will choose r to maximize $v(r) - r$. Consequently, the buyer will select an inefficiently high r ; the problem is that the buyer does not take into account that investment does not have any value when performance does not occur⁹⁹. Under a sophisticated expectation measure based on efficient investment, namely $d = v(r^*)$, however, investment as well as performance can be shown to be efficient¹⁰⁰.

Another damage measure that has been examined is known as *reliance damages*, according to which the buyer would receive a return of his initial payment p plus his reliance investment r if the seller breaches. Under this measure, if there is a breach,

⁹⁹ We observe that the problem of an inefficiently high r does not arise under the expectation measure where the seller makes the reliance investment to lower his production cost and also faces uncertainty about it. Specifically, suppose that production cost is $c(r, \theta)$, where θ is an uncertain state of nature, $c_\theta > 0$, and $c_r < 0$. In this case, the seller will choose the efficient r ; the explanation in essence is that the seller obtains the benefit of his reliance only when there is performance. The efficient r is that maximizing

$$\int_0^{\theta(v, r)} (v - c(r, \theta)) g(\theta) d\theta - r,$$

where $\theta(v, r)$ is the θ such that $c(r, \theta) = v$. Thus, r^* is determined by $-\int c_r(r, \theta) g(\theta) d\theta = 1$. Now under the expectation measure, the seller will perform when $c < v$ and pay v otherwise. Thus, the seller chooses r to maximize

$$p - \int_0^{\theta(v, r)} c(r, \theta) g(\theta) d\theta - v(1 - G(\theta(v, r))) - r,$$

and differentiation of this yields the same condition as that which determines r^* . This point is noted in Shavell (1980b).

¹⁰⁰ If $d = v(r^*)$, the seller will perform when $c < v(r^*)$, so the buyer will maximize $v(r) G(v(r^*)) + v(r^*)(1 - G(v(r^*))) - r$. Accordingly, r will be determined by $v'(r) G(v(r^*)) = 1$, and this condition is clearly satisfied at r^* . The explanation is that the buyer's choice of r affects his return only when he obtains performance. Hence, r^* will be chosen and performance will also be efficient. This point was first mentioned by Cooter (1985). (Observe that the tribunal does not need to observe r to enforce $d = v(r^*)$, as the parties can name $v(r^*)$ in the contract.) The analysis would change, however, if the buyer does not know $G(\cdot)$. See Craswell (1988).

the buyer will be placed in the position he would have enjoyed had he not invested and made the contract. It can be shown that, under the reliance measure, investment would be even more excessive than under the expectation measure, and there would be too little performance. (Note, however, that to apply the reliance measure, courts must be able to verify investment r , and that if this is so, r^* could be achieved simply by the parties naming it in their contract.) Finally, under specific performance, there is excessive performance, but r is chosen optimally given that level of performance (because performance always occurs).

Next assume that the parties do renegotiate after the reliance investment is made and before the seller decides about production, so that, assuming symmetric information, there will always be efficient performance. This version of the model of production contracts was originally studied by Rogerson (1984). Here, damage remedies may influence investment through their effect on the outcome of renegotiation. To illustrate, consider what would occur under specific performance. Under this remedy, as suggested earlier, there will be renegotiation in which the seller pays the buyer to be allowed not to perform whenever $c > v(r)$, since then performance would be inefficient. In particular, the assumption is that the seller would pay the buyer $v(r) + (1 - \alpha)(c - v(r))$ to be allowed not to perform; for $v(r)$ is needed to compensate the buyer for not receiving performance, $1 - \alpha$ is the buyer's share of the surplus from renegotiation, and $c - v(r)$ is that surplus. Anticipating this, the buyer can be shown to choose an r exceeding the efficient level¹⁰¹. The features of the results about reliance investment in the case with renegotiation are very close to those where there is no renegotiation. Indeed, they are identical under the expectation measure, essentially because there is no renegotiation under the expectation measure; thus, with $d = v(r)$, investment will be excessive because the buyer will always be compensated for his investment. Furthermore, under the sophisticated expectation measure based on efficient investment, $d = v(r^*)$, investment will be efficient. See Spier and Whinston (1995).

Both parties make reliance investments and both the value of performance and production costs are uncertain. Here let $v = v(r, \theta)$ and $c = c(s, \theta)$, where s is reliance investment of the seller and θ is the state of nature; s lowers c given θ . In this more general situation, what occurs can be understood in many respects by analogy to the case just discussed. For example, under the expectation measure, investment will tend to be excessive for both parties, but performance will be efficient.

Much recent literature, beginning with Hart and Moore (1988), has focused on this general situation, assuming that parties can renegotiate after reliance investments are

¹⁰¹ In general, the buyer will choose an r in between the excessive level he would choose under the expectation measure (determined by $v'(r) = 1$) and r^* . This can be explained as follows. If the buyer's fraction of surplus is 0, he will receive $v(r)$ whether or not there is renegotiation, so his situation will be the same as under the expectation measure. If the buyer's fraction of surplus is 1, he will clearly choose r^* . This suggests what can be shown, that if the buyer's fraction of surplus is positive and less than 1, he will choose an r exceeding r^* and less than the r chosen under the expectation measure.

made and θ is revealed, and that they will always then agree on efficient production decisions because information is symmetric. The literature in question, furthermore, usually supposes that none of the variables (costs, values of performance, reliance investments) are verifiable by the tribunal. Thus, a contract can depend only on what is recorded in it, certain subsequent communications between the parties, whether there has been performance, and, if not, who committed breach.

Of note are a number of results establishing the existence of contracts that will produce efficient outcomes, that is, both parties choosing efficient levels of reliance investment (performance will always be efficient). Aghion, Dewatripont and Rey (1994) and Chung (1991) demonstrate the efficiency result using a contract in which one party is effectively given the right to make a single take-it-or-leave-it offer to the other in renegotiation. It is evident that this party will invest efficiently, as he can extract in bargaining the full marginal return from his investment. For instance, if the buyer has the right to make an offer and is paying the seller to perform, he will pay only the minimum needed to induce the seller to do so, and will obtain any increase in value $v(r, \theta)$ due to his having chosen a higher r . Less apparent is how the other party is given an incentive to invest efficiently; that is accomplished by properly choosing the quantity of the good or the probability of delivery. (For instance, if the named quantity of the good is chosen to be higher than is likely to be efficient, the buyer will usually pay the seller to agree to lower the quantity. The amount the buyer will pay must compensate the seller for the profits he would have made at that higher contracted-for quantity. But the profits the seller would have made at that quantity will depend on his investment in lowering production costs – thereby giving the seller an incentive to invest in lowering his production costs, and an incentive that is greater the higher the contracted-for quantity.) Also, Noldeke and Schmidt (1995) establish that a simple option contract will induce efficient investments for reasons that are closely related to those just reviewed. Additionally, Edlin and Reichelstein (1996) and Hermalin and Katz (1993) adduce contracts leading to efficiency under somewhat different conditions, and Rogerson (1992) shows that efficiency can be achieved under wide circumstances, but assuming that parties can commit not to renegotiate their contracts.

Cooperative reliance investments. It has been assumed above that a reliance investment benefits directly only the party who makes it. Another possibility is that a reliance investment benefits the other party to the contract; importantly, suppose that a seller's investment raises product quality and, thus, value for the buyer. Such cooperative reliance investment is studied in Che and Chung (1999)¹⁰². As they emphasize, when cooperative investment cannot be verified by courts, then under the expectation measure, there will be too little investment (in contrast to the usual case under the expectation measure, where investment is excessive). Indeed, there will be no investment if the seller who makes a cooperative investment will not benefit directly

¹⁰² Cooperative investment is also studied in MacLeod and Malcomson (1993).

or in damages he receives in the event of breach. Moreover, there is no contract that will result in efficient cooperative investment (again in contrast to the usual case); this point is stressed in Che and Hausch (1999), who also demonstrate that contracting offers no advantage over no contracting in a wide set of circumstances¹⁰³.

4.2.3. Further considerations

Risk-bearing. We have not discussed in this section on production contracts the allocation of risk among possibly risk-averse contracting parties, about which several comments should be made. First, if all variables are verifiable by a tribunal, the presence of risk-averse parties does not affect when it is Pareto efficient to perform; it continues to be efficient to perform if and only if $c < v$. However, efficiency requires that the resulting risk be allocated appropriately; for instance, if the seller is risk averse and the buyer risk neutral, the seller would be insured against fluctuations in c by the buyer's paying him c plus a constant. In addition, the level of efficient reliance investment will generally be affected by considerations of risk-bearing.

Second, when variables of relevance are not verifiable, then damage measures and other mechanisms that may be employed to induce efficient behavior when parties are risk neutral have to be reconsidered. For instance, the expectation measure imposes risk on the party who might breach and pay these damages; if that party is risk averse, the expectation measure would become less attractive relative to lower measures of damages. Furthermore, as we earlier noted, renegotiation does not generally lead to efficient risk bearing, even though it may lead to efficient performance.

Asymmetric information. Another factor about production contracts that we have not examined is asymmetric information between the parties. When parties are asymmetrically informed, renegotiation of contracts might not be successful, so that it becomes more important that the initial contract induces efficiency. Hermalin and Katz (1993) show that efficiency can be achieved under certain types of asymmetry of information using a relatively complicated mechanism in the contract.

New entrants. We have not examined the possibility that new buyers would appear and bid for the seller's good (a similar possibility is that new sellers would appear and make offers to the buyer). In this regard, it should be noted that it is Pareto efficient for the initial contracting parties that a sale be made to a new buyer if and only if his bid exceeds the contract buyer's valuation. Moreover, the contracting parties will want to maximize the amount that they can extract from a new buyer if he purchases the good. This observation raises the possibility that the buyer and the seller may wish to set damages for seller breach at a high level in order to induce a new party to bid more (which he would have to do to make it in the seller's interest to commit breach). Such

¹⁰³ For recent attempts to provide a unified theoretical framework for the various problems concerning reliance discussed in Section 4.2.2, see Maskin and Moore (1999) and Segal and Whinston (forthcoming).

an incentive of contracting parties to set damages at high levels can, though, result in too little breach and sale to new parties; thus, at least in principle, the incentive in question is a ground for tribunals not to enforce the high damage level specified by the contracting parties. This point was first made in Diamond and Maskin (1979) and has been refined in a number of articles; see Aghion and Bolton (1987) and Chung (1992). However, Spier and Whinston (1995) observe that three-way renegotiation would seem to vitiate the advantage to the contracting parties of setting high damages. Yet they emphasize another reason (concerning induced reliance investment) that the parties will, after all, benefit from setting high damages.

Precautions and probabilistic breach. It has been supposed throughout that breach occurs when a party decides not to perform, but often breach does not occur in this way: rather a party chooses a level of precaution that affects the likelihood of performance, and a random factor then determines whether breach or performance results. For example, a shipper's care in packing dishes affects the likelihood that they will arrive unbroken, and a chance event (a jolt) determines whether they arrive broken or unbroken. In this setting, the conclusions reached about damage measures in the absence of renegotiation continue to apply: the expectation measure results in efficient precautions, the buyer's reliance investment is excessive, and so forth. The very issue of renegotiation is made moot because the precautions are chosen before breach might occur (if the dishes arrive broken, it is too late for renegotiation). See Bebchuk and Png (1999), Cooter (1985), Craswell (1988), and Kornhauser (1983).

4.3. *Other types of contract*

4.3.1. *Contracts for transfer of possession*

A different contractual context from production is where something that already exists is to be conveyed to a buyer. Examples include contracts for transfer of real estate, goods in inventory, and durable goods. Here a major uncertainty of interest concerns bids by new parties. With regard to these bids, the points just discussed concerning new entrants apply; the parties would like for there to be a sale to a new buyer when he will pay more than the contract buyer's valuation, and so forth.

It is of interest to explore why contracting parties often adopt specific performance as the remedy for breach of contracts for transfer of possession, even though damage measures are commonly employed for other types of contract. Initially, suppose that the contract buyer and the contract seller have equal access to bids from new parties. Then the buyer's always receiving the good does not result in any loss of opportunity to sell to a new party willing to bid a high amount; that is, specific performance does not suffer from any clear disadvantage relative to damage measures that would allow the seller to breach and sell to a new party. Moreover, specific performance offers the parties an advantage over damage measures. Namely, because under specific performance it will always be the buyer who will be bargaining with a new party, the good will never be sold to a new party bidding less than the buyer's valuation. In contrast, such a sale

could occur if the seller might pay damages, commit breach, and bargain with a new party (suppose that bargaining does not involve the contract buyer as well). And, after such a sale, the buyer would have to obtain the good through repurchase from the new party, but in general this will be at a higher price than the seller obtained – meaning that some of the surplus would be shared with the new party. (Although the contracting parties would be worse off, if the buyer repurchases at a higher price, society would not be worse off as the good would still be allocated to the user who places the highest value on it.) See Shavell (1984b) and Bishop (1985).

The foregoing advantage of specific performance in preventing inefficient sales to new parties is clearly reduced if the buyer does not have equal access to bids from new parties (suppose that the seller is a dealer and the buyer is not). Also, the use of specific performance might increase transaction costs, if the new party purchases after delivery of the good to the buyer.

Notice too that some of the disadvantages of specific performance in the production context are less significant in the present context of transfer of possession. In production contracts, specific performance imposes a possibly large risk of loss on sellers whose production costs might be very high; here, specific performance only reallocates a beneficial risk (of a sale at a high price) from seller to buyer. In addition, enforcement of specific performance in the context of contracts for transfer of possession is often easier than in the production context, where enforcement might involve policing the quality of production or services.

4.3.2. *Donative contracts*

An important category of contractual arrangement is donative, concerning gifts. Assuming that the motivation for gifts is altruism¹⁰⁴, a basic question is why a donor would want to defer his gift rather than make it immediately (in which case no contract would be required). The answers include the possibilities that the donor may face liquidity constraints and that he may wish to wait for resolution of uncertainties concerning, among other factors, his own needs and future income and the donee's needs, future income, and character traits.

Given that a donor does desire to defer making a gift, would he want to make a contract that would in some way bind him? The disadvantage of so doing is that it may not be feasible for him to limit as he wishes the conditions under which he makes the gift (due to the costs of specifying these conditions and to the problems that courts would have in verifying them). The donor's principal benefit from entering into a contract is that it may induce the donee to engage in reliance activities that will increase the value of the gift to the donee (a high-school student might study more if he anticipates a gift that will finance his college education). Such reliance activities

¹⁰⁴ Other motives for gift giving exist (such as obtaining utility from expressions of gratitude from donees); some have similar implications to those of altruism.

will in turn inure to the benefit of the donor because of his altruism. However, if the donee knows about the altruism of the donor, a contract may not be necessary to induce donee reliance activity; if so, a contract would be disadvantageous for the donor. On these issues, see Goetz and Scott (1980) and Shavell (1991a); and see also R.A. Posner (1977b) and E.A. Posner (1997).

4.3.3. *Additional types of contract*

In this section, mention has not been made of many additional types of contract, including principal–agent contracts, even though they have been studied, often intensively, in the economics literature. The omission of such contracts from consideration here is explained in part by convention (by what is and is not considered to be a law and economics topic) and in part by the relative inattention that has been paid to contract enforcement.

5. Litigation

In this section, we consider civil litigation, that is, the bringing of lawsuits by private actors to enforce their rights in the areas of civil law that we have just discussed. Until now, we have largely assumed that the operation of the legal system is frictionless, in the sense that the bringing and adjudication of lawsuits is without cost. We now analyze the implications of the expense involved in the operation of the legal system.

We begin with what may be called the basic theory of litigation: the choice of a party who has suffered a loss whether to sue; if suit is brought, the choice of the litigants whether to settle with each other or instead go to trial; and the choice of litigants, before or during trial, of how much to expend on litigation. Then we discuss various extensions to the basic theory of litigation, including nuisance suits, the shifting of legal fees, lawyers and agency problems in litigation, and legal discovery. We subsequently consider the provision of legal advice, the appeals process, alternative dispute resolution, and the formulation of legal rules.

5.1. *Suit*

5.1.1. *Private incentive to sue*

As a general matter, the plaintiff will sue when the cost of suit c_P is less than the expected benefits from suit. The expected benefits from suit incorporate potential settlements and trial outcomes, but in this section we usually assume for simplicity that if suit is brought, the plaintiff obtains as a judgment a certain amount h equal to harm suffered. Thus the plaintiff will sue if and only if his litigation cost, c_P , is less than h . (Obviously, if there is only a probability p of winning this amount, the plaintiff, if risk neutral, would sue if and only if $c_P < ph$; and if the plaintiff is risk

averse, he would be less likely to sue.) The effect on the private incentive to sue of many variations in the legal environment is straightforward to identify, as we will note below.

5.1.2. Socially optimal suit versus the private incentive to sue

The private incentive to bring suit is fundamentally misaligned with the socially optimal incentive to do so, given the social costs and social benefits of suit. The deviation between the privately motivated and socially appropriate level of suit could be in either direction. The general reasons for these conclusions may be understood as follows.

On one hand, there is a divergence between social and private costs that can lead to socially excessive suit. Specifically, when a plaintiff contemplates bringing suit, he bears only his own costs; he does not take into account the defendant's costs or the state's costs that his suit will engender.

On the other hand, there is a difference between the social and private benefits of suit that can either lead to a socially inadequate level of suit or reinforce the cost-related tendency toward excessive suit. Namely, the plaintiff does not recognize as a benefit to himself the social benefit of suit, its deterrent effect on the behavior of injurers. But the plaintiff does consider his private benefit, the gain he would obtain from prevailing. This private gain is not a social benefit but instead is a transfer from the defendant; it could be either larger or smaller than the social benefit. The contrast between the socially optimal and private incentive to sue was initially examined in Shavell (1982b)¹⁰⁵.

Let us consider the foregoing in more detail. Suppose that liability is strict. As stated, victims will sue if and only if

$$c_P < h.$$

Let x be the precaution expenditures that injurers will be induced to make if there is suit, q the probability of harm h if suit is not brought, and q' the probability of harm if suit is brought. (Thus, q' will be less than q if x is positive.) Suit will be socially worthwhile if and only if

$$q'(c_P + c_D + c_S) < (q - q')h - x,$$

where c_D is the defendant's litigation cost and c_S is the state's cost. In other words, suit is socially worthwhile if the expected litigation costs are less than the net deterrence benefits of suit. It is clear that the two foregoing conditions, for victims to sue and for suit to be socially optimal, are very different. Whether victims will sue does not

¹⁰⁵ See also subsequent examination of the issue in Menell (1983), Kaplow (1986b), Rose-Ackerman and Geistfeld (1987), and Shavell (1997).

depend on the costs c_D and c_S . Moreover, the private benefit of suit is h , the amount of harm (conditional on harm occurring), because this is what the victim will receive as a damages award; in contrast, the social benefit depends on the harm weighted by the reduction in the accident probability, $q - q'$, net of the cost of precautions x . It is evident, therefore, that victims might sue when suit is not socially optimal, and that victims might not sue even when suit would be socially optimal.

To illustrate the possibility of socially excessive suit, suppose that the losses a victim would suffer in an accident are \$10000; that a victim's cost of bringing suit will be \$3000 and an injurer's cost of defending \$2000; that the probability of accidents is 10%, and that there is no precaution that injurers can take to lower the accident risk. Victims will then bring suits whenever accidents occur, for suing will cost a victim only \$3000 and yield him \$10000. From the social perspective, this outcome is undesirable. Suit creates no beneficial deterrent, as injurers cannot do anything to lower risk. Yet suit does generate legal costs: expected legal costs are $10\% \times (\$3000 + \$2000) = \$500$. The bringing of suits is not socially desirable in this example because there are no incentives toward safety created by the suits. Yet this fact is of no moment to victims; nor are other parties' litigation costs. Victims bring suits for their private gain of \$10000.

To illustrate the opposite possibility, suppose that the losses victims suffer in accidents are now \$1000, and an expenditure of only \$10 by injurers will reduce the probability of accidents from 10% to 1%. The costs of suit and of defending against suits are as in the previous example. In this case victims will not bring suits, as doing so will cost a victim \$3000 but yield him only \$1000. Hence injurers will have no reason to take care to reduce risk, and total expected social costs will be $10\% \times \$1000 = \100 . It would be desirable for victims to bring suit, however. If they did, injurers would be led to spend \$10 to lower risk to 1%, and total expected social costs would be only $\$10 + 1\% \times (\$1000 + \$5000) = \70 . Here the bringing of suits is socially worthwhile because of the significant reduction in the risk of accident losses that would result. (And observe that this is true even though the total legal costs of \$5000 exceed the victim's losses of \$1000.) But victims do not take the deterrence-related benefits of suit into account. Each victim looks only to his own gain from suit, which is small.

Under the negligence rule, the conclusions are qualitatively similar to those under strict liability, but the problem of excessive suit is less likely. To explain, assume initially that a victim would not sue a non-negligent injurer, because he would know that he would lose. Then it is socially desirable for victims always to be willing to bring suit against negligent injurers, however great the legal costs of suit would be. For if victims always stand ready to sue negligent injurers, injurers will be induced to act non-negligently. Thus, there will never actually be any suits for negligence – given the assumption that no one sues a non-negligent injurer – and thus no legal costs will be borne; deterrence of negligence will be achieved without legal cost. Although it is socially desirable for victims always to be willing to sue negligent injurers, victims of course will not do so if the cost of making claims exceeds their losses. Consequently, there might be a problem of too few suits.

Now assume, more realistically, that victims might sometimes bring suit against non-negligent injurers (or that injurers cannot perfectly control their behavior and sometimes act negligently). Then legal costs will in fact be incurred under the negligence rule. The situation will therefore be qualitatively similar to that under strict liability; there may be too many suits as well as too few, although one might suppose the problem of too many suits to be less severe than under strict liability.

It should be clear from our discussion that the point that the private and social incentives to bring suit may diverge is robust. On one hand, it will always be the case that the private cost of use of the system will be less than the social. And, on the other hand, the private benefits from suit will be what the plaintiff will win from suit, usually money, whereas the social benefits from suit will ordinarily be different: they will always include deterrence benefits and may also include compensation of victims (if insurance is unavailable) and the setting of precedent. These benefits litigants either will not take into account or will tend to weigh differently from their social importance¹⁰⁶.

5.1.3. *Implications of the social and private divergence*

The main implication of the social and private divergence is that state intervention may be desirable, either to correct a problem of excessive suit – notably, by taxing suit or barring it in some domain – or a problem of inadequate suit – by subsidizing suit in some way¹⁰⁷. For the state to determine optimal policy, however, requires it to determine the effects of suit on injurer behavior and weigh them against the social costs of suit. It is thus not correct for the state to base policy on some simple, even though superficially appealing, criterion, notably, whether the plaintiff's expected gains from suit would have exceeded the aggregate litigation costs.

It should also be emphasized that the importance of the private–social divergence in incentives to sue may be substantial. This is suggested by the fact that the costs of use of the legal system are high; indeed, legal costs may on average actually equal the amounts received by those who sue¹⁰⁸. Hence, the incentives created by the legal system must be significant to justify its use. However, regardless of whether the legal system creates valuable incentives, the private motive to bring suit may be great,

¹⁰⁶ Whereas this section was concerned with the implications of litigation costs for the frequency of suit, another litigation-cost related issue has to do with the level of precautions taken by injurers. The optimal level of precautions should reflect not only the direct harm that would be caused by an accident, but also litigation costs. To induce this higher level of precautions, injurers who are sued should pay not only the harm, but also (perhaps as a penalty) litigation costs borne by the plaintiff and by the state. On this issue, see Hylton (1990), Polinsky and Rubinfeld (1988a), and Shavell (1997, 1999).

¹⁰⁷ An additional response to the problem of inadequate suit that is sometimes possible is the use of class actions. See Dam (1975). Other approaches to encourage suits involve pro-plaintiff fee-shifting (see Section 5.4.2), damage multipliers [see Kaplow (1993)], and “decoupling” [see Polinsky and Che (1991) and Shavell (1997, 1999)].

¹⁰⁸ See Section 2.3.1.

giving rise to a reason for social intervention. Conversely, it may be important in some domains to create deterrence because this would have a significant effect on behavior, even though the money benefits of suit are too small for most victims to bring suit. This would justify the state's supporting litigation.

5.2. *Settlement versus trial*

Assuming that suit has been brought, we now take up the question whether parties will reach a settlement or go to trial¹⁰⁹. A settlement is a legally enforceable agreement, usually involving a payment from the defendant to the plaintiff, in which the plaintiff agrees not to pursue his claim further. If the parties do not reach a settlement, we assume that they go to trial, that is, that some tribunal determines the outcome of their case. In fact, the vast majority of cases settle¹¹⁰. We discuss here two different models describing whether settlement occurs and then consider the socially optimal versus the private decision whether to settle.

5.2.1. *Exogenous beliefs model*

One model of settlement versus trial presumes that parties have each somehow come to a belief about the probability of the trial outcome; let p_P represent the probability of the plaintiff prevailing in his opinion, and let p_D be that same probability in the defendant's opinion. Let w be the amount that would be won (for simplicity assume that they agree about w). Assume also that the parties are risk neutral.

The plaintiff's expected gain from trial, net of litigation costs, is $p_P w - c_P$. This is the minimum amount he would accept as a settlement, rather than go to trial. The defendant's expected loss from trial, including his litigation costs, is $p_D w + c_D$; this is the maximum amount he would pay in settlement rather than go to trial. Hence, a settlement is possible if and only if $p_P w - c_P \leq p_D w + c_D$, in which case the settlement amount will be in the settlement range $[p_P w - c_P, p_D w + c_D]$. Note that if the parties agree on the probability p , the settlement range will be positive and $c_P + c_D$ in length. A settlement range does not exist, and trial will occur, when $p_P w - p_D w > c_P + c_D$. This means that the expected award in the plaintiff's opinion exceeds the expected award in the defendant's opinion by more than the sum of litigation costs. Thus, trial will tend to occur when the plaintiff is sufficiently more optimistic about winning than the defendant believes he should be¹¹¹.

¹⁰⁹ Cooter and Rubinfeld (1989), Daughety (2000), and Hay and Spier (1998) survey this general topic; Farmer and Pecorino (1996) review the asymmetric information literature on settlement versus trial.

¹¹⁰ In fiscal year 1992, over 96% of civil cases in state courts did not go to trial; see Ostrom and Kauder (1996). In fiscal year 1995, almost 97% of federal civil cases were resolved without trial; see Administrative Office of the United States Courts (1995). These figures, however, overstate the settlement rate because many of the cases not tried were dismissed by a court rather than being settled. On the other hand, many disputes are settled before any complaint is filed.

¹¹¹ Loewenstein et al. (1993) and Mnookin (1993) suggest that litigant overoptimism is plausible.

Risk aversion of the parties will generally increase the size of the settlement range and thus, one presumes, makes settlement more likely. If the plaintiff is risk averse, he will be willing to settle for less than $p_P w - c_P$; and if the defendant is risk averse, he will be willing to pay more than $p_D w + c_D$.

The model under discussion originated with Friedman (1969), Gould (1973), Landes (1971), and R.A. Posner (1973), and was further elaborated in Shavell (1982c). It has the virtue of clarifying several basic intuitions: that settlement is fostered by litigation cost savings and by risk aversion, and that trial might result when plaintiffs expect to gain more than defendants expect to lose. The model also helps to explain the striking predominance of settlement in actuality. First, lawyers, who are experts on the law, are typically advising both litigants, and much information is acquired and comes to be shared by the opposing sides; we should thus expect beliefs of the two sides to be similar. Second, the costs of trial tend to be substantial. These observations suggest that a settlement range typically exists and thus that settlement would be likely to occur¹¹².

The model has the additional virtue of being simple and easy to manipulate because it focuses on the calculation of the settlement range¹¹³. The model is unsatisfying, however, in two respects. It does not explain the origin of parties' beliefs. And it does not include a description of rational bargaining between the parties; thus, it does not explain whether there will be a settlement when there is a positive settlement range or the amount of any settlement within the range.

5.2.2. *Asymmetric information model*

A second type of model of settlement versus litigation presumes that there is asymmetry of information between litigants, and includes an explicit account of bargaining. The simplest of such models is that of Bebchuk (1984), in which there is one-sided asymmetry of information and bargaining consists of a single take-it-or-leave-it settlement offer made by the party without private information¹¹⁴. Suppose,

¹¹² The observations also raise interesting questions about the timing of settlement – will it occur early or late? (In fact, many cases settle early, but many also settle late, on the eve of trial or even during trial.) A reason for settlement to be delayed is that at the outset of settlement negotiations, information may be disparate; but, as noted, over time, as information is acquired and shared, the parties' beliefs tend to converge. A reason for settlement to occur early, however, is that this maximizes the parties' savings in litigation costs. To express the point differently, as time passes, more litigation costs are sunk, meaning that the savings from settlement are lowered, tending to decrease the chances of settlement. For analysis of the timing of settlement (in a model of asymmetric information), see Spier (1992a).

¹¹³ For example, the larger the possible judgment amount w , the greater the chance of trial, for a larger judgment magnifies the effect of differences of opinion in the likelihood of trial outcomes ($p_P w - p_D w$, when positive, is increasing in w). However, larger judgments tend to reduce the likelihood of trial if litigants are risk averse.

¹¹⁴ Asymmetric information models of trial versus settlement have been refined and extended in various ways. See, for example, Daughety and Reinganum (1994) (in which asymmetry of information is two-

for example, that the defendant has private information about the probability p that the plaintiff will win at trial (perhaps the defendant possesses private information bearing on whether he will be found negligent)¹¹⁵. The plaintiff makes a settlement offer x , knowing that low- p defendants will reject his offer and high- p defendants will accept; specifically, if $pw + c_D < x$, the defendant will reject and the plaintiff will therefore obtain only $pw - c_P$, but if $pw + c_D \geq x$, the defendant will accept and pay x . The plaintiff, who knows the probability distribution over p , chooses x to maximize his expected payoff from settlement or trial¹¹⁶. The higher his offer x , the more he will obtain if his offer is accepted, but the greater is the likelihood of rejection and thus of his bearing trial costs. At the optimal offer for the plaintiff, there will generally be a positive probability of trial and also of settlement. Furthermore, it can be shown that the higher are litigation costs, the more likely is settlement, and that risk aversion increases the likelihood of settlement.

This model, note, is roughly consistent with the previous one of Section 5.2.1 in the sense that trial occurs due to disparate beliefs (arising out of the asymmetry of information). In particular, the plaintiff's opinion of the probability of winning is the mean probability $E(p)$ over the distribution of defendants, and trial will occur if the defendant's p is sufficiently low in the distribution. In addition, the comparative statics of the present model are similar to that of the previous one (for instance, as just noted, higher litigation costs make settlement more likely).

The primary virtues of such asymmetric information models are twofold. First, they include an explicit account of bargaining and thus of the probability of settlement and the magnitude of the settlement offer. (But the ability to predict the probability of settlement and the magnitude of the settlement offer is to some extent specious. Under the bargaining models studied, essentially arbitrary modeling choices are made over such matters as who makes the offer, the informed or the uninformed party; these choices substantially influence the probability of settlement and the settlement offers¹¹⁷.) Second, the models explain differences of opinion that give rise to trial in terms of differences in possession of information. (However, the models do not explain

sided), Hay (1995) (in which unobservable case preparation contributes to asymmetry of information), Reinganum and Wilde (1986) (in which the informed party makes the offer, and the uninformed party makes an inference from it), Schweizer (1989) (in which asymmetry of information is two-sided), and Spier (1992a) (in which there are multiple rounds of bargaining and, as discussed in note 112, the focus is on the timing of settlement). For a useful survey of asymmetric information models of litigation, see Farmer and Pecorino (1996), and for a general survey of asymmetric information models of bargaining, see Kennan and Wilson (1993). For empirical investigations of litigation that emphasize asymmetric information, see Farber and White (1991), Osborne (1999), and Sieg (2000).

¹¹⁵ Asymmetric information could also concern the magnitude of the judgment or factors independent of the trial itself, such as parties' degree of risk aversion, their short-run need for funds, their tastes for litigation, and, as mentioned in the previous note, unobservable aspects of case preparation.

¹¹⁶ Specifically, the plaintiff's expected payoff as a function of x is $\int_0^z (pw - c_P) f(p) dp + (1 - F(z))x$, where $z = (x - c_D)/w$, f is the density of p , and F is the cumulative distribution of p .

¹¹⁷ For an attempt to address this problem, see Daughety and Reinganum (1993).

why there should be such differences in information, especially given the incentives for sharing of information and the possibility of its forced disclosure through legal discovery; we discuss these issues below in Sections 5.4.7 and 5.4.8.)

5.2.3. *Socially optimal versus privately determined settlement*

The private and the social incentive to settle may diverge for reasons related to those explaining the difference between the private and the social incentive to sue (see Section 5.1.2)¹¹⁸. First, because the parties involved in litigation do not bear all the costs of a trial – the salaries of judges and ancillary personnel, the forgone value of juror time, implicit rent on court buildings – the parties save less by settling than society does, which suggests that the private incentive to settle is socially inadequate. Second, when there is asymmetric information, parties will fail to settle – and thus litigation costs will be incurred – when their demands turn out to have been too aggressive. But their desire to obtain from each other a greater share of their litigation cost savings does not itself translate into any social benefit. Third, the prospect of settlement may reduce deterrence because defendants gain from settlement. This need not, however, be socially undesirable because settlement lowers the real total social cost of harmful acts, making less deterrence appropriate¹¹⁹. Also, the division of surplus in settlement may affect deterrence. Fourth, the prospect of settlement may increase deterrence because it lowers plaintiffs' expected litigation costs and thus increases the chance of suit. These latter two factors are not, of course, taken into account by the parties to settlement negotiations. Finally, by averting trial, settlement may have other effects on social welfare. For example, trials may reveal socially valuable information (such as about product hazards that consumers could guard against) or lead to new precedents. These are also factors that parties may ignore or treat inappropriately (a firm might have a socially perverse incentive to avoid trial to conceal information about product hazards).

The state can act to correct a divergence between private and social incentives to settle. A point that should be stressed in considering optimal social policy is that if settlements were to reduce deterrence undesirably, this does not imply that trial should be fostered; deterrence could be enhanced by raising damages to induce settlements for greater amounts or by imposing a tax on defendants (regardless of whether they settle). Trial is desirable only when there is no less costly way to raise social welfare, and a conjecture is that the usual social problem is that there are too many trials, not too few¹²⁰.

¹¹⁸ The normative question concerning the social versus the private value of settlement has received little attention relative to the positive question of when parties will settle. On the normative question, see Polinsky and Rubinfeld (1988b), Shavell (1997, 1999), and Spier (1997).

¹¹⁹ Settlement reduces ex post social costs by the sum of the plaintiff's, defendant's, and court's costs, but deterrence is reduced in a way that reflects only the fraction of the savings in the plaintiff's and defendant's costs that is captured by the defendant in settlement bargaining.

¹²⁰ In fact, courts attempt to promote settlement in a variety of ways.

5.3. *Litigation expenditures*

5.3.1. *Private incentives to spend on litigation*

Here we focus on litigant expenditures given that suit has been brought. (We should note that litigation expenditures are made prior to trial as well as during trial; indeed, most are incurred in cases that settle.) Suppose that each litigant's expenditures are made noncooperatively, as in Braeutigam, Owen and Panzar (1984), Katz (1987, 1988), and R.A. Posner (1973). Under this assumption, a plaintiff will make litigation expenditures as long as this raises his expected return from settlement or trial (net of litigation costs), and a defendant will make such expenditures as long as this lowers his expected total outlays. The effects of each litigant's expenditures will generally depend on what the other does; indeed, the two will often be spending to rebut one another¹²¹.

5.3.2. *Social versus private incentives to make litigation expenditures*

There are several sources of divergence between social and private incentives to spend during litigation. First, as just noted, the litigants may well be spending in ways that offset each other. To the extent that their expenditures do not alter trial or settlement outcomes, the expenditures constitute a social waste. Second, even if expenditures are not offsetting, they may mislead the tribunal rather than enhance the accuracy of outcomes. Such expenditures have negative social value.

Third, even if expenditures do improve the accuracy of outcomes, they may not be socially optimal in magnitude. By analogy to what we stressed in Section 5.1.2, the parties decide on their expenditures based on how they influence the litigation outcome, without regard to their influence (if any) on incentives. This could lead to expenditures that are too great or too small, relative to what is socially correct¹²².

An important instance of the possibility that expenditures could be socially excessive concerns the assessment of damages. See Kaplow and Shavell (1996b). Suppose that the presently estimated harm deviates from the truth by \$100. Then one of the litigants will be willing to spend up to \$100 to prove the correct amount (it will be the defendant if the estimate exceeds the correct level, and the plaintiff if the estimate falls below the correct level). It can be shown that the social value of the more accurate estimate tends, however, to be lower than \$100, because the social value of accuracy is based on its effects on incentives. Indeed, there will sometimes be no beneficial incentive effect from more accurate assessment of harm, such as when errors (in the absence of

¹²¹ One may contrast a system in which a single authority, perhaps the tribunal, makes decisions about litigation expenditures. This is done to an extent in many European countries for criminal proceedings. In the United States, federal trial court judges occasionally use special masters or court-appointed experts to perform similar functions.

¹²² See generally Kaplow (1994a).

additional expenditures) are unbiased and not predictable *ex ante* by potential injurers. In particular, potential injurers, at the time they choose their precautions, will often know only a probability distribution of possible harm, so litigation expenditures *ex post* that provide a precise assessment of a particular victim's actual harm would not affect incentives¹²³.

Expenditures on determining whether a party is liable (as opposed to the magnitude of damages) could be socially excessive or inadequate¹²⁴. To illustrate the latter possibility, suppose that the cost of establishing that a defendant was negligent exceeds the amount of harm suffered. Plaintiffs will not have an incentive to make the necessary expenditure, with the result that negligence might not be discouraged. But if the deterrent effect of liability were significant, that result would be undesirable. (Suppose that deterrence would eliminate most negligently caused harm, so that *ex post* litigation costs would not often have to be incurred; see Section 5.1.2.)

Because private and social incentives to spend on litigation may diverge, it may be beneficial for expenditures to be either controlled or encouraged. In practice, courts often act to restrict the legal effort that parties can undertake, for example, by limiting the extent of discovery and the number of testifying experts.

5.4. *Extensions of the basic theory*

We consider here various extensions of the basic theory discussed above; for the most part, these extensions are concerned with the description of litigation rather than with its normative analysis¹²⁵.

5.4.1. *Nuisance suits*

A nuisance suit is often defined as a suit that the plaintiff brings even though he would not actually pursue his case to trial, because the expected award he would obtain is less than the trial cost; in this sense a nuisance suit is a negative expected value suit. We should first point out that we cannot infer that nuisance suits should not be brought: as we stressed in Section 5.1.2, it is possible that the social deterrence benefits of a type of suit make it desirable to bring even though litigation costs exceed the expected

¹²³ See Section 2.4.4. If compensation of risk-averse victims were important due to the unavailability of insurance (see Section 2.2), more accurate compensation will have social value, although parties' incentives to make litigation expenditures would still tend to be excessive because the insurance benefit of avoiding a \$1 error in compensation is less than the maximum of \$1 that a party would be willing to expend to correct the error. See Kaplow (1994b).

¹²⁴ The determination of liability in the context of law enforcement (see Section 6) is analyzed in Kaplow and Shavell (1994a).

¹²⁵ For empirical studies that bear on the theory of litigation, see, for example, Hughes and Snyder (1995), Ramseyer and Nakazato (1989), and Viscusi (1986a, 1988).

judgment¹²⁶. (Nor, as we emphasized, can it be assumed that non-nuisance suits – positive expected value suits – ought to be brought.)

A major question about nuisance suits, and the one to which primary attention has been given, is why they are brought in view of their negative expected value. One important explanation concerns asymmetric information: that plaintiffs who are not willing to go to trial are not identifiable to defendants and ride on the coattails of plaintiffs who would be willing to go to trial. As a consequence, the plaintiffs who are unwilling to go to trial are able to settle for a positive amount with defendants; see Bebchuk (1988) and Katz (1990a). Another possibility, not premised on asymmetric information, is that a plaintiff can initiate a suit at low cost and, although he would lose if the defendant undertook substantial litigation effort, he would prevail if the defendant did not. In this case, the defendant might prefer to settle to avoid paying defense costs; see Rosenberg and Shavell (1985). An additional reason concerns the point that, as plaintiffs spend continuously on litigation, their willingness to go to trial increases (because the amount that they would then save by not going forward diminishes); see Bebchuk (1996)¹²⁷.

5.4.2. *Shifting of legal fees*

Thus far, we have assumed that parties bear their own legal costs, a regime referred to as the *American rule*. By contrast, under the *English rule*, the loser pays the legal costs of both sides. Fee-shifting may also be one-way, favoring the plaintiff (that is, shifted only to the defendant, if the plaintiff wins) or favoring the defendant (shifted only to the plaintiff, if the defendant wins)¹²⁸. Fee-shifting has clear implications for the incentive to sue; for example, under the English rule, suit is encouraged, relative to a regime of no fee-shifting, if the plaintiff's probability of winning is sufficiently high, because then his expected costs of trial fall¹²⁹.

Fee-shifting may increase the chance of trial given that suit has been brought, essentially because it accentuates differences in litigant estimates of the expected gains and losses from trial; see R.A. Posner (1977a) and Shavell (1982c). Under the English rule (the effects under one-way fee-shifting are similar), if the plaintiff and the defendant are each optimistic about winning, then each will be optimistic about

¹²⁶ This is obviously not to deny that some types of nuisance suits are undesirable. For example, where plaintiffs would not prevail because their cases are fictitious, their bringing of suits would tend to distort incentives as well as waste resources on litigation.

¹²⁷ There is also a literature on how nuisance suits (or, relatedly, suits with a low probability of success) might be discouraged. See Bebchuk and Chang (1996), Katz (1990a), and Polinsky and Rubinfeld (1993, 1996).

¹²⁸ For a description of the use of fee-shifting, see Derfner and Wolf (1995). Fee-shifting favoring only plaintiffs is used to stimulate suits where the private incentive is thought to be inadequate, whereas fee-shifting favoring only defendants is usually proposed as a means to discourage frivolous litigation.

¹²⁹ Fee-shifting may also affect injurers' incentives to reduce risk and thus the number of harmful outcomes that could lead to suits. See van Wijck and van Velthoven (2000).

passing on his legal expenses to the other, which tends to reduce the settlement range and increase the chances of trial¹³⁰. However, fee-shifting tends to raise the amounts the parties will spend at trial, as a party's expenditure will only be a cost to him with a probability rather than with certainty¹³¹. This increase in anticipated litigation expenditures attenuates the rise in the chance of trial. Also, fee-shifting makes trial riskier, so that if parties are risk averse, it may reduce the chance of trial. Because of the variousness of the effects of fee-shifting (and because of the divergences between private and social incentives to sue, to settle, and to spend on litigation), its influence on social welfare is generally ambiguous, as is emphasized, for example, in Gravelle (1993).

A variant of simple fee-shifting is an offer-of-settlement scheme, according to which fees are shifted only if a settlement proposal is rejected and the amount actually awarded differs in a specified way from the rejected proposal. For instance, if a defendant rejects a plaintiff's offer and the actual trial award exceeds that offer, fees might be shifted to the defendant. The effects of such schemes on settlement are complex and not readily summarized. See Bebchuk and Chang (1999), Miller (1986), and Spier (1994a).

5.4.3. *Additional elements of trial outcomes*

We have assumed that the only outcome of a trial is a judgment paid by the defendant and received by the plaintiff, but there are other possibilities. First, a trial outcome may have implications for a litigant beyond the immediate judgment. For example, a firm may believe that a loss at trial would invite a string of future lawsuits; thus, a loss would be more costly for it than the judgment¹³². This would tend to make settlement more likely, as it would raise the amount the defendant firm would be willing to pay in settlement. Second, a litigant may care whether a trial is held *per se*: a plaintiff might, say, wish the defendant to be exposed to public scrutiny. This would make trial more likely. Or a party might want to avoid a trial because it would result in the airing of embarrassing facts or the disclosure of valuable business information, which would tend to make trial less likely. Third, in cases such as child-custody disputes, the combination of indivisibilities and wealth constraints may make settlement less likely¹³³.

¹³⁰ The same conclusion holds, for closely related reasons, in the asymmetric information model of Bebchuk (1984).

¹³¹ See Braeutigam, Owen and Panzar (1984), Hause (1989), and Katz (1987). In Katz's simulation, the English rule increases costs by 125 percent.

¹³² See, for example, Che and Yi (1993). A party's willingness to settle and the amount of settlement may also have reputational effects. See Miceli (1993).

¹³³ See Shavell (1993b). For instance, suppose that for each parent in a custody dispute, the value of custody is equivalent to \$1 000 000, each parent believes custody would be awarded with probability 50%, and the cost of trial for each is \$10 000. Then to induce either parent to settle and give up the opportunity

5.4.4. *Statistical inference from cases that go to trial*

A question of interest is whether cases that go to trial are representative of the underlying population of cases, and notably, whether the likelihood of plaintiff victory at trial or the amounts won are typical of the cases that settled. This question is important because, often, the most readily available data is on cases that go to trial, whereas the great majority of cases settle. As Priest and Klein (1984) first emphasized, the cases that go to trial may be quite different from settled cases. For example, if in 99% of cases defendants would be found liable for a certain amount, but in 1% of cases defendants would prevail, then, if plaintiffs cannot distinguish the two groups, plaintiffs will likely insist on a settlement amount that the former defendants would pay and the latter would reject. Hence, defendants would win all cases that go to trial, which would be wholly unrepresentative of the cases that settled. In general, cases that go to trial are not representative of the underlying population of cases, and the proper manner of making inferences from trial outcomes is complex¹³⁴.

5.4.5. *Lawyers as agents of litigants*

Because clients and their lawyers are in a principal and agent relationship, the general problems of principals and agents are relevant for clients and lawyers. Consequently, to the degree that clients cannot observe lawyers' effort levels and lack legal expertise, a fee arrangement linked to lawyers' performance might have joint value to them, but it would impose risk on lawyers (although it would simultaneously reduce clients' risk, and many clients – particularly individuals or small entities – may be risk averse)¹³⁵.

In fact, lawyers often are compensated at an hourly rate for time spent, without regard to legal outcomes. The only important explicit exception is that plaintiffs' lawyers in tort actions frequently are paid a fraction of the amount they obtain for their clients under a so-called contingent fee agreement¹³⁶. In addition, lawyers are

of custody, an offer of at least \$490,000 would have to be made, yet neither parent may have assets nearly equal to that amount. Thus, despite the fact that the parents agree about the likelihood of trial outcomes and could save litigation costs by settling, they would go to trial.

¹³⁴ Priest and Klein (1984) suggested that cases that go to trial would be won by plaintiffs approximately 50% of the time, regardless of the underlying population of cases. This somewhat surprising conclusion of theirs is correct given their assumptions; but it is not borne out in fact, and does not hold under general assumptions about the population of cases and bargaining over settlement and trial. See Eisenberg (1990), Eisenberg and Farber (1997), Hylton (1993), Shavell (1996), Waldfogel (1995b), and Wittman (1985).

¹³⁵ Another problem in the agency relationship is that, at the time of contracting, the client may not know the lawyer's quality, and there may also be asymmetric information regarding the strength of the client's case. For a discussion of how these problems may affect fee arrangements, see Dana and Spier (1993), Emons (2000), and Rubinfeld and Scotchmer (1993).

¹³⁶ See generally Rubinfeld and Scotchmer (1998) on contingent fees. We note that payment arrangements that are contingent upon outcomes may be common because lawyers who nominally charge hourly rates may submit higher bills when successful and may trim their bills when they lose.

implicitly rewarded on the basis of performance in the sense that they (and their firms) acquire reputations, so that their future business depends on performance. Lawyers' conduct is also controlled to some extent by the threat of suit by clients for malpractice, by court-mandated penalties, and by bar association discipline. See Wilkins (1992).

Principal-agent problems that are specific to the legal context arise in the decisions to sue and to settle versus go to trial. See Miller (1987). For example, when lawyers are paid on a contingent fee basis, they might have perverse incentives to favor not bringing suits or to settle, because their own gain would be only a fraction of the total gain from winning. See Danzon (1983) and Hay (1996)¹³⁷. When lawyers are paid on an hourly basis, it is often said that they have an excessive incentive to sue and to reject settlement offers in favor of trial. (This claim, however, assumes that their hourly rate exceeds their opportunity costs; if, for example, additional, more profitable work comes into the office after the hourly rate is set, then hourly-compensated lawyers may have an excessive incentive to settle.)

5.4.6. *Insurers as agents of litigants*

Insurers often play a role in litigation. In accident suits, for example, plaintiffs may own medical or disability insurance policies with clauses giving their insurers the right to bring suit and conduct litigation, and defendants frequently hold liability insurance policies that give insurers a role in litigation. Conflicts may arise between litigants and their insurers as their agents in litigation when the coverage ceiling is less than the amount at stake in litigation. See Meurer (1992) and Sykes (1994). To illustrate, suppose that there is a 20% chance that trial would result in a finding of liability and that losses are \$500 000; assume also that the defendant's liability coverage ceiling is \$150 000. The liability insurer would prefer to reject a settlement offer of \$75 000, even though the offer falls below the expected judgment of \$100 000. (If the settlement offer is accepted, the insurer pays \$75 000 for sure, whereas if there is a trial the insurer makes a payment of \$150 000 only 20% of the time, which has an expected cost of \$30 000.) By similar reasoning, a plaintiff's insurer would tend to want to settle for less than plaintiffs would like, which would increase the chance of settlement. Note, however, that reputational interests of insurers, as well as the possibility of renegotiation between insurers and insureds, serve to mitigate their conflicts of interest.

5.4.7. *Voluntary sharing of information*

In the discussion of settlement versus trial in Section 5.2, we assumed that the information of parties was somehow exogenously determined: either information was

¹³⁷ Hay (1997) discusses how bifurcated contingent fees (which pay a higher rate if there is no settlement) can help to address this problem. Additional principal-agent problems arise in the context of class actions, where many plaintiffs are joined in a class and there is a free-rider difficulty with regard to supervision of attorneys. See Coffee (1986) and Macey and Miller (1991).

in the background and formed parties' perhaps disparate beliefs, or else information was explicitly presumed to be asymmetric. However, litigants in general have strong motives to share information. See Shavell (1989a). Most obviously, parties will want to share favorable information in order to foster settlement and to improve its terms. A plaintiff, for example, would want to show the defendant information establishing that his losses were in fact higher than the defendant otherwise believes; in this way, the plaintiff can induce the defendant to pay more in settlement and perhaps avoid an impasse leading to trial. Likewise, a defendant would want to show the plaintiff evidence pointing toward his lack of responsibility, in order to convince the plaintiff to accept a lower settlement offer.

In addition, parties will want to reveal information to avoid negative inferences that would be made from their silence. If a plaintiff says nothing about the magnitude of his losses, the defendant will be likely to infer that the plaintiff is withholding information that his losses are lower than average, and if this inference is made, the defendant will not be willing to make an average offer. Both this incentive to avoid negative inferences and the incentive to reveal favorable information tend to produce significant voluntary disclosure and help to explain the high rate of settlement¹³⁸.

Nevertheless, certain information will not be shared, and this helps to explain why some cases do not settle. First, a party may decide against disclosing information because revealed information can often be countered at trial if the opposing side has foreknowledge of it. Second, information may be difficult to share, even though a party wants to do that. For instance, a plaintiff might know that his business losses from a breach of contract will be high, but not be able to demonstrate this during settlement negotiations (because, say, experts will have to be hired for trial to verify the losses). Another difficulty faced by a party who wants to reveal favorable information is that it may consist of the absence of unfavorable information. (For example, if the defendant was not drinking before a traffic accident, his favorable information may be the nonexistence of anyone who saw him drinking, and he may have no way to demonstrate this¹³⁹.) Third, information may not be shared because it is unfavorable and the negative inference drawn from silence is not too strong. Note that the negative inference from silence will be weakened to the extent that some parties do not disclose favorable information for the first two reasons just given.

5.4.8. Required disclosure of information – legal discovery

The courts may require that a litigant disclose certain information to the other side; this practice is known as discovery. It is commonly believed that discovery

¹³⁸ Farber and White (1991) find that many malpractice cases settle after plaintiffs obtain information from defendants.

¹³⁹ If the case does not settle, the plaintiff may ultimately be able to verify the defendant's claim implicitly: investigations may fail to locate any person who saw the defendant drinking (whereas if there really is a witness, there is some probability that the witness would be located).

significantly increases the likelihood of settlement because it reduces differences in parties' information. But, as just emphasized, there may well be substantial voluntary sharing of information, so the influence of compulsory disclosure will not be so great and is in fact nonexistent in a natural model of disclosure. See generally Shavell (1989a), and see also Hay (1994)¹⁴⁰.

Discovery will, nevertheless, tend to increase the rate of settlement and also will affect the terms of settlements. First, when parties would otherwise withhold favorable information to disable the opponent from countering it at trial, discovery will force disclosure, which in turn will make settlement more likely. Second, when parties would otherwise withhold unfavorable information (because the negative inference from so doing would not be too strong), discovery will mandate disclosure and lead to settlement on less favorable terms. It should be noted, however, that such parties with unfavorable information would have settled in the absence of discovery¹⁴¹. Settlement will increase overall because, when those with unfavorable information are required to disclose it, more generous offers will be made to those who remain silent in the face of discovery (perhaps those with favorable information who cannot verify the strength of their cases). Third, the prospect of legal sanctions for false statements may make more credible parties' insistence that they lack certain unfavorable information (such as the assertion that there is no witness who could testify to the party having been drinking before an accident); this would encourage settlement of such cases¹⁴².

Discovery may also be used strategically. Obeying discovery requests is often expensive because significant time and resources may be needed to produce the desired information. This fact raises questions about the use of discovery requests as a threat, for the costs of compliance with discovery requests are, under our current system, generally borne by the side asked to comply. It also raises questions about the socially optimal amount of discovery.

¹⁴⁰ Other models of discovery are Sobel (1989) and Mnookin and Wilson (1998); because these articles do not compare outcomes when there is discovery with outcomes with voluntary sharing of information, they are hard to interpret. See also Cooter and Rubinfeld (1994), a model of discovery without an explicit treatment of asymmetric information, Schrag (1999), a model in which there is judicial control of discovery, and Shepherd (1999), an empirical study of discovery.

¹⁴¹ Those who withhold unfavorable information when there is no compulsory discovery seek to mimic others with favorable information who remain silent (whether because they strategically withhold information or because they cannot credibly verify their favorable information). Accordingly, they receive settlement offers that reflect the average characteristics of the silent group; being those in the group with the least favorable cases, they will be the ones who settle, on terms that are better than they can expect if they were to disclose their unfavorable information. See Shavell (1989a).

¹⁴² There may, however, be limitations on the feasibility of enforcement of discovery obligations. If a side fails to divulge unfavorable information, often this will not come to light (because the case may settle beforehand or because, even if there is a trial, the other side may never learn the truth in any event). Accordingly, very high sanctions for misrepresentations and possibly selective investigation (perhaps by the state) of the veracity of discovery responses may be necessary, although the present system does not follow either course.

5.4.9. *Criminal adjudication*

The analysis of suit and settlement for criminal adjudication [see, for example, Landes (1971) and the literature cited in Section 6.3.8 on plea bargaining] is in some respects similar to that for civil adjudication, but there are differences in parties' incentives that are worthy of note. First, in criminal cases the complaining party is a public prosecutor. Accordingly, litigation decisions will not be based on a simple comparison of litigation costs and the expected gain because the prosecutor neither directly bears these costs nor benefits monetarily from winning (costs are borne by the state and there is no actual recovery). Instead, a prosecutor's decisions will be dictated by the complex of factors determining his salary and his professional future. Nevertheless, one expects there to be a rough congruence between prosecutorial behavior and what the basic theory suggests. For example, prosecutors should tend to bring cases that have higher prospects of success and are less costly, and asymmetric information may impede settlement.

Second, a criminal defendant is often impecunious and will have been assigned a public defender. Not having to pay for his defense, such a defendant will not save legal expenses by settling, making him less willing to settle than otherwise. But those who serve as public defenders or are appointed to represent indigent defendants will often have limited budgets and receive low compensation, so they may exert less effort than what defendants would demand were they not liquidity constrained. Also, criminal defendants, and especially first-time defendants, may not care so much about the magnitude of punishment as about the fact of a criminal conviction or about having to spend any time in prison. If so, they would be less willing to settle than otherwise. There are other possibilities, of course, but our main point is that the basic theory provides only a very rough prediction of suit (prosecution) and settlement in the criminal context.

5.4.10. *Additional aspects of legal procedure*

There are many aspects of legal procedure that merit study but which we do not examine here, due mainly to their having received only limited treatment in the literature. Topics include the burden of proof¹⁴³, rules of evidence (and tribunals' making inferences from evidence)¹⁴⁴, the use of juries¹⁴⁵, the behavior of judges¹⁴⁶, summary adjudication¹⁴⁷, class actions¹⁴⁸, sequential versus joint

¹⁴³ See Davis (1994), Hay and Spier (1997), Kaplow (1994a), R.A. Posner (1973), Rubinfeld and Sappington (1987), and Sobel (1985).

¹⁴⁴ See Daughety and Reinganum (1995), Froeb and Kobayashi (1996), Lewis and Poitevin (1997), R.A. Posner (1999), Schrag and Scotchmer (1994), and Shavell (1989b).

¹⁴⁵ See Klevorick, Rothschild and Winship (1984) (jury deliberations), and Schwartz and Schwartz (1996) (peremptory challenges).

¹⁴⁶ See Cohen (1992), Elder (1987), Higgins and Rubin (1980), Kimenyi et al. (1993), Kornhauser (1992a, 1992b), R.A. Posner (1993a), Ramseyer (1998), and Rasmusen (1994).

¹⁴⁷ See R.A. Posner (1986) on summary jury trials.

¹⁴⁸ See Che (1996), Dam (1975), Miller (1998), and Silver (2000).

adjudication of multiple issues in a single case¹⁴⁹, the sharing of liability among multiple defendants¹⁵⁰, the use of lawyers as advocates for clients¹⁵¹, and the adversarial system of adjudication (in which each side substantially controls its litigation activity) versus the European inquisitorial model (in which the tribunal controls much litigation activity)¹⁵².

5.5. *Legal advice*

Because legal advice is costly, individuals must make decisions whether or not to obtain it, and questions about its social desirability also arise. In discussing the topic of legal advice, it is useful to consider separately *ex ante* legal advice – obtained when a party is contemplating an action – and *ex post* legal advice – secured after a party has acted or someone has been harmed, which is to say, at the stage of possible or actual litigation. A notable difference between the types of advice is that *ex ante* advice can channel behavior directly in conformity with law, whereas *ex post* advice comes too late to accomplish that (although it has indirect effects on behavior). *Ex ante* legal advice was first studied from an economic perspective in Shavell (1988) and Kaplow and Shavell (1992); *ex post* legal advice was initially investigated from this standpoint in Kaplow and Shavell (1989, 1990)¹⁵³.

5.5.1. *Ex ante legal advice: when acts are contemplated*

Advice has private value to a party who is considering taking some action with a possible legal consequence if the advice might lead him to alter his decision. The private value of legal advice is just an instance of the conventional definition of the expected value of information to a decisionmaker, as presented for instance in Raiffa (1968).

The social, as opposed to the private, value of *ex ante* legal advice inheres in the social desirability of advice-induced changes in parties' behavior. In general, advice has positive social value because it promotes adherence to legal rules. The specific nature of the comparison between the social and the private values of legal advice depends on the form of liability. When liability is strict, the private value of legal advice is the same as its social value. This basic conclusion follows essentially because a party's liability burden equals the harm he causes. When, however, liability is based on negligence, the private value of legal advice can be shown to exceed its

¹⁴⁹ See Landes (1993).

¹⁵⁰ See Easterbrook et al. (1980), Klerman (1996), Kornhauser and Revesz (1994), and Polinsky and Shavell (1981).

¹⁵¹ See Dewatripont and Tirole (1999).

¹⁵² See, for example, Daughety and Reinganum (2000a), Langbein (1985), and Shin (1998), and, closely related, Milgrom and Roberts (1986).

¹⁵³ Legal advice was further studied in Bundy and Elhauge (1991, 1993) and Fischel (1998).

social value. The explanation is in part that if a person avoids negligence because of advice, his liability saving will generally be larger than the reduction in expected harm he accomplishes, for he will escape liability entirely even though his non-negligent behavior might still cause harm.

5.5.2. *Ex post legal advice: at the stage of litigation*

The private value of ex post legal advice resides in the possibility that the advice will lead a party to change his decisions about suit, settlement, and trial.

In considering the social value of ex post legal advice, observe first that because such advice is, by its nature, imparted to parties only after they have acted, it cannot have aided them initially in conforming with the law. A firm that does not know whether discharging a chemical waste into a river will violate an antipollution statute obviously cannot be led to behave appropriately by learning what the law is after it decides about discharging the chemical. This simple but fundamental observation means that ex post advice does not raise social welfare in the direct way that ex ante advice does. Nonetheless, ex post advice certainly may influence behavior and social welfare.

Ex post advice that defendants obtain in the course of a lawsuit may affect social welfare by lowering sanctions for those who knowingly violate the law, that is, ex post advice may dilute deterrence of undesirable conduct. Lawyers may lower expected sanctions by advantageous use of legal strategy and, importantly, by counseling defendants on the selection of evidence to present and to suppress. Given that individuals anticipate that their expected sanctions for causing harm will be reduced due to the subsequent availability of legal advice, fewer individuals will be deterred from engaging in undesirable behavior. Thus, legal advice may have negative social value, a point that was early emphasized by Bentham (1827). This reasoning, however, is incomplete, in part because the state may be able to raise overall sanctions to offset the dilution of deterrence due to advice.

Ex post advice may, however, enhance social welfare by increasing otherwise inadequate sanctions that would be imposed on those who knowingly commit sanctionable acts. Specifically, advice may raise expected sanctions because lawyers may help plaintiffs to obtain higher judgments, better reflecting the harms they have sustained. Additionally, ex post advice may raise social welfare by lowering sanctions for defendants who did not violate the law, or who face higher sanctions than they should.

There is thus no way on the basis of logic alone to conclude whether or not ex post advice provided during litigation is on balance socially desirable – whether or not its socially undesirable effect, due to dilution of deterrence, is less important than its desirable effect, due to increased accuracy of legal outcomes for the guilty and for the innocent. Moreover, it is not obvious whether the net effect of advice will be to increase or to decrease the accuracy of adjudication.

Let us, however, restrict attention to ex post legal advice that does increase the accuracy of legal outcomes, and ask how the typically positive social value of this

advice compares to its private value. The general answer to this question is that either the private value of the advice or its social value could be larger, so that the private incentive to spend on the advice could be socially excessive or it could be inadequate. The reason is essentially that explained in Section 5.1.2; the social value of legal advice that increases accuracy inheres in its incentive effect on prior behavior of parties, and this has little connection to the private incentive to spend on advice, for that derives from the amount at stake in litigation. In some contexts, however, the private value of accuracy-enhancing advice will tend to exceed the social value and too much will be purchased. Notably this may often be true of advice about proving the extent of harm, for the reasons we explain in Section 5.3.2.

In sum, the social value of ex post legal advice is complicated to determine, possibly negative and possibly positive, and not closely related to its private value. In certain domains, a plausible conjecture is that, in an appropriate average sense, the private value of ex post advice exceeds its social value.

5.5.3. *Other aspects of legal advice*

Subversion of the law. One issue that we have not mentioned is that advice may directly subvert the law. Lawyers may lower the effective magnitude of sanctions by helping clients to hide assets, and lawyers may also decrease the likelihood of sanctions if they have knowledge of enforcement strategies (such as how the tax authorities choose whom to audit). Of course, lawyers are not supposed to thwart law enforcement, but they have an economic incentive to do so and can fairly easily avoid punishment for it (lawyers give advice in private and can phrase their advice in hypothetical but readily understood terms). From the social perspective, legal advice that frustrates law enforcement is obviously undesirable.

Confidentiality of legal advice. The legal system protects the confidentiality of communications between lawyers and their clients under wide circumstances. Confidentiality of legal advice will benefit clients when there is a positive probability that disclosure of advice would lower its value to them. This would usually be true of advice about the selection of evidence to present in litigation: such advice generally would be robbed of effectiveness if it were disclosed to the opposing side and the court. Confidentiality is also of obvious importance to those obtaining advice subversive of the law. By contrast, confidentiality often should not matter to parties obtaining advice about the legality of an act or about the magnitude or the likelihood of sanctions, because disclosure of such advice will usually not disadvantage them. (This is because they seek advice with the intention of following the law¹⁵⁴.) Still, whatever is the character of strictly legal advice, maintaining the confidentiality of much business or personal information about clients themselves will frequently be of importance to the clients.

¹⁵⁴ See Shavell (1988).

Because protection of confidentiality can benefit clients (and never is a disadvantage to them), it encourages clients to consult with and reveal information to their lawyers. This in itself is sometimes thought to imply that confidentiality is socially desirable. That reasoning, however, is mistaken: confidentiality is socially desirable only if the legal advice that confidentiality encourages is socially desirable, and, as has been explained above, that may not be the case.

Confidentiality of legal work product. The legal system also protects the confidentiality of legal work product (documents and other records of lawyers' effort) that they generate on behalf of clients in connection with litigation. The protection of work product is accomplished principally by denying opposing litigants the right to legal discovery of it. As Easterbrook (1981) stressed, protection of work product encourages lawyers to engage in research on their clients' cases, for much of the value of the research would be lost if it became immediately known to the other side. But whether protection of work product is socially desirable is not evident a priori; for it depends on whether or not the legal advice that the work product supports is socially desirable¹⁵⁵. A further complication is that, even when the advice is socially desirable, the private value of advice, and thus the amount of work product, may be socially excessive.

Quality and truthfulness of advice. To the degree that poor or dishonest advice would be discovered and that lawyers would suffer penalties for having provided such advice, they will have reason not to do so. There are two basic types of penalty lawyers face for furnishing unsound legal advice: loss of business because of damage to reputation, and legal sanctions, in the form of damage judgments arising from malpractice actions, fines assessed by courts, or punishments imposed by professional associations. See Wilkins (1992).

5.6. Appeals

The appeals process – the process whereby a litigant disappointed with the decision of a first-order tribunal can seek reconsideration before a higher tribunal – is a widely observed feature of adjudication; in virtually all legal systems today, there exists a fairly general right of appeal of trial court decisions.

An important social justification for the appeals process concerns correction of error. See Shavell (1995b). Suppose that litigants possess information about the occurrence of error and that appeals courts can frequently verify it. Then litigants may be induced to bring appeals when errors are likely to have been made but not otherwise, because of the costs of appeals. This outcome may be fostered by imposition of fees for bringing appeals, so as to discourage appeal when decisions were likely to have been correct

¹⁵⁵ To illustrate, investigation may be necessary to determine or document facts that will improve accuracy, but the social value of greater accuracy may or may not exceed the cost of investigation. Also, with work-product protection, two parties may engage in duplicative efforts.

and thus unlikely to be reversed¹⁵⁶. In other words, if there is an appropriate price for pursuing appeals, the appeals process can harness the information that litigants have about the occurrence of error and tend to remedy it.

When this process functions well, appeals not only result in error correction, they also do so cheaply, for the legal system is burdened with reconsidering only the subset of cases in which errors were more probably made. This may render society's investment in the appeals process inexpensive in comparison to the alternative it has of improving the accuracy of the trial process (by investing in the length and quality of trial court adjudication). Under that alternative approach, extra expenditure would be required in all cases rather than only in the subset of cases that are appealed. The appeals process, in other words, may be an economical way of correcting error by taking advantage of litigants' information that it has occurred¹⁵⁷.

5.7. *Alternative dispute resolution*

When parties need to resolve a dispute, they may turn not only to the state-sanctioned method of dispute resolution, namely, trial before a court, but also to arbitration and other forms of alternative dispute resolution (ADR)¹⁵⁸. In examining ADR, it is helpful to distinguish between *ex ante* agreements to employ ADR – arrangements made before disputes arise – and *ex post* resort to ADR – use of ADR after disputes have arisen. See Shavell (1995a)¹⁵⁹.

5.7.1. *Ex ante ADR agreements*

Ex ante ADR agreements may be adopted because they are to the mutual benefit of the parties to a contract¹⁶⁰. In particular, ADR may lower the cost of resolving disputes or reduce risk. Second, ADR may engender superior incentives for the parties through greater accuracy of results. Suppose, for instance, that substandard performance of a contract would be correctly assessed by expert arbitrators under ADR but not by courts. Then the parties to the contract might well prefer to adopt ADR because it would induce good performance, thereby raising the willingness of the promisee to pay for the contract. Third, ADR may beneficially affect the volume of adjudication. For example, it may be that the number of disputes brought under the legal process

¹⁵⁶ In fact, however, public fees for appeal are nominal, although private costs may be nontrivial.

¹⁵⁷ Daughety and Reinganum (2000b) and Spitzer and Talley (2000) consider models of the appeals process that include factors different from pure error correction.

¹⁵⁸ We focus on binding ADR; nonbinding ADR, such as mediation, is often used to foster settlement.

¹⁵⁹ On other issues raised by ADR, see Landes and R.A. Posner (1979). It should also be mentioned that there is a literature considering arbitration alone, not arbitration as an alternative to state-authorized litigation. See, for example, Ashenfelter (1989, 1992), Ashenfelter and Bloom (1984), and Farber (1980).

¹⁶⁰ We observe that to obtain many of the benefits noted in the text, the agreement to use ADR must be made *ex ante*; if the parties wait until a dispute arises, it will often be in the interest of one of the parties to refuse to accept ADR.

would be excessive, dissipating substantial resources of the parties without instigating mutually desirable changes in behavior; thus an ADR agreement that would serve to limit the number of disputes would be advantageous.

Because *ex ante* ADR agreements made by knowledgeable parties raise their well-being, it seems that *ex ante* ADR agreements should ordinarily be enforced by the legal system, as they are in fact. It is sometimes suggested that society should go further and subsidize ADR. A subsidy might be justified on second-best grounds, because the state already subsidizes ordinary litigation by not charging litigants for its full costs. It would seem, however, that the optimal solution is to remove the latter subsidy, unless it is justified on the ground of inadequate private incentives to sue.

5.7.2. *Ex post ADR agreements*

Parties will tend to make *ex post* ADR agreements in order to reduce dispute resolution costs and risk. On this account, *ex post* ADR would also tend to be socially desirable. A full evaluation of *ex post* ADR, however, must recognize other effects, notably, how the prospect that parties would adopt ADR *ex post* would affect their *ex ante* behavior. The proper analysis is similar to that bearing on the private versus the social value of settlements, in Section 5.2.3¹⁶¹.

5.8. *Formulation of legal rules*

Economic analysis of the operation of the legal system often takes the legal rules that are enforced as given. The formulation of legal rules itself, however, raises interesting economic issues¹⁶². One issue concerns the optimal level of detail of rules. On one hand, greater detail allows better-tailored control of behavior. On the other hand, greater detail involves higher compliance and litigation costs. Moreover, it cannot be assumed that parties will become informed of the precise content of more detailed rules. See Diver (1983), Ehrlich and Posner (1974), and Kaplow (1995a)¹⁶³.

Another issue is whether rules should be formulated fully *ex ante*, or instead should be incompletely specified initially and fully articulated only *ex post*, during adjudication of particular disputes. Fuller *ex ante* specification is more costly for the state, but may provide greater predictability for parties and hence induce better behavior, and it also may reduce adjudication costs. See Diver (1983), Ehrlich and Posner (1974), and Kaplow (1992c). Full *ex ante* specification of legal rules tends to be advantageous when the governed behavior is frequent and has common characteristics,

¹⁶¹ Indeed, parties' adoption of ADR can be seen as a form of out-of-court settlement because use of ADR means that there will be no trial and instead the parties will be bound by the alternative they have chosen.

¹⁶² For a survey of the literature, see Kaplow (2000).

¹⁶³ The subject of legal complexity has received particular attention in the context of the income tax. See, for example, Blumenthal and Slemrod (1992) and Kaplow (1996).

essentially because of economies of scale (the rule is formulated only once). For infrequent, heterogeneous behavior, leaving the specification of details until the stage of adjudication may save the state expense because many situations for which details may have been provided will never arise. A closely related subject is the issuance of precedents by courts; for example, major disagreements about issuing precedents concern the degree to which details of rulings beyond those necessary to decide the case before a court should be specified and when courts should take the opportunity to announce new legal rules or modify existing ones¹⁶⁴.

Additional issues are presented by the frequent need to modify legal rules. New rules, if fully and immediately applicable, will typically affect the returns to previous investments. The prospect of such application of new rules imposes risk on actors and also affects their investment decisions, but the latter effect tends to be efficient when the legal reform reflects certain economically relevant information or changed circumstances¹⁶⁵. See generally Kaplow (1986a, 1992a).

5.9. Relevance to general incentive schemes

In closing, we suggest that the topic of the operation of the legal system should in substantial respects be viewed as a basic one in the theory of incentives. This is because incentive schemes often require that parties come before authorities who apply rules – that is, the incentive schemes – and this adjudication process is costly. (Even if the adjudication is informal, it will involve expense.) Therefore, many of the issues that we address are of relevance. Notably, questions arise concerning private versus system-appropriate motives to come before authorities who apply rules, for individuals will not take into account total adjudication costs nor the incentive effects of adjudication (such as whether the bringing of employee complaints will induce better behavior in a firm). Also, many of the more particular issues that we consider about litigation and the design of legal procedure – including settlement, discovery, and appeals – have general analogs in other incentive systems¹⁶⁶.

6. Law enforcement

In this section we consider the theory of public enforcement of law – the use of hired agents (inspectors, tax auditors, police) to detect violators of legal rules and

¹⁶⁴ See also Landes and Posner (1976), who analyze the body of precedent as a capital stock that depreciates over time.

¹⁶⁵ For example, suppose that the government learns that a type of emission is harmful and, accordingly, imposes some sort of regulation or corrective tax. It will tend to be efficient for the new rule to apply to preexisting sources of the emissions because the prospect of such application will induce actors to take into account the probability that the emissions will turn out to be harmful. (Compare the discussion of compensation for government takings in Section 3.4.3.)

¹⁶⁶ The discussion in the text concerning the cost and process of verifying variables in incentive schemes obviously bears on whether parties should include variables that analysts might treat as unverifiable but that they understand to be verifiable, to some degree of accuracy, if the parties incur sufficient costs.

to impose sanctions. Outside the scope of our discussion are many other factors that affect compliance with the law, including public programs (such as job training for low-income individuals, which affects their opportunity cost of crime) and private behavior (such as the carrying of guns and the use of locks to prevent theft)¹⁶⁷.

We begin by noting the justification for using public law enforcement rather than relying exclusively on private suits. Then, we analyze basic issues concerning the optimal probability, magnitude, and form of sanctions and the rule of liability. Next, we examine a variety of extensions of the central theory, including accidental harms, error, marginal deterrence, repeat offenders, self-reporting, and incapacitation. Finally, we briefly discuss criminal law in the light of the theory.

Before proceeding, we observe that economically oriented analysis of public law enforcement dates from the eighteenth-century contributions of Montesquieu (1748), Beccaria (1770), and, especially, Bentham (1789), whose investigation of deterrence was sophisticated and expansive. But, curiously, after Bentham, the subject of law enforcement lay essentially dormant in economic scholarship until the late 1960s, when Gary Becker (1968) published a highly influential article, which has led to a voluminous literature¹⁶⁸.

6.1. Rationale for public enforcement

A basic question is why there is a need for public enforcement of law in the light of the availability of private suits brought by victims. The answer depends very much on the locus of information about the identity of injurers. When victims of harm naturally possess knowledge of the identity of injurers, allowing private suits for damages will motivate victims to sue and thus harness the information they have for purposes of law enforcement. This may help to explain why the enforcement of contractual obligations and of accident law is primarily private¹⁶⁹.

When victims do not know who caused harm and penalizing wrongdoing is difficult, society tends to rely instead on public investigation and prosecution; this is broadly true of crimes and of many violations of environmental and safety regulations. Even in contexts where sanctioning violators is difficult, however, we should ask why society cannot rely on inducements to private parties – rewards of some type – to supply information or otherwise to help in sanctioning. One difficulty with such private enforcement is that if a reward is available to anyone, there might be wasteful effort devoted to finding violators (akin to excessive fishing activity). Another problem

¹⁶⁷ On other public policies, see, for example, Donohue and Siegelman (1998) and Wilson and Herrnstein (1985). On private behavior and crime, see, for example, Ayres and Levitt (1998), Cook et al. (1995), Lott and Mustard (1997), and Shavell (1991c).

¹⁶⁸ For surveys and references, see Garoupa (1997), Mookherjee (1997), and Polinsky and Shavell (2000) (which is similar to this section).

¹⁶⁹ It may not be the case, however, that private incentives to bring suit are optimal, as we discuss in Section 5.1.2.

is that the best technologies for finding liable parties often require coordination of many individuals, sometimes on a vast scale. Additionally, it may be advantageous for expensive information systems (fingerprint records, data banks on offenders) to be developed and maintained, even though their benefits would be hard for the private sector to capture fully; such enforcement technologies may constitute natural monopolies. An additional obstacle to private enforcement is that force (or the threat of it) may be needed to gather information, capture violators, and prevent reprisal, yet the state for various reasons may not want to permit private parties to use force. Thus, there appear to exist arguments favoring public enforcement when effort is required to identify and sanction violators¹⁷⁰.

6.2. *Basic theory of enforcement*

Suppose that an individual (or a firm) chooses whether to commit an act that causes harm with certainty (see Section 6.3.1 on uncertain harm). If he commits the act, he obtains some gain and also faces the risk of being caught, found liable, and sanctioned. The rule of liability could be either strict – under which the individual is definitely sanctioned – or fault-based – under which he is sanctioned only if his behavior fell below a fault standard¹⁷¹. The sanction that he suffers could be a monetary fine, a prison term, or a combination of the two.

Whether an individual commits a harmful act is determined by an expected utility calculation. He will commit the act if that would raise his expected utility, taking into account the gain he would derive and the probability, form, and level of sanction that he would then face¹⁷². We will usually first examine the case in which individuals are risk neutral with respect to sanctions, but we will also consider other possibilities.

We assume, as is conventional, that fines are socially costless to employ because they are mere transfers of money, whereas imprisonment involves positive social costs because of the expense associated with the operation of prisons and the disutility

¹⁷⁰ The comparison between public and private enforcement has received some, but modest, attention in the literature. See Becker and Stigler (1974), Landes and Posner (1975), and Polinsky (1980a); see also Friedman (1995) and Shavell (1993a).

¹⁷¹ Fault-based liability is often employed in accident law (the negligence rule) and in many regulatory schemes (which penalize only parties that fail to meet regulatory standards). On reflection, criminal law may be seen to be fault-based; it only punishes certain harmful acts whose characteristics make them almost always undesirable.

¹⁷² We assume that the probability and magnitude of the sanction are known. See, however, Bebchuk and Kaplow (1992) on the case where individuals misperceive the probability of sanctions, Kaplow (1990b) on the case where individuals may make expenditures to acquire information about sanctions, Garoupa (1999) on both of these cases, and Sah (1991) on the process by which individuals learn about actual levels of enforcement. For empirical evidence on offenders' knowledge of expected sanctions, see Wilson and Herrnstein (1985).

due to imprisonment (which is not naturally balanced by gains to others)¹⁷³. We also assume that the higher is the probability of detecting and sanctioning violators, the more resources the state must devote to enforcement¹⁷⁴.

Social welfare generally is presumed to equal the sum of individuals' expected utilities. An individual's expected utility depends on whether he commits a harmful act, on whether he is sanctioned, on whether he is a victim of someone else's harmful act, and on his tax payment, which will reflect the costs of law enforcement, less any fine revenue collected. If individuals are risk neutral, social welfare can be expressed simply as the gains individuals obtain from committing their acts, less the harms caused, and less the costs of law enforcement. (The assumption that individuals' gains are always credited in social welfare could be relaxed without affecting most of our conclusions. The principal difference that altering the assumption would make is that more acts would be treated as socially undesirable and that optimal sanctions and enforcement effort would thus be higher.)

The enforcement authority's problem is to maximize social welfare by choosing enforcement expenditures, or, equivalently, a probability of detection, and also the level of sanctions, their form (a fine, prison term, or combination), and the rule of liability (strict or fault-based).

6.2.1. *Optimal enforcement given the probability of detection*

We consider here optimal enforcement given the assumption that the probability of detection is fixed. Thus, we ask about the optimal form and level of sanctions under strict and fault-based liability and about how the two liability rules compare.

Strict liability. Assume initially that fines are the form of sanction and that individuals are risk neutral. Then the optimal fine f is h/p , the harm divided by the probability of detection, for then the expected fine equals the harm. This fine is optimal because, when the expected fine equals the harm, an individual will commit a harmful act if, and only if, the gain he would derive from it exceeds the harm he would cause. Essentially this basic and fundamental formula was noted by Bentham (1789, p. 173), and it has been observed by many others since.

¹⁷³ Imposing fines may, in fact, be costly, due to the need for adjudication and fine collection. Were we to take this into account, the main effect on our conclusions would be that the optimal expected sanction would be higher because harmful acts would cause not only direct harm but also, if detected, additional administrative costs. (Note, however, that any legal costs borne by the actor are already included in his calculus, so they do not affect the optimal expected sanction.) See, for example, Becker (1968) and Polinsky and Shavell (1992).

¹⁷⁴ We note that when the method of enforcement involves investigating particular acts after they have been committed (rather than auditing or monitoring, such as when police walk a beat), raising the probability of apprehension may, in some ranges, involve lower costs on account of greater deterrence, which reduces the number of acts that need to be investigated to maintain a given probability of detection.

If individuals are risk averse, one might expect the optimal fine to be lower than in the risk-neutral case for two reasons. First, because risk-averse individuals are more easily deterred than risk-neutral individuals, the fine does not need to be as high as before to achieve any desired degree of deterrence. Second, lowering the fine reduces the bearing of risk by individuals who commit the harmful act. However, lowering the fine also increases the number of individuals who commit the harmful act and hence bear risk¹⁷⁵.

Next, assume that imprisonment is the form of sanction, so that social costs will be incurred in imposing sanctions. In this case, there is not a simple formula for the optimal imprisonment term. See Polinsky and Shavell (1984). The optimal term could be such that there is either underdeterrence or overdeterrence, compared to socially ideal behavior. On one hand, a relatively low imprisonment term, implying underdeterrence, might be socially desirable because it means that imprisonment costs are reduced for those individuals who commit harmful acts. On the other hand, a relatively high term, implying overdeterrence, might be socially desirable because it means that imprisonment costs are reduced due to fewer individuals committing harmful acts¹⁷⁶. (For reasons that we will discuss below and because of factors outside the model, our conjecture is that overdeterrence is unlikely to be optimal.)

Now consider the combined use of fines and imprisonment. Here, the main point is that fines should be employed to the maximum extent feasible before resort is made to imprisonment. In other words, it is not optimal to impose a positive imprisonment term unless the fine is maximal. (The maximal fine might be interpreted as the wealth of an individual.) The rationale for this conclusion is that fines are socially costless to impose, whereas imprisonment is socially costly, so deterrence should be achieved through the cheaper form of sanction first. This point is noted by Bentham (1789, p. 183) and Becker (1968); see also Polinsky and Shavell (1984). To amplify, suppose that the fine f is less than the maximal fine f_m and that a positive prison term t is employed. Raise f toward f_m and lower t so as to keep the disutility of the combined sanctions constant. Then deterrence and the amount of harm will be unchanged, but the cost of imposing the imprisonment sanction will fall, raising social welfare. Hence, it must be optimal for the fine to be maximal before imprisonment is used¹⁷⁷. (It can be shown that this argument holds regardless of individuals' attitudes toward risk in either fines or imprisonment.)

¹⁷⁵ An additional complication, which might favor a higher optimal fine when individuals are risk averse, is that individuals who commit a harmful act might obtain such great benefits that they would be wealthier (and thus have a lower marginal utility of wealth) than others even if they paid a fine equal to h/p . Then, raising the fine above h/p would tend to raise social welfare by transferring wealth from those who are sanctioned to others, who have a relatively higher marginal utility of wealth.

¹⁷⁶ See also Kaplow (1990a), who notes that extreme sanctions (zero or the maximal sanction) may well be optimal in the standard model.

¹⁷⁷ See Levitt (1997b) on why it may be optimal to rely more on imprisonment when offenders' wealth cannot be observed. For empirical evidence on the use of fines versus imprisonment, see Lott (1992) and Waldfogel (1995a).

Fault-based liability. As we explained in Section 2.4.1 on accident law, damages equal to harm, in excess of harm, or even somewhat less than harm, will be sufficient to induce optimal behavior under fault-based liability. The same logic is applicable here, where a sanction of h/p – implying that the expected sanction equals expected harm – plays the role of damages in our prior discussion. However, if errors occur in the legal process, deterrence may not be optimal, and excessive deterrence may result. See Section 2.1.1.

When individuals are risk averse or imprisonment is used as a sanction, fault liability has an advantage over strict liability: individuals who behave optimally, but nevertheless cause harm, will not be sanctioned. The socially costly imposition of sanctions is thus avoided. (That is, with fines, individuals who behave properly will not actually bear any risk, and with imprisonment, resources will not be wasted on such individuals.) See Shavell (1982a, 1985a, 1987b). The primary qualification to this argument is that fault-based liability is more difficult to administer because the enforcement authority must determine the fault standard and it must ascertain whether the fault standard was met. See Section 2.1.1. (Moreover, for reasons we discuss in Section 6.3.2 below, strict liability encourages better decisions by injurers regarding their level of participation in harm-creating activities.)

6.2.2. *Optimal enforcement including the probability of detection*

We now consider the optimal system of enforcement when expenditures on enforcement, and hence the probability of detection, are allowed to vary. Consideration of this issue originated with Becker (1968).

Strict liability. Assume first that the sanction is a fine and that individuals are risk neutral. Then the optimal level of the fine is maximal, f_m , and the optimal probability is low (in a sense to be described). The explanation is that if the fine were not maximal, society could save enforcement costs by simultaneously raising the fine and lowering the probability without affecting the level of deterrence: if $f < f_m$, then raise the fine to f_m and lower the probability from p to $(f/f_m)p$; the expected fine is still pf , so that deterrence is maintained, but expenditures on enforcement are reduced, implying that social welfare rises. Becker (1968) suggested this result (although much of his analysis implicitly presumes that the fine is not maximal); Carr-Hill and Stern (1979) and Polinsky and Shavell (1979) note it explicitly.

The optimal probability is low in that there is some underdeterrence: the optimal p is such that the expected fine pf_m is less than the harm h . See Polinsky and Shavell (1984). The reason for this result is that if pf_m equals h , behavior will be ideal, meaning that the individuals who are just deterred obtain gains just equal to the harm. These are the individuals who would be led to commit the harmful act if p were slightly reduced. Decreasing p , in turn, must be socially beneficial because these individuals cause no net social losses (because their gains essentially equal the harm), but reducing p saves enforcement costs.

If individuals are risk averse, the optimal fine may well be less than maximal, as shown in Polinsky and Shavell (1979). This is because the use of a very high fine would impose a substantial risk-bearing cost on individuals who commit harmful acts¹⁷⁸. For further discussion of the optimal fine when individuals are risk averse, see Kaplow (1992b)¹⁷⁹.

Next, assume that the sanction is imprisonment and that individuals are risk neutral in imprisonment, that is, the disutility of a year of imprisonment is the same for each additional year¹⁸⁰. Then the optimal imprisonment term is maximal for essentially familiar reasons: if the imprisonment term is raised and the probability of detection lowered so as to keep the expected sanction constant, neither individual behavior nor the costs of imposing imprisonment are affected (by construction, the expected prison term is the same), but enforcement expenditures fall. See Shavell (1991b). If, instead, individuals are risk averse in imprisonment (the disutility of each year of imprisonment grows with the number of years in prison), there is a stronger argument for setting the imprisonment sanction maximally than when individuals are risk neutral: when the imprisonment term is raised, the probability of detection can be lowered even more than in the risk-neutral case without reducing deterrence. Thus, not only are there greater savings in enforcement expenditures, but also the social costs of imposing imprisonment sanctions decline because the expected prison term falls. See Polinsky and Shavell (1999).

Last, suppose that individuals are risk preferring in imprisonment (the disutility of each year of imprisonment falls with the number of years in prison). This possibility seems particularly important: the first years of imprisonment may create special disutility, due to brutalization of the prisoner or due to the stigma of having been imprisoned at all, and potential offenders may have unusually high discount rates. In this case, the optimal sanction may well be less than maximal: if the sanction were raised, the probability that maintains deterrence could not be lowered proportionally, implying that the expected prison term would rise. See Polinsky and Shavell (1999).

¹⁷⁸ A more particular explanation involves reconsidering the argument that we used in the risk-neutral case. If the fine f is less than f_m , it is still true that f can be raised to f_m and p lowered so that prospective violators' expected utility remains constant; hence, everyone's behavior will be unchanged. However, because of risk aversion, this adjustment implies that pf falls, meaning that fine revenue falls. (The reduction in fine revenue reflects the disutility caused by imposing greater risk on risk-averse individuals.) If individuals are sufficiently risk averse, the decline in fine revenue associated with greater risk-bearing could more than offset the savings in enforcement expenditures from reducing the probability of detection, implying that taxes would have to rise to make up the shortfall; accordingly, social welfare would be lower.

¹⁷⁹ Another reason that optimal fines may not be maximal is that higher sanctions may induce violators to expend additional resources to avoid punishment. See Malik (1990). Further reasons are discussed below.

¹⁸⁰ We did not discuss individuals' attitudes toward the risk of imprisonment above because the points we made there did not depend on this consideration.

Now consider the situation when both fines and imprisonment are employed as sanctions. Recall that under the optimal enforcement policy, the fine must be maximal, for otherwise it cannot be desirable to employ imprisonment. The main point we wish to add is that, unlike when imprisonment is used alone, the optimal imprisonment term may not be maximal even if individuals are risk neutral or risk averse in imprisonment. The basic reason is that, if the imprisonment term is raised and the probability of detection is lowered so as to keep deterrence constant, there will be relatively greater reliance on imprisonment than on fines, which is more socially costly¹⁸¹.

Fault-based liability. The least expensive way to accomplish compliance with the fault standard is to use the highest possible sanction and, given this sanction, the lowest probability of detection that deters individuals who would be at fault. The reason is that, if all individuals who would be at fault are deterred, the only cost incurred is associated with the setting of the probability; this cost is minimized by using the maximal sanction and a correspondingly low probability. Observe that this is true regardless of whether the sanction is a fine or imprisonment and regardless of individuals' attitudes toward the risk of fines or of imprisonment. As noted above, however, determining fault may be difficult. Errors will affect deterrence and will result in some imposition of sanctions that may be socially costly.

6.3. *Extensions of the basic theory*

6.3.1. *Accidental harms*

We initially assumed that individuals consider committing acts that cause harm with certainty. In many circumstances, however, individuals cause harms only by accident – harm occurs only with a probability. For instance, if someone drives while intoxicated, he only creates a likelihood of a collision; or if a firm stores toxic chemicals in a substandard tank, the firm only creates the probability of a harmful spill.

Essentially all that we have said above applies in a straightforward manner when harms are accidental. There is, however, an additional issue that arises when harm is uncertain: a sanction can be imposed either on the basis of the commission of a dangerous act that increases the chance of harm – storing chemicals in a substandard tank – or on the basis of the actual occurrence of harm – only if the tank ruptures and results in a spill. In principle, either approach can achieve optimal deterrence: when individuals are risk-neutral, the sanction for committing a dangerous act would equal the expected harm, and the sanction for causing harm would simply equal the magnitude of the harm itself.

Several factors are relevant to the choice between act-based and harm-based sanctions. See Shavell (1993a). First, act-based sanctions, being based only on expected

¹⁸¹ Reducing the probability reduces the expected disutility attributable to fines (which are constant in nominal amount, at the maximum level); to keep deterrence constant, expected disutility attributable to imprisonment must rise. See Shavell (1991b).

harm, need not be as high to accomplish a given level of deterrence, and thus offer an underlying advantage over harm-based sanctions because of limitations in parties' assets. See Section 2.6. Such lower sanctions will also be beneficial when parties are risk averse. Second, act-based sanctions and harm-based sanctions may differ in the ease with which they can be applied. In some circumstances, act-based sanctions may be simpler to impose (it might be easier to determine whether an oil shipper properly maintains its vessels' holding tanks than to detect whether one of the vessels leaked oil into the ocean). In other circumstances, harm-based sanctions may be more readily applied (it may be easy to identify that a truck exploded but may be difficult to detect a truck illegally carrying explosives). Third, calculation of the appropriate sanction may be less difficult in one context or the other: actual harm may be apparent when it occurs, whereas the probability may be difficult to assess at the time of an act; or expected harm may be statistically determinable, but identifying actual harm (for example, tracing particular pollutants to particular victims) may be nearly impossible.

6.3.2. *Level of activity*

We have been assuming that the sole decision that an individual makes is whether to act in a way that causes harm when engaging in some activity. In many contexts, however, an individual also chooses whether to engage in that activity, or, more generally, at what level to do so. Thus, as we discussed in Section 2.1.3 on liability for accidents, individuals decide both how carefully to drive and how much to drive. Similarly, firms decide on a pollution technology and a level of production. And, as we observed previously, even parties who act with appropriate care may impose harm; hence, their activity levels will tend to be optimal only if they bear the cost of that residual harm. Thus, under strict liability, choices about activity levels tend to be correct, but under fault-based liability, parties generally will participate in activities to a socially excessive extent. An important application of this point concerns safety and environmental regulation. Such regulation is typically framed in terms of standards that have to be met, but which, if met, free regulated parties from liability. Under such regulation, levels of regulated activities tend to be excessive.

6.3.3. *Enforcement error*

Errors of the two classic types can occur in law enforcement: an individual who should be found liable might mistakenly not be found liable, and an individual who should not be found liable might mistakenly be found liable. Let the probabilities of these errors be ε_1 and ε_2 , respectively, for an individual who has been detected. Thus, an individual will commit the wrongful act when his gain g net of his expected fine if he does commit it leaves him better off than paying the expected fine if he does not commit it, namely, when $g - p(1 - \varepsilon_1)f > -p\varepsilon_2f$, or, equivalently, when $g > (1 - \varepsilon_1 - \varepsilon_2)pf$.

The first point to note is that, as emphasized in Png (1986), both types of error reduce deterrence: the term $(1 - \varepsilon_1 - \varepsilon_2)pf$ is declining in both ε_1 and ε_2 . The first

type of error diminishes deterrence because it lowers the expected fine if an individual violates the law. The second type of error, when an individual is mistakenly found liable, also lowers deterrence because it reduces the marginal benefit of complying with the law. Because errors dilute deterrence, they reduce social welfare. Specifically, to achieve any level of deterrence, the probability p must be higher to offset the effect of errors. Also, when sanctions are socially costly, greater sanctioning costs may be incurred to achieve a given level of deterrence¹⁸². See generally Kaplow and Shavell (1994a).

Now consider the optimal choice of the fine. Given any probability of detection, the dilution in deterrence caused by errors requires a higher fine to restore deterrence. If the probability and the fine are variable, then, as before, the optimal fine is maximal for the now familiar reason.

Next, consider the possible risk aversion of individuals. As we emphasized, the optimal fine under strict liability may well be less than maximal when individuals are risk averse, in part because lowering the fine from the maximum level reduces the bearing of risk. Introducing the possibility of errors may increase the desirability of lowering the fine because individuals who do not violate the law are subject to the risk of having to pay a fine¹⁸³. Indeed, because the number of persons who do not violate the law often would far exceed the number who do, the desire to avoid imposing risk on the former group can lead to a substantial reduction in the optimal fine.

The possibility of error has analogous effects on our analysis of nonmonetary sanctions. The effect of error on the performance of fault-based liability was already noted in Section 6.2.1.

Finally, observe that, although we have treated the probabilities of error as fixed, they can be influenced by procedural choices: generally, increasing resources devoted to investigation and adjudication tends to decrease errors, and adjusting the burden of proof affects the tradeoff between the two types of errors. Because both types of error reduce deterrence and increase the imposition of socially costly sanctions for a given level of deterrence, expenditures made to reduce errors may be socially beneficial. See Kaplow and Shavell (1994a).

6.3.4. General enforcement

Enforcement sometimes is general in the sense that several different types of violations will be detected by an enforcement agent's activity. For example, a police officer waiting at the roadside may notice a driver who litters as well as one who goes through a red light or who speeds, and a tax auditor may detect a variety of infractions when he examines a tax return. To analyze such situations, suppose that a single probability

¹⁸² First, sanctions will sometimes be imposed on those who did not commit the harmful act. Second, to maintain a given gap in disutility from sanctions due to committing the harmful act, the expected sanction for actual injurers must rise as well.

¹⁸³ See Block and Sidak (1980).

of detection applies uniformly to all harmful acts, regardless of the magnitude of the harm. (The contrasting assumption is that enforcement is specific, meaning that the probability is chosen independently for each type of harmful act¹⁸⁴.)

The main point that we want to make is that in contexts in which enforcement is general, the optimal sanction rises with the severity of the harm and is maximal only for relatively high harms. See Shavell (1991b); Mookherjee and Png (1992) is closely related. To explain, assume that liability is strict, the sanction is a fine, and injurers are risk neutral. Let $f(h)$ be the fine given harm h . Then, for any given general probability of detection p , the optimal fine schedule is h/p , provided that h/p is feasible; otherwise – for high h (all h such that $h/p > f_m$) – the optimal fine is maximal. This schedule is obviously optimal given p because it implies that the expected fine equals harm, thereby inducing ideal behavior, whenever that is possible.

The question remains whether it would be desirable to lower p and raise fines to the maximal level for the range of relatively low-harm acts for which h/p is less than maximal. The answer is that if p is reduced for the relatively low-harm acts (and the fine raised for them), then p , being general, is also reduced for the high-harm acts for which the fine is already maximal, which worsens the extent of underdeterrence of these acts. The decline in deterrence of high-harm acts may cause a greater social loss than the savings in enforcement costs from lowering p . To express this point differently, p must be sufficiently high to avoid significant underdeterrence of high-harm acts (for which fines are maximal). But since this p also applies to less harmful acts, the fines for them do not need to be maximal in order to deter them appropriately.

6.3.5. Marginal deterrence

Sometimes a person may consider which of several harmful acts to commit, for example, whether to release only a small amount of a pollutant into a river or a large amount, or whether only to kidnap a person or also to kill him. In such contexts, the threat of sanctions influences not only whether individuals are deterred from committing harmful acts but also, for those who are not deterred, which harmful acts they will then choose to commit. Notably, undeterred individuals will have a reason to commit less harmful rather than more harmful acts if expected sanctions rise with harm – a phenomenon that is sometimes referred to as marginal deterrence, named by Stigler (1970). The benefits of achieving marginal deterrence were noted long ago by Beccaria (1770, p. 32) and Bentham (1789, p. 171). There may, however, be costs of accomplishing marginal deterrence: for sanctions to rise with the magnitude of harm it may be necessary to apply lower sanctions to less harmful acts, which will reduce the deterrence of such acts.

¹⁸⁴ These assumptions correspond to different law enforcement technologies. Investigation (the police following leads after the commission of a particular crime) tends to be specific, whereas auditing and monitoring tend to be general.

Two additional observations should be made about marginal deterrence. First, marginal deterrence can be promoted by adjusting the probability of detection as well as the magnitude of sanctions. (Thus, rather than achieving marginal deterrence by lowering the sanction for the less harmful act, the state can lower the probability of detection for that act; this accomplishes the same result with regard to deterrence and saves enforcement resources¹⁸⁵.) Second, marginal deterrence is naturally accomplished if the expected sanction equals harm for all levels of harm; for instance, if a polluter's expected fine would rise from \$100 to \$500 if he dumps five gallons instead of one gallon of waste into a lake, where each gallon causes \$100 of harm, his marginal incentives to pollute will be correct. For formal analyses of marginal deterrence, see Friedman and Sjostrom (1993), Mookherjee and Png (1994), Shavell (1992), and Wilde (1992).

6.3.6. *Repeat offenders*

In practice, the law often sanctions repeat offenders more severely than first-time offenders. Note initially, however, that sanctioning repeat offenders more severely cannot be socially advantageous if deterrence always induces first-best behavior¹⁸⁶. Thus, it is only when there is underdeterrence (which is often optimal even when acts are always undesirable, as with many crimes) that it might be optimal to punish repeat offenders more severely.

The main justification for a greater sanction for repeat offenders is that repeat offenders may reveal themselves to be different in some manner that bears on the optimal sanction¹⁸⁷. Another reason to raise sanctions is if additional imprisonment has less deterrent effect per unit, as discussed in Section 6.2.2. We also note that greater sanctions for repeat offenders not only deter repeat offenses but also initial offenses (if these might be followed by later offenses). For analyses of repeat offenses, see Chu et al. (2000), Landsberger and Meilijson (1982), Polinsky and Rubinfeld (1991), Polinsky and Shavell (1998b), and Rubinstein (1979).

¹⁸⁵ When enforcement is general, the same probability will be applicable to a range of offenses, in which case adjusting sanctions may be the only way to achieve marginal deterrence.

¹⁸⁶ If the sanction for polluting and causing a \$1000 harm is \$1000, then any person who pollutes and pays \$1000 is a person whose gain from polluting (say, the savings from not installing pollution control equipment) must have exceeded \$1000. Social welfare therefore is higher as a result of his polluting. If such an individual polluted and was sanctioned in the past, that only means that it was socially desirable for him to have polluted previously. Raising the current sanction because of his having a record of sanctions would overdeter him now.

¹⁸⁷ Observe that the mere fact that the sanction for the first offense was inadequate to deter repeat offenders is not enough to justify a higher sanction, for this fact was known at the time they were sanctioned for the first offense.

6.3.7. *Self-reporting*

We have thus far assumed that individuals are subject to sanctions only if they are detected by an enforcement agent, but in fact parties sometimes disclose their own violations to enforcement authorities. For example, firms often are required to report their violations of environmental and safety regulations, individuals sometimes notify police of their involvement in traffic accidents, and even criminals occasionally turn themselves in. We explain here why it is generally socially desirable for the structure of enforcement to be such as to encourage self-reporting. See Kaplow and Shavell (1994b) and Malik (1993).

Self-reporting can be induced by the state lowering the sanction for individuals who disclose their own infractions. Moreover, the reward for self-reporting can be made small enough that deterrence is only negligibly reduced. To amplify, assume for simplicity that the sanction is a fine f , that the probability of detection is p , and that individuals are risk neutral. If an individual commits a violation and does not self-report, his expected fine is pf . Suppose the fine if a violator self-reports is set just below pf , say at $pf - \varepsilon$, where $\varepsilon > 0$ is arbitrarily small. Then the violator will want to self-report but the deterrent effect of the sanction will be (approximately) the same as if he did not self-report.

Given that self-reporting can be induced, essentially without compromising deterrence, why exactly is self-reporting socially advantageous? One reason is that self-reporting reduces enforcement costs: when a party self-reports, the enforcement authority does not have to identify and prove who the violator was; if a polluter or a burglar turns himself in, investigatory resources are saved¹⁸⁸. Second, self-reporting reduces risk, and thus is advantageous if injurers are risk averse. Drivers bear less risk because they know that if they cause an accident, they can (and will be led to) report this to the police and suffer a lower and certain sanction, rather than face a substantially higher sanction (such as for hit and run driving) imposed only with some probability. Third, the magnitude of harm caused sometimes will be mitigated as a consequence of self-reporting; for example, when firms are induced to report leaks of toxic substances when they occur, prompt remediation is more likely to take place¹⁸⁹.

6.3.8. *Plea bargaining*

Plea bargaining refers to settlement negotiations between a public prosecutor and a criminal defendant. We examined this subject in Section 5.4.9. On plea bargaining, see, for example, Froeb (1993), Grossman and Katz (1983), Kobayashi and Lott (1996), Miceli (1996), and Reinganum (1988, 2000).

¹⁸⁸ When enforcement is accomplished by means of auditing or monitoring, self-reporting results in only modest savings. For self-reporting then only reduces, perhaps slightly (if most individuals comply with the law), the population of individuals to be audited or monitored.

¹⁸⁹ See Innes (1999).

6.3.9. *Corruption of law enforcement agents*

An enforcement agent and a potential violator might well find it mutually profitable to make an agreement under which the violator pays the agent to keep silent. This problem of corruption would seem to be worse the larger is the sanction faced by a violator. To combat corruption and the undermining of deterrence that it brings about, two general approaches can be employed. One is to raise the overall level of sanctions, so that bargained-for payments will also rise. This, however, is a gross strategy, and also suffers from the limit on the magnitude of sanctions that can actually be imposed. The second approach is to attempt to control corruption by use of sanctions against those who participate in it. This approach is expensive and involves issues of enforcement in its own right¹⁹⁰.

6.3.10. *Principal-agent relationship*

Although we have assumed that an injurer is a single actor, the injurer is often an agent of some principal. For example, the agent could be an employee of a firm, or the agent could be a subcontractor working for a contractor.

When harm is caused by the behavior of principals and their agents, many of the conclusions of our prior analysis carry over to the sanctioning of principals. Notably, if a risk-neutral principal faces an expected fine equal to harm done, he will in effect be in the same position vis-à-vis his agent as society is vis-à-vis a single potential violator of law. See Newman and Wright (1990). Consequently, the principal will behave socially optimally in controlling his agents and, in particular, will contract with them and monitor them in ways that will give the agents socially appropriate incentives to reduce harm¹⁹¹.

A question about enforcement that arises when there are principals and agents is the allocation of financial sanctions between the two parties¹⁹². It is apparent, however, that the particular allocation of sanctions does not matter when, as would be the natural presumption, the parties can reallocate the sanctions through their own contract. For example, if the agent finds that he faces the risk of a large fine but is more risk averse than the principal, the principal can assume the risk; conversely, if the risk of the fine is imposed on the principal, he will retain it. Thus, the post-contract sanctions that the agent bears are not affected by the particular division of sanctions initially selected by the enforcement authority.

¹⁹⁰ For fairly general, mainly informal discussions of corruption, see Becker and Stigler (1974), Klitgaard (1988), Rose-Ackerman (1978, 1999), and Shleifer and Vishny (1993); and for models of various aspects of corruption, see, for example, Bowles and Garoupa (1997), Cadot (1987), Mookherjee and Png (1995), and Polinsky and Shavell (2001). In the principal-agent context, analogous problems are examined in Kofman and Lawarree (1993) and Tirole (1986).

¹⁹¹ But, as Arlen (1994) indicates, firms' internal monitoring of agents might be discouraged if such monitoring makes firms' exposure to external sanctions more likely.

¹⁹² See Kraakman (1984).

The allocation of monetary sanctions between principals and agents does matter if some allocations allow the pair to reduce their total burden. An important example is when a fine is imposed only on the agent and he is unable to pay it because his assets are less than the fine; see Kornhauser (1982) and Sykes (1981). Then, he and the principal (who often would have higher assets) would jointly escape part of the fine, diluting deterrence. Imposing the fine on the principal rather than on the agent avoids this problem¹⁹³.

A closely related point is that the imposition of imprisonment sanctions on agents may be desirable when their assets are less than the harm that they can cause, even if the principal's assets are sufficient to pay the optimal fine. See Polinsky and Shavell (1993). The fact that an agent's assets are limited means that the principal may be unable to control him adequately through use of contractually determined penalties. For example, a firm may not be able, despite the threat of salary reduction or dismissal, to induce its employees never to rig bids. In such circumstances, it may be socially valuable to use the threat of personal criminal liability and a jail sentence to improve the control of agents' misconduct.

6.3.11. *Incapacitation*

Our discussion of public enforcement has focused on the deterrent effect of sanctions. However, a different way for society to reduce harm is by imposing sanctions that remove parties from positions in which they are able to cause harm – that is, by incapacitating them. Imprisonment is the primary incapacitative sanction, although there are other examples: individuals can lose their drivers' licenses; businesses can lose their right to operate in certain domains, and the like. Here, we consider imprisonment, but what we say applies to incapacitative sanctions generally. On the economic theory of incapacitation, see Shavell (1987c).

To better understand the role of public enforcement when sanctions are incapacitative, suppose that the sole function of sanctions is to incapacitate; that is, sanctions do not deter. In this case, continued imprisonment will be desirable as long as the reduction in crime from incapacitation exceeds the costs of imprisonment. Observe that this condition could hold for a long period, even for offenses that are not the most serious. There is, however, evidence that the proclivity to commit crimes declines sharply with age after a certain point. We also note that, as a matter of logic, the incapacitative rationale might imply that a person should be imprisoned even if he has not committed a crime – because the danger he poses to society makes incapacitating him worthwhile. In practice, however, the fact that a person has committed a harmful act may be the best basis for predicting his future behavior, in

¹⁹³ The converse problem, when the principal has insufficient assets, may also arise. Then, it may be optimal to hold agents or other contracting parties, such as lawyers or lenders, liable as well. See, for example, Kraakman (1986) and Pitchford (1995).

which case the incapacitation rationale would suggest imprisoning an individual only if he has committed such an act¹⁹⁴.

Several comments may be made on the relationship between optimal enforcement when incapacitation is the goal versus when deterrence is the goal. First, when incapacitation is the goal, the optimal magnitude of the sanction is independent of the probability of apprehension, which contrasts with the case when deterrence is the goal. Second, when deterrence is the goal, the probability and magnitude of sanctions depend on the ability to deter, and if this ability is limited (as, for instance, with the insane), a low expected sanction may be optimal, whereas a high sanction still might be called for to incapacitate.

6.3.12. *Empirical evidence on law enforcement*

A great deal of empirical work has been devoted to controlling criminals. See, for example, Anderson (1999), Blumstein et al. (1978), Cook and Zarkin (1985), Dilulio and Piehl (1991), Ehrlich (1973, 1975), Eide (1994, 2000), Grogger (1991), Kessler and Levitt (1999), Levitt (1996, 1997a, 1998a,b), Nagin (1978), Pyle (1983), Tauchen et al. (1994), Viscusi (1986b), and Witte (1980). Some of this literature, however, does not distinguish between deterrence and incapacitation as the source of reductions in crime following from greater law enforcement. A problematic issue with which the literature grapples is the simultaneity problem. Notably, when greater law enforcement is not associated with a significant reduction in crime, the explanation could be either that deterrence and incapacitation are relatively unimportant, or else that their importance is masked because enforcement effort and sanctions are increased in response to higher crime rates.

6.4. *Criminal law*

The subject of criminal law may be viewed in the light of the theory of public law enforcement. See R.A. Posner (1985) and Shavell (1985a). First, that the acts in the core area of crime – robbery, murder, rape, and so forth – are punished by the sanction of imprisonment makes basic sense. Were society to rely on monetary penalties alone, deterrence of the acts in question would be grossly inadequate. Notably, the probability of sanctions for many of these acts is small, making the money sanction necessary for deterrence large, but the assets of many individuals who might commit these acts are quite limited; hence, the threat of prison is needed for deterrence. Moreover, the incapacitative aspect of imprisonment is valuable because of difficulties in deterring many of the individuals who are prone to commit criminal acts.

Second, many of the doctrines of criminal law appear to enhance social welfare. This seems true of the basic feature of criminal law that punishment is not imposed

¹⁹⁴ An exception may arise for certain types of mental illness, in which case we do in fact incapacitate individuals.

on all harmful acts but instead is usually confined to those that are undesirable. (For example, murder is subject to criminal sanctions, but not all accidental killing.) As we have stressed, when the socially costly sanction of imprisonment is employed, the fault system is desirable because it results in less frequent imposition of punishment than strict liability. The focus on intent in criminal law, another of its defining features, may be sensible with regard to deterrence because those who intend to do harm are more likely actually to cause harm, may be more inclined to conceal their acts, and may be harder to discourage because of the benefits they anticipate. That unsuccessful attempts to do harm are punished in criminal law is an implicit way of raising the likelihood of sanctions for undesirable acts¹⁹⁵. Study of specific doctrines of criminal law seems to afford a rich opportunity for economic analysis¹⁹⁶.

Third, the level of sanctions commonly employed is in some respects in accord with the theory concerning optimal enforcement; notably, offenses that are relatively more serious or more difficult to detect tend to be punished more severely than others. Sanctions, however, do not always seem to be as high as the theory suggests would be optimal. To be sure, the theory surveyed above provides many reasons that optimal sanctions may be less than maximal. Yet some sanctions appear to be substantially lower than can readily be explained by the theory¹⁹⁷. An important reason for this may be that the public would view as unfair the imposition of punishment that was disproportionate to the magnitude of an offense, although the precise nature of this constraint on the use of sanctions is difficult to ascertain¹⁹⁸.

7. Criticism of economic analysis of law

Many observers, and particularly non-economists, view economic analysis of law with skepticism. In this section, we briefly note some of the most common criticisms.

7.1. *Positive analysis*

It is often claimed that individuals and firms do not respond to legal rules as rational maximizers of their well-being. Sometimes this criticism of the conventional economic approach verges on an outright rejection of the use of models. Such an extreme view reflects a failure to appreciate the role of simplifying assumptions, and, accordingly,

¹⁹⁵ See Shavell (1990).

¹⁹⁶ See R.A. Posner (1985) and Shavell (1985a) on the various doctrines of criminal law, and Kaplow (1990b) and Murphy and O'Hara (1997) on whether ignorance of the law should excuse criminal liability. See also Fischel and Sykes (1996) and Khanna (1996) on corporate criminal liability.

¹⁹⁷ For example, fines for many traffic violations are quite low; increasing them might achieve significant savings in enforcement resources without raising other serious problems.

¹⁹⁸ An implication is that jurors may not always be willing to convict if sanctions are viewed as excessive. For a model of optimal enforcement in this case, see Andreoni (1991).

it can be largely dismissed. Frequently, however, the criticism is limited to particular contexts. For example, it is often asserted that decisions to commit crimes are not governed by economists' usual assumptions. Ultimately, such criticisms raise questions that can only be answered by empirical investigation.

It is also suggested that, in predicting individuals' behavior, certain standard assumptions should be modified. For example, in predicting compliance with a law, the assumption that preferences be taken as given would be inappropriate if a legal rule would change people's preferences, as some say was the case with civil rights laws. In addition, laws may frame individuals' understanding of problems, which could affect their probability assessments or willingness to pay. See, for example, Kahneman et al. (1990) on the assignment of entitlements and the Coase theorem. The emerging field of behavioral economics and work in various disciplines that address social norms is beginning to examine these sorts of issues¹⁹⁹.

7.2. Normative analysis

7.2.1. Distribution of income

A frequent criticism of economic analysis of law concerns its focus on efficiency, to the exclusion of the distribution of income. The claim of critics is that legal rules – such as the choice between strict liability and negligence to govern automobile–pedestrian accidents – should be selected in a manner that reflects their effects on the rich and the poor.

There is not a good reason, however, to employ legal rules to accomplish redistributive objectives given the general alternative of achieving sought-after redistribution through the income tax and transfer programs. Such direct methods of redistribution tend to be superior to redistribution through the choice of legal rules: selecting legal rules other than those that are most efficient in order to effect redistribution is itself costly, and it also will distort individuals' labor–leisure decision in the same manner as does the income tax. See Shavell (1981) and Kaplow and Shavell (1994c)²⁰⁰.

Moreover, it is difficult to redistribute income systematically through the choice of legal rules. In the first place, many individuals are never involved in litigation. Also, for those who are, there is substantial income heterogeneity both among plaintiffs and

¹⁹⁹ See, for example, Baron (1994), Jolls et al. (1998), Kahneman et al. (1982), and Rabin (1998).

²⁰⁰ There are subtle qualifications to this claim of the sort identified in the optimal income tax literature (for example, that activities that make leisure relatively more attractive than work should be disfavored), but these qualifications are largely independent of the main criticism at issue, holding that legal rules should be designed to favor whichever party has lower income. On one such qualification, see Sanchirico (2000) and Kaplow and Shavell (2000). Furthermore, this claim assumes that redistribution will have the same effect on labor effort however it is accomplished, whereas it is possible that individuals would over- or underestimate the extent of redistribution accomplished by legal rules. See Jolls (1998).

among defendants. Additionally, in contractual contexts, the choice of a legal rule often will not have any effect on distribution because contract terms, notably, the price, will adjust so that any agreement into which parties enter will continue to reflect the initial bargaining power of each party.

7.2.2. Victim compensation

Another major criticism of economic analysis of law is that it usually emphasizes the effects of legal rules on behavior, but not the compensation of victims – which, some believe, is the main purpose of private law. Economic analysis does not, though, ignore victim compensation per se; victim compensation is relevant to social welfare if victims are risk averse. However, as we have discussed, if victims can obtain insurance, as is often possible, then the legal system need not be relied on to provide compensation. Moreover, providing compensation through legal rules tends to be significantly more expensive than doing so through insurance²⁰¹.

7.2.3. Concerns for fairness

An additional source of criticism is that the welfare-economic approach slights important concerns about fairness, justice, and rights. Some of these notions refer implicitly to the appropriateness of the distribution of income and, accordingly, are encompassed by our preceding remarks. Also, to some degree, the notions are motivated by instrumental concerns. For example, the attraction of just punishment must inhere in part in its deterrent effect, and the appeal of obeying contractual promises must rest in part on the beneficial effect this has on production and exchange. To this extent, critics' concerns are already taken into account in standard welfare-economic analysis.

However, many who advance ideas of fairness and cognate notions do not regard them merely as some sort of proxy for attaining instrumental objectives. Instead, they believe that satisfying the notions is intrinsically valuable. This view too can be partially reconciled with economists' conception of social welfare: if individuals have a taste for a legal rule or institution because they regard it as fair, that should be credited in the determination of social welfare, just as any taste should. (Note that, in this case, the importance of fairness is converted from a philosophical issue to an empirical question about individuals' tastes.)

But many uphold the view that notions of fairness are important as ethical principles in themselves, without regard to any possible relationship the principles may have to individuals' welfare. This opinion is, of course, the subject of longstanding debate

²⁰¹ The reader will also recall from Section 2.2 that there may be additional reasons that providing compensation through the legal system is undesirable, including adverse effects of victims' incentives and imposing risk on injurers.

among moral philosophers²⁰². Some readers (along with us) may be skeptical of normative views that are not grounded in individuals' well-being because embracing such views entails a willingness to sacrifice individuals' well-being. Indeed, consistently pursuing any non-welfarist principle will sometimes result in everyone being made worse off²⁰³. Nevertheless, it is clear that such views will be reflected in criticism of economic analysis of law for the foreseeable future²⁰⁴.

7.3. Purported efficiency of judge-made law

Also criticized is the contention of some economically-oriented academics – notably, R.A. Posner (1975) and Landes and Posner (1987a) – that judge-made law tends to be efficient (in contrast to legislation, which is said to reflect the influence of special interest groups)²⁰⁵. Instead, critics believe that the judge-made law is guided by notions of fairness, justice, and rights, and thus will not necessarily be efficient. Several observations about these competing views may be made. First, one would certainly expect legal rules to promote efficiency, at least in a very approximate sense, for that is consistent with many notions of fairness and with common sense. But second, one would not expect that legal rules would be efficient in a detailed sense for a variety of reasons²⁰⁶. Third, we note that judge-made law is peculiar to “common-law” countries, those of the former British Commonwealth, yet common-law legal rules are not markedly different from those in the civil-law countries of Continental Europe, which rely more on statutes and less on judicial development than common-law countries. Moreover, to the extent that legal rules in common-law and civil-law systems differ, it is hardly clear that the typical civil-law rules are less efficient²⁰⁷. Finally, it should be emphasized that the economic efficiency thesis is a particular descriptive claim about the law, and its validity does not bear on the power of

²⁰² We note, however, that much of the philosophical debate is about what principles should guide personal behavior in everyday life, which may not be applicable to the determination of what principles should guide social policy. A related distinction is emphasized in Hare (1981).

²⁰³ See Kaplow and Shavell (2001).

²⁰⁴ Kaplow and Shavell (2002a) provides an extensive investigation of the issues discussed in this section. See generally Sen and Williams (1982) for representative views of leading economists and philosophers on normative analysis.

²⁰⁵ The argument is advanced by examining particular common-law rules and presenting arguments that they are efficient. In addition, the argument has been examined in the context of models of common-law evolution. See Cooter and Kornhauser (1980), Priest (1977), and Rubin (1977).

²⁰⁶ Legal rules are arguably influenced by notions of fairness, which only loosely reflect instrumental objectives, and are also determined by a multiplicity of institutional and historically contingent factors. Moreover, even if lawmakers were attempting to promote efficiency over the course of history, they would have had a limited understanding of the relevant theory and little empirical evidence to guide them.

²⁰⁷ See, for example, Shavell (1987a), who makes some comparisons of typical common-law and civil-law rules on accident law.

economics to predict behavior in response to legal rules or on the merits of normative economic analysis of law.

8. Conclusion

Having surveyed the basic areas of economic analysis of law, let us comment on possible directions for future research. Although accident liability has been fairly well explored, relatively little formal work has been done on the subject of property law. With regard to contract law, most analysis has concerned remedies for breach, but little attention has been paid to contract formation. In the area of litigation, research effort so far has focused on settlement versus trial, whereas other aspects of litigation, including its adversarial character and its optimal design, merit study. With regard to law enforcement, an issue worthy of further consideration is the incentives of enforcers (including the problem of corruption); also, many of the doctrines of criminal law deserve investigation.

Moreover, there is a very general need for empirical work on the legal system to be undertaken. One area of study is suggested by the fact that, as we emphasized, the private and the social incentives to use the legal system can be expected to diverge. Consequently, society needs estimates of the benefits and costs of legal activity in broad domains (such as auto accidents, product liability) in order to devise appropriate policy. Another potential research area is investigation of the provisions in actual contracts in various settings, in order to test and to inform the extensive theoretical work in the field. An additional subject that seems ripe for empirical study is the litigation process, especially the determinants of suit and settlement decisions and the effects of procedures for the conduct of discovery and trial. Overall, in the areas of law considered in this survey, relatively few of the particular topics that were covered have been the subject of serious empirical work, and the opportunity for progress appears to be substantial.

Acknowledgments

We thank Alan Auerbach, Steven Levitt, A. Mitchell Polinsky, Kathryn Spier, and Tanguy van Ypersele for comments, Judson Berkey, Jerry Fang, Bo Li, Eric Selmon, and Chad Shirley for research assistance, and the John M. Olin Center for Law, Economics, and Business at Harvard Law School for financial support.

References

Administrative Office of the United States Courts (1995), "Judicial business of the United States Courts", Report of the Director (Administrative Office of the United States Courts, Washington, D.C.).

- Aghion, P., and P. Bolton (1987), "Contracts as a barrier to entry", *American Economic Review* 77:388–401.
- Aghion, P., and B.E. Hermalin (1990), "Legal restrictions on private contracts can enhance efficiency", *Journal of Law, Economics, & Organization* 6:381–409.
- Aghion, P., M. Dewatripont and P. Rey (1994), "Renegotiation design with unverifiable information", *Econometrica* 62:257–282.
- Alchian, A.A., and H. Demsetz (1972), "Production, information costs, and economic organization", *American Economic Review* 62:777–795.
- Alston, L.J., G.D. Libecap and R. Schneider (1996), "The determinants and impact of property rights: land titles on the Brazilian frontier", *Journal of Law, Economics, & Organization* 12:25–61.
- Anderson, D.A. (1999), "The aggregate burden of crime", *Journal of Law and Economics* 42:611–642.
- Andreoni, J. (1991), "Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime?" *Rand Journal of Economics* 22:385–395.
- Arlen, J. (1994), "The potentially perverse effects of corporate criminal liability", *Journal of Legal Studies* 23:833–867.
- Ashenfelter, O. (1989), "Evidence on U.S. experiences with dispute resolution systems", in: W.-C. Huang, ed., *Organized Labor at the Crossroads* (W.E. Upjohn Institute for Employment Research, Kalamazoo, MI) pp. 139–162.
- Ashenfelter, O. (1992), "An experimental comparison of dispute rates in alternative arbitration systems", *Econometrica* 60:1407–1433.
- Ashenfelter, O., and D.E. Bloom (1984), "Models of arbitrator behavior: theory and evidence", *American Economic Review* 74:111–124.
- Atkinson, A.B. (1976), "The income tax treatment of charitable contributions", in: R.E. Grieson, ed., *Public and Urban Economics: Essays in Honor of William S. Vickrey* (D.C. Heath, Lexington, MA) pp. 13–29.
- Atwood, D.A. (1990), "Land registration in Africa: the impact on agricultural production", *World Development* 18:659–671.
- Ayres, I., and R. Gertner (1989), "Filling gaps in incomplete contracts: an economic theory of default rules", *Yale Law Journal* 99:87–130.
- Ayres, I., and S.D. Levitt (1998), "Measuring positive externalities from unobservable victim precaution: an empirical analysis of Lojack", *Quarterly Journal of Economics* 113:43–77.
- Baron, J. (1994), *Thinking and Deciding* (Cambridge University Press, New York).
- Barton, J.H. (1972), "The economic basis of damages for breach of contract", *Journal of Legal Studies* 1:277–304.
- Bebchuk, L.A. (1984), "Litigation and settlement under imperfect information", *Rand Journal of Economics* 15:404–415.
- Bebchuk, L.A. (1988), "Suing solely to extract a settlement offer", *Journal of Legal Studies* 17:437–450.
- Bebchuk, L.A. (1996), "A new theory concerning the credibility and success of threats to sue", *Journal of Legal Studies* 25:1–25.
- Bebchuk, L.A., and O. Ben-Shahar (2001), "Precontractual reliance", *Journal of Legal Studies* 30: 423–457.
- Bebchuk, L.A., and H.F. Chang (1996), "An analysis of fee shifting based on the margin of victory: on frivolous suits, meritorious suits, and the role of rule 11", *Journal of Legal Studies* 25:371–403.
- Bebchuk, L.A., and H.F. Chang (1999), "The effect of offer-of-settlement rules on the terms of settlement", *Journal of Legal Studies* 28:489–513.
- Bebchuk, L.A., and L. Kaplow (1992), "Optimal sanctions when individuals are imperfectly informed about the probability of apprehension", *Journal of Legal Studies* 21:365–370.
- Bebchuk, L.A., and I.P.L. Png (1999), "Damage measures for inadvertent breach of contract", *International Review of Law and Economics* 19:319–331.
- Bebchuk, L.A., and S. Shavell (1991), "Information and the scope of liability for breach of contract: the rule of *Hadley v. Baxendale*", *Journal of Law, Economics, & Organization* 7:284–312.

- Beccaria, C. (1872), *An Essay on Crimes and Punishments* (W.C. Little, Albany); originally published 1770.
- Becker, G.S. (1968), "Crime and punishment: an economic approach", *Journal of Political Economy* 76:169–217.
- Becker, G.S., and G.J. Stigler (1974), "Law enforcement, malfeasance, and compensation of enforcers", *Journal of Legal Studies* 3:1–18.
- Bentham, J. (1789), "An introduction to the principles of morals and legislation", in: *The Utilitarians*, 1973 edition (Anchor Books, Garden City, NY) pp. 5–398.
- Bentham, J. (1827), *Rationale of Judicial Evidence*, Vol. 5 (Hunt and Clarke, London).
- Bentham, J. (1830), *The Theory of Legislation*, 1931, edited by C.K. Ogden, translated by R. Hildreth (Kegan Paul and Co., London).
- Bernstein, L. (1992), "Opting out of the legal system: extralegal contractual relations in the diamond industry", *Journal of Legal Studies* 21:115–157.
- Bernstein, L. (1998), "Private commercial law", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 3 (Macmillan, London) pp. 108–114.
- Besen, S.M., and L.J. Raskind (1991), "An introduction to the law and economics of intellectual property", *Journal of Economic Perspectives* 5(1):3–27.
- Besley, T. (1995), "Property rights and investment incentives: theory and evidence from Ghana", *Journal of Political Economy* 103:903–937.
- Besley, T. (1998), "Investment incentives and property rights", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 2 (Macmillan, London) pp. 359–365.
- Biblowit, C. (1991), "International law and the allocation of property rights in common resources", *New York International Law Review* 4:77–85.
- Birmingham, R.L. (1970), "Breach of contract, damage measures, and economic efficiency", *Rutgers Law Review* 24:273–292.
- Bishop, W. (1985), "The choice of remedy for breach of contract", *Journal of Legal Studies* 14:299–320.
- Blackstone, W. (1765–1769), *Commentaries on the Laws of England*; 1992 reprint of 1st edition (William S. Hein and Co., Buffalo).
- Block, M.K., and J.G. Sidak (1980), "The cost of antitrust deterrence: Why not hang a price fixer now and then?", *Georgetown Law Journal* 68:1131–1139.
- Blume, L., D.L. Rubinfeld and P. Shapiro (1984), "The taking of land: when should compensation be paid?" *Quarterly Journal of Economics* 99:71–92.
- Blumenthal, M., and J. Slemrod (1992), "The compliance cost of the U.S. individual income tax system: a second look after tax reform", *National Tax Journal* 45:185–202.
- Blumstein, A., J. Cohen and D. Nagin, eds (1978), *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates* (National Academy of Science, Washington, D.C.).
- Bouckaert, B., and G. De Geest (1995), "Private takings, private taxes, private compulsory services: the economic doctrine of quasi contracts", *International Review of Law and Economics* 15:463–487.
- Bouckaert, B., and G. De Geest, eds (2000), *Encyclopedia of Law and Economics*, Vols. I–V (Edward Elgar, Cheltenham, UK).
- Bovenberg, A.L., and L.H. Goulder (2002), "Environmental taxation and regulation", in: A.J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*, Vol. 3 (North-Holland, Amsterdam) ch. 23, this volume.
- Bowles, R., and N. Garoupa (1997), "Casual police corruption and the economics of crime", *International Review of Law and Economics* 17:75–87.
- Braeutigam, R., B. Owen and J. Panzar (1984), "An economic analysis of alternative fee shifting systems", *Law and Contemporary Problems* 47:173–185.
- Brown, J.P. (1973), "Toward an economic theory of liability", *Journal of Legal Studies* 2:323–349.
- Bundy, S.McG., and E.R. Elhauge (1991), "Do lawyers improve the adversary system? A general theory of litigation advice and its regulation", *California Law Review* 79:313–420.

- Bundy, S.McG., and E.R. Elhauge (1993), "Knowledge about legal sanctions", *Michigan Law Review* 92:261–335.
- Cadot, O. (1987), "Corruption as a gamble", *Journal of Public Economics* 33:223–244.
- Calabresi, G. (1970), *The Costs of Accidents* (Yale University Press, New Haven).
- Calabresi, G. (1975), "Concerning cause and the law of torts", *University of Chicago Law Review* 43:69–108.
- Calabresi, G., and A.D. Melamed (1972), "Property rules, liability rules, and inalienability: one view of the cathedral", *Harvard Law Review* 85:1089–1128.
- Carr-Hill, R.A., and N.H. Stern (1979), *Crime, the Police and Criminal Statistics: An Analysis of Official Statistics for England and Wales Using Econometric Methods* (Academic Press, London).
- Chang, H.F. (1995), "Patent scope, antitrust policy, and cumulative innovation", *Rand Journal of Economics* 26:34–57.
- Charny, D. (1990), "Nonlegal sanctions in commercial relationships", *Harvard Law Review* 104:373–467.
- Che, Y.-K. (1996), "Equilibrium formation of class action suits", *Journal of Public Economics* 62: 339–361.
- Che, Y.-K., and T.-Y. Chung (1999), "Contract damages and cooperative investments", *Rand Journal of Economics* 30:84–105.
- Che, Y.-K., and D.B. Hausch (1999), "Cooperative investments and the value of contracting", *American Economic Review* 89:125–147.
- Che, Y.-K., and J.G. Yi (1993), "The role of precedents in repeated litigation", *Journal of Law, Economics, & Organization* 9:399–424.
- Chu, C.Y.C., S.-C. Hu and T.-Y. Huang (2000), "Punishing repeat offenders more severely", *International Review of Law and Economics* 20:127–140.
- Chung, T.-Y. (1991), "Incomplete contracts, specific investments, and risk sharing", *Review of Economic Studies* 58:1031–1042.
- Chung, T.-Y. (1992), "On the social optimality of liquidated damage clauses: an economic analysis", *Journal of Law, Economics, & Organization* 8:280–305.
- Coase, R.H. (1937), "The nature of the firm", *Economica* 4:386–405.
- Coase, R.H. (1960), "The problem of social cost", *Journal of Law and Economics* 3:1–44.
- Coffee, J.C. (1986), "Understanding the plaintiffs' attorney: the implications of economic theory for private enforcement of law through class and derivative actions", *Columbia Law Review* 86:669–727.
- Cohen, M.A. (1992), "The motives of judges: empirical evidence from antitrust sentencing", *International Review of Law and Economics* 12:13–30.
- Cook, P.J., and G. Zarkin (1985), "Crime and the business cycle", *Journal of Legal Studies* 14:115–128.
- Cook, P.J., S. Molliconi and T.B. Cole (1995), "Regulating gun markets", *Journal of Criminal Law and Criminology* 86:59–92.
- Cooter, R.D. (1985), "Unity in tort, contract, and property: the model of precaution", *California Law Review* 73:1–51.
- Cooter, R.D. (1989), "Punitive damages for deterrence: when and how much?" *Alabama Law Review* 40:1143–1196.
- Cooter, R.D., and L.A. Kornhauser (1980), "Can litigation improve the law without the help of judges?" *Journal of Legal Studies* 9:139–163.
- Cooter, R.D., and D.L. Rubinfeld (1989), "Economic analysis of legal disputes and their resolution", *Journal of Economic Literature* 27:1067–1097.
- Cooter, R.D., and D.L. Rubinfeld (1994), "An economic model of legal discovery", *Journal of Legal Studies* 23:435–463.
- Cooter, R.D., and T. Ulen (1997), *Law and Economics*, 2nd edition (Addison-Wesley, Reading, MA).
- Craswell, R. (1988), "Contract remedies, renegotiation, and the theory of efficient breach", *Southern California Law Review* 61:629–670.
- Craswell, R. (1996), "Offer, acceptance, and efficient reliance", *Stanford Law Review* 48:481–553.

- Craswell, R., and J.E. Calfee (1986), "Deterrence and uncertain legal standards", *Journal of Law, Economics, & Organization* 2:279–303.
- Croson, R., and J.S. Johnston (2000), "Experimental results on bargaining under alternative property rights regimes", *Journal of Law, Economics, & Organization* 16:50–73.
- Dam, K.W. (1975), "Class actions: efficiency, compensation, deterrence, and conflict of interest", *Journal of Legal Studies* 4:47–73.
- Dana, J.D., and K.E. Spier (1993), "Expertise and contingent fees: the role of asymmetric information in attorney compensation", *Journal of Law, Economics, & Organization* 9:349–367.
- Danzon, P.M. (1983), "Contingent fees for personal injury litigation", *Bell Journal of Economics* 14:213–224.
- Danzon, P.M. (1985), *Medical Malpractice: Theory, Evidence, and Public Policy* (Harvard University Press, Cambridge, MA).
- Daughety, A.F. (2000), "Settlement", in: B. Bouckaert and G. De Geest, eds., *Encyclopedia of Law and Economics*, Vol. V (Edward Elgar, Cheltenham, UK) pp. 95–158.
- Daughety, A.F., and J.F. Reinganum (1993), "Endogenous sequencing in models of settlement and litigation", *Journal of Law, Economics, & Organization* 9:314–348.
- Daughety, A.F., and J.F. Reinganum (1994), "Settlement negotiations with two-sided asymmetric information: model duality, information distribution, and efficiency", *International Review of Law and Economics* 14:283–298.
- Daughety, A.F., and J.F. Reinganum (1995), "Keeping society in the dark: on the admissibility of pretrial negotiations as evidence in court", *Rand Journal of Economics* 26:203–221.
- Daughety, A.F., and J.F. Reinganum (2000a), "On the economics of trials: adversarial process, evidence, and equilibrium bias", *Journal of Law, Economics, & Organization* 16:365–394.
- Daughety, A.F., and J.F. Reinganum (2000b), "Appealing judgments", *Rand Journal of Economics* 31:502–525.
- Davis, M.L. (1994), "The value of truth and the optimal standard of proof in legal disputes", *Journal of Law, Economics, & Organization* 10:343–359.
- Demsetz, H. (1967), "Toward a theory of property rights", *American Economic Association Papers and Proceedings* 57:347–359.
- Derfner, M.F., and A.D. Wolf (1995), *Court Awarded Attorney Fees* (Matthew Bender, New York).
- DeVany, A.S., R.D. Eckert, C.J. Meyers, D.J. O'Hara and R.C. Scott (1969), "A property system for market allocation of the electromagnetic spectrum: a legal-economic-engineering study", *Stanford Law Review* 21:1499–1561.
- Devlin, R.A. (1990), "Some welfare implications of no-fault automobile insurance", *International Review of Law and Economics* 10:193–205.
- Dewatripont, M., and J. Tirole (1999), "Advocates", *Journal of Political Economy* 107:1–39.
- Deweese, D., D. Duff and M.J. Trebilcock (1996), *Exploring the Domain of Accident Law: Taking the Facts Seriously* (Oxford University Press, New York).
- Diamond, P.A. (1974), "Single activity accidents", *Journal of Legal Studies* 3:107–164.
- Diamond, P.A. (1997), "Efficiency effects of punitive damages", Working Paper 97-17 (MIT Department of Economics).
- Diamond, P.A., and E. Maskin (1979), "An equilibrium analysis of search and breach of contract, I: steady states", *Bell Journal of Economics* 10:282–316.
- Dilulio, J.J., and A.M. Piehl (1991), "Does prison pay?: The stormy national debate over the cost-effectiveness of imprisonment", *Brookings Review* 9(4):28–35.
- Diver, C.S. (1983), "The optimal precision of administrative rules", *Yale Law Journal* 93:65–109.
- Donohue, J.J., and P. Siegelman (1998), "Allocating resources among prisons and social programs in the battle against crime", *Journal of Legal Studies* 27:1–43.
- Easterbrook, F.H. (1981), "Insider trading, secret agents, evidentiary privileges, and the production of information", *Supreme Court Review* 1981:309–365.

- Easterbrook, F.H., W.M. Landes and R.A. Posner (1980), "Contribution among antitrust defendants: a legal and economic analysis", *Journal of Law and Economics* 23:331–370.
- Eckert, R.D. (1979), *The Enclosure of Ocean Resources: Economics and the Law of the Sea* (Hoover Institution, Stanford).
- Edlin, A.S. (1996), "Cadillac contracts and up-front payments: efficient investment under expectation damages", *Journal of Law, Economics, & Organization* 12:98–118.
- Edlin, A.S., and S. Reichelstein (1996), "Holdups, standard breach remedies, and optimal investment", *American Economic Review* 86:478–501.
- Ehrlich, I. (1973), "Participation in illegitimate activities: a theoretical and empirical investigation", *Journal of Political Economy* 81:521–565.
- Ehrlich, I. (1975), "The deterrent effect of capital punishment: a question of life and death", *American Economic Review* 65:397–417.
- Ehrlich, I., and R.A. Posner (1974), "An economic analysis of legal rulemaking", *Journal of Legal Studies* 3:257–286.
- Eide, E. (1994), *Economics of Crime: Deterrence and the Rational Offender* (Elsevier, New York).
- Eide, E. (2000), "Economics of criminal behavior", in: B. Bouckaert and G. De Geest, eds., *Encyclopedia of Law and Economics*, Vol. V (Edward Elgar, Cheltenham, UK) pp. 345–389.
- Eisenberg, T. (1990), "Testing the selection effect: a new theoretical framework with empirical tests", *Journal of Legal Studies* 19:337–358.
- Eisenberg, T., and H.S. Farber (1997), "The litigious plaintiff hypothesis: case selection and resolution", *Rand Journal of Economics* 28:S92–S112.
- Eisenberg, T., J. Goerdts, B.J. Ostrom, D. Rottman and M.T. Wells (1997), "The predictability of punitive damages", *Journal of Legal Studies* 26:623–661.
- Elder, H.W. (1987), "Property rights structures and criminal courts: an analysis of state criminal courts", *International Review of Law and Economics* 7:21–32.
- Ellickson, R.C. (1989), "A hypothesis of wealth-maximizing norms: evidence from the whaling industry", *Journal of Law, Economics, & Organization* 5:83–97.
- Ellickson, R.C. (1991), *Order Without Law: How Neighbors Settle Disputes* (Harvard University Press, Cambridge, MA).
- Ellickson, R.C. (1993), "Property in land", *Yale Law Journal* 102:1315–1400.
- Emons, W. (1990), "Efficient liability rules for an economy with non-identical individuals", *Journal of Public Economics* 42:89–104.
- Emons, W. (2000), "Expertise, contingent fees, and insufficient attorney effort", *International Review of Law and Economics* 20:21–33.
- Emons, W., and J. Sobel (1991), "On the effectiveness of liability rules when agents are not identical", *Review of Economic Studies* 58:375–390.
- Farber, H.S. (1980), "An analysis of final-offer arbitration", *Journal of Conflict Resolution* 24:683–705.
- Farber, H.S., and M.J. White (1991), "Medical malpractice: an empirical examination of the litigation process", *Rand Journal of Economics* 22:199–217.
- Farmer, A., and P. Pecorino (1996), "Issues of informational asymmetry in legal bargaining", in: D.A. Anderson, ed., *Dispute Resolution: Bridging the Settlement Gap* (JAI Press, Greenwich, CT) pp. 79–105.
- Farnsworth, E.A. (1982), *Contracts* (Little, Brown, Boston).
- Feder, G., and D. Feeny (1991), "Land tenure and property rights: theory and implications for development policy", *World Bank Economic Review* 5:135–153.
- Feess, E. (1999), "Lender liability for environmental harm: an argument against negligence based rules", *European Journal of Law and Economics* 8:231–250.
- Fischel, D.R. (1998), "Lawyers and confidentiality", *University of Chicago Law Review* 65:1–33.
- Fischel, D.R., and A.O. Sykes (1996), "Corporate crime", *Journal of Legal Studies* 25:319–349.
- Fischel, W.A. (1995), *Regulatory Takings: Law, Economics, and Politics* (Harvard University Press, Cambridge, MA).

- Fishman, M.J., and K.M. Hagerty (1990), "The optimal amount of discretion to allow in disclosure", *Quarterly Journal of Economics* 105:427–444.
- Friedman, A.E. (1969), "An analysis of settlement", *Stanford Law Review* 22:67–100.
- Friedman, D.D. (1988), "Does altruism produce efficient outcomes? Marshall versus Kaldor", *Journal of Legal Studies* 17:1–13.
- Friedman, D.D. (1995), "Making sense of English law enforcement in the eighteenth century", *The University of Chicago Law School Roundtable* 2:475–505.
- Friedman, D.D., and W. Sjostrom (1993), "Hanged for a sheep – the economics of marginal deterrence", *Journal of Legal Studies* 22:345–366.
- Friedman, D.D., W.M. Landes and R.A. Posner (1991), "Some economics of trade secret law", *Journal of Economic Perspectives* 5(1):61–72.
- Froeb, L.M. (1993), "The adverse selection of cases for trial", *International Review in Law and Economics* 13:317–324.
- Froeb, L.M., and B.H. Kobayashi (1996), "Naive, biased, yet Bayesian: can juries interpret selectively produced evidence?" *Journal of Law, Economics, & Organization* 12:257–276.
- Fudenberg, D., and J. Tirole (1990), "Moral hazard and renegotiation in agency contracts", *Econometrica* 58:1279–1319.
- Gambetta, D. (1993), *The Sicilian Mafia: The Business of Private Protection* (Harvard University Press, Cambridge, MA).
- Garoupa, N. (1997), "The theory of optimal law enforcement", *Journal of Economic Surveys* 11:267–295.
- Garoupa, N. (1999), "Optimal law enforcement with dissemination of information", *European Journal of Law and Economics* 7:183–196.
- Gilbert, R., and C. Shapiro (1990), "Optimal patent length and breadth", *Rand Journal of Economics* 21:106–112.
- Ginsburg, D.H., and P. Shechtman (1993), "Blackmail: an economic analysis of the law", *University of Pennsylvania Law Review* 141:1849–1876.
- Goetz, C.J., and R.E. Scott (1980), "Enforcing promises: an examination of the basis of contract", *Yale Law Journal* 89:1261–1322.
- Goldberg, V.P. (1974), "The economics of product safety and imperfect information", *Bell Journal of Economics* 5:683–688.
- Gordon, H.S. (1954), "The economic theory of a common property resource: the fishery", *Journal of Political Economy* 62:124–142.
- Gordon, W.J., and R.G. Bone (2000), "Copyright", in: B. Bouckaert and G. De Geest, eds., *Encyclopedia of Law and Economics*, Vol. II (Edward Elgar, Cheltenham, UK) pp. 189–215.
- Gould, J.P. (1973), "The economics of legal conflicts", *Journal of Legal Studies* 2:279–300.
- Grady, M.F. (1983), "A new positive economic theory of negligence", *Yale Law Journal* 92:799–829.
- Gravelle, H.S.E. (1993), "The efficiency implications of cost-shifting rules", *International Review of Law and Economics* 13:3–18.
- Green, J. (1976), "On the optimal structure of liability laws", *Bell Journal of Economics* 7:553–574.
- Green, J., and S. Scotchmer (1995), "On the division of profit in sequential innovation", *Rand Journal of Economics* 26:20–33.
- Greif, A. (1998), "Informal contract enforcement: lessons from medieval trade", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 2 (Macmillan, London) pp. 287–295.
- Grogger, J. (1991), "Certainty vs. severity of punishment", *Economic Inquiry* 29:297–309.
- Grossman, G., and M.L. Katz (1983), "Plea bargaining and social welfare", *American Economic Review* 73:749–757.
- Grossman, S.J. (1981), "The informational role of warranties and private disclosure about product quality", *Journal of Law and Economics* 24:461–483.
- Grossman, S.J., and O.D. Hart (1986), "The costs and benefits of ownership: a theory of vertical and lateral integration", *Journal of Political Economy* 94:691–719.

- Grout, P.A. (1984), "Investment and wages in the absence of binding contracts: a Nash bargaining approach", *Econometrica* 52:449–460.
- Hadfield, G.K. (1994), "Judicial competence and the interpretation of incomplete contracts", *Journal of Legal Studies* 23:159–184.
- Hansmann, H. (1996), *The Ownership of Enterprise* (Harvard University Press, Cambridge, MA).
- Hansmann, H., and R.H. Kraakman (1991), "Toward unlimited shareholder liability for corporate torts", *Yale Law Journal* 100:1879–1934.
- Hardin, G. (1968), "The tragedy of the commons", *Science* 162:1243–1248.
- Hare, R.M. (1981), *Moral Thinking: Its Levels, Method and Point* (Oxford University Press, New York).
- Hart, O.D. (1987), "Incomplete contracts", in: J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave Dictionary of Economics*, Vol. 13 (Macmillan Press, New York) pp. 752–759.
- Hart, O.D. (1989), "An economist's perspective on the theory of the firm", *Columbia Law Review* 89:1757–1774.
- Hart, O.D. (1995), *Firms, Contracts, and Financial Structure* (Oxford University Press, New York).
- Hart, O.D., and B.R. Holmström (1987), "The theory of contracts", in: T.F. Bewley, ed., *Advances in Economic Theory: Fifth World Congress* (Cambridge University Press, New York) pp. 71–155.
- Hart, O.D., and J. Moore (1988), "Incomplete contracts and renegotiation", *Econometrica* 56:755–785.
- Hart, O.D., and J. Moore (1990), "Property rights and the nature of the firm", *Journal of Political Economy* 98:1119–1158.
- Hart, O.D., A. Shleifer and R.W. Vishny (1997), "The proper scope of government: theory and an application to prisons", *Quarterly Journal of Economics* 112:1127–1161.
- Hause, J.C. (1989), "Indemnity, settlement, and litigation, or I'll be suing you", *Journal of Legal Studies* 18:157–179.
- Hay, B.L. (1994), "Civil discovery: its effects and optimal scope", *Journal of Legal Studies* 23:481–515.
- Hay, B.L. (1995), "Effort, information, settlement, trial", *Journal of Legal Studies* 24:29–62.
- Hay, B.L. (1996), "Contingent fees and agency costs", *Journal of Legal Studies* 25:503–533.
- Hay, B.L. (1997), "Optimal contingent fees in a world of settlement", *Journal of Legal Studies* 26:259–278.
- Hay, B.L., and K.E. Spier (1997), "Burdens of proof in civil litigation: an economic perspective", *Journal of Legal Studies* 26:413–431.
- Hay, B.L., and K.E. Spier (1998), "Settlement of litigation", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 3 (Macmillan, London) pp. 442–451.
- Heller, M.A., and R.S. Eisenberg (1998), "Can patents deter innovation? The anticommons in biomedical research", *Science* 280:698–701.
- Hermalin, B.E., and M.L. Katz (1993), "Judicial modification of contracts between sophisticated parties: a more complete view of incomplete contracts and their breach", *Journal of Law, Economics, & Organization* 9:230–255.
- Higgins, R.S. (1978), "Producers' liability and product-related accidents", *Journal of Legal Studies* 7:299–321.
- Higgins, R.S., and P.H. Rubin (1980), "Judicial discretion", *Journal of Legal Studies* 9:129–138.
- Hirshleifer, J. (1971), "The private and social value of information and the reward to inventive activity", *American Economic Review* 61:561–574.
- Hobbes, T. (1651), *Leviathan*, Parts One and Two, 1958 ed. by H.W. Schneider (Liberal Arts Press, New York).
- Hoffman, E., and M. Spitzer (1982), "The Coase theorem: some experimental tests", *Journal of Law and Economics* 25:73–98.
- Holmström, B.R., and J. Tirole (1989), "The theory of the firm", in: R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization*, Vol. 1 (Elsevier, New York) pp. 61–133.
- Hughes, J.W., and E.A. Snyder (1995), "Litigation and settlement under the English and American rules: theory and evidence", *Journal of Law and Economics* 38:225–250.

- Hylton, K.N. (1990), "The influence of litigation costs on deterrence under strict liability and under negligence", *International Review of Law and Economics* 10:161–171.
- Hylton, K.N. (1993), "Asymmetric information and the selection of disputes for litigation", *Journal of Legal Studies* 22:187–210.
- Innes, R. (1999), "Remediation and self-reporting in optimal law enforcement", *Journal of Public Economics* 72:379–393.
- Jensen, M., and W.H. Meckling (1976), "Theory of the firm: managerial behavior, agency costs, and capital structure", *Journal of Financial Economics* 3:305–360.
- Jolls, C. (1997), "Contracts as bilateral commitments: a new perspective on contract modification", *Journal of Legal Studies* 26:203–237.
- Jolls, C. (1998), "Behavioral economic analysis of redistributive legal rules", *Vanderbilt Law Review* 51:1653–1677.
- Jolls, C., C.R. Sunstein and R.H. Thaler (1998), "A behavioral approach to law and economics", *Stanford Law Review* 50:1471–1550.
- Joskow, P.L. (1977), "Commercial impossibility, the uranium market and the Westinghouse case", *Journal of Legal Studies* 6:119–176.
- Kahan, M. (1989), "Causation and incentives to take care under the negligence rule", *Journal of Legal Studies* 18:427–447.
- Kahneman, D., P. Slovic and A. Tversky, eds (1982), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, New York).
- Kahneman, D., J.L. Knetsch and R.H. Thaler (1990), "Experimental tests of the endowment effect and the Coase theorem", *Journal of Political Economy* 98:1325–1348.
- Kakalik, J.S., P.A. Ebener, W.L.F. Felstiner and M.G. Shanley (1983), "Costs of asbestos litigation", Report R-3042-ICJ (Rand Corporation, Santa Monica, CA).
- Kaplow, L. (1984), "The patent–antitrust intersection: a reappraisal", *Harvard Law Review* 97:1813–1892.
- Kaplow, L. (1986a), "An economic analysis of legal transitions", *Harvard Law Review* 99:509–617.
- Kaplow, L. (1986b), "Private versus social costs in bringing suit", *Journal of Legal Studies* 15:371–385.
- Kaplow, L. (1990a), "A note on the optimal use of nonmonetary sanctions", *Journal of Public Economics* 42:245–247.
- Kaplow, L. (1990b), "Optimal deterrence, uninformed individuals, and acquiring information about whether acts are subject to sanctions", *Journal of Law, Economics, & Organization* 6:93–128.
- Kaplow, L. (1992a), "Government relief for risk associated with government action", *Scandinavian Journal of Economics* 94:525–541.
- Kaplow, L. (1992b), "The optimal probability and magnitude of fines for acts that definitely are undesirable", *International Review of Law and Economics* 12:3–11.
- Kaplow, L. (1992c), "Rules versus standards: an economic analysis", *Duke Law Journal* 42:557–629.
- Kaplow, L. (1993), "Shifting plaintiffs' fees versus increasing damage awards", *Rand Journal of Economics* 24:625–630.
- Kaplow, L. (1994a), "The value of accuracy in adjudication: an economic analysis", *Journal of Legal Studies* 23:307–401.
- Kaplow, L. (1994b), "Optimal insurance contracts when establishing the amount of loss is costly", *Geneva Papers on Risk and Insurance Theory* 19:139–152.
- Kaplow, L. (1995a), "A model of the optimal complexity of legal rules", *Journal of Law, Economics, & Organization* 11:150–163.
- Kaplow, L. (1995b), "A note on subsidizing gifts", *Journal of Public Economics* 58:469–477.
- Kaplow, L. (1996), "How tax complexity and enforcement affect the equity and efficiency of the income tax", *National Tax Journal* 49:135–150.
- Kaplow, L. (2000), "General characteristics of rules", in: B. Bouckaert and G. De Geest, eds., *Encyclopedia of Law and Economics*, Vol. V (Edward Elgar, Cheltenham, UK) pp. 502–528.

- Kaplow, L., and S. Shavell (1989), "Legal advice about information to present in litigation: its effects and social desirability", *Harvard Law Review* 102:565–615.
- Kaplow, L., and S. Shavell (1990), "Legal advice about acts already committed", *International Review of Law and Economics* 10:149–159.
- Kaplow, L., and S. Shavell (1992), "Private versus socially optimal provision of ex ante legal advice", *Journal of Law, Economics, & Organization* 8:306–320.
- Kaplow, L., and S. Shavell (1994a), "Accuracy in the determination of liability", *Journal of Law and Economics* 37:1–15.
- Kaplow, L., and S. Shavell (1994b), "Optimal law enforcement with self-reporting of behavior", *Journal of Political Economy* 102:583–606.
- Kaplow, L., and S. Shavell (1994c), "Why the legal system is less efficient than the income tax in redistributing income", *Journal of Legal Studies* 23:667–681.
- Kaplow, L., and S. Shavell (1996a), "Property rules versus liability rules: an economic analysis", *Harvard Law Review* 109:713–790.
- Kaplow, L., and S. Shavell (1996b), "Accuracy in the assessment of damages", *Journal of Law and Economics* 39:191–210.
- Kaplow, L., and S. Shavell (2000), "Should legal rules favor the poor? Clarifying the role of legal rules and the income tax in redistributing income", *Journal of Legal Studies* 29:821–835.
- Kaplow, L., and S. Shavell (2001), "Any non-welfarist method of policy assessment violates the Pareto principle", *Journal of Political Economy* 109:281–286.
- Kaplow, L., and S. Shavell (2002a), *Fairness versus Welfare* (Harvard University Press, Cambridge, MA); also in *Harvard Law Review* 114:961–1388.
- Kaplow, L., and S. Shavell (2002b), "On the superiority of corrective taxes to quantity regulation", *American Law and Economics Review* 4, forthcoming.
- Karpoft, J.M., and J.R. Lott (1999), "On the determinants and importance of punitive damage awards", *Journal of Law and Economics* 42:527–573.
- Katz, A. (1987), "Measuring the demand for litigation: is the English rule really cheaper?" *Journal of Law, Economics, & Organization* 3:143–176.
- Katz, A. (1988), "Judicial decisionmaking and litigation expenditure", *International Review of Law and Economics* 8:127–143.
- Katz, A. (1990a), "The effect of frivolous lawsuits on the settlement of litigation", *International Review of Law and Economics* 10:3–27.
- Katz, A. (1990b), "The strategic structure of offer and acceptance: game theory and the law of contract formation", *Michigan Law Review* 89:215–295.
- Katz, A. (1990c), "Your terms or mine? The duty to read the fine print in contracts", *Rand Journal of Economics* 21:518–537.
- Katz, A. (1993), "Transaction costs and the legal mechanics of exchange: when should silence in the face of an offer be construed as acceptance?" *Journal of Law, Economics, & Organization* 9:77–97.
- Katz, A. (1996), "When should an offer stick? The economics of promissory estoppel in preliminary negotiations", *Yale Law Journal* 105:1249–1309.
- Keeton, W.R., and E. Kwerel (1984), "Externalities in automobile insurance and the underinsured driver problem", *Journal of Law and Economics* 27:149–179.
- Kennan, J., and R. Wilson (1993), "Bargaining with private information", *Journal of Economic Literature* 31:45–104.
- Kessler, D., and S.D. Levitt (1999), "Using sentence enhancements to distinguish between deterrence and incapacitation", *Journal of Law and Economics* 42:343–363.
- Kessler, D., and M. McClellan (1996), "Do doctors practice defensive medicine?" *Quarterly Journal of Economics* 111:353–390.
- Khanna, V.S. (1996), "Corporate criminal liability: what purpose does it serve?" *Harvard Law Review* 109:1477–1534.

- Kimenyi, M.S., W.F. Shughart and R.D. Tollison (1993), "What do judges maximize?" in: C.K. Rowley, ed., *Public Choice Theory*, Vol. 3, *The Separation of Powers and Constitutional Political Economy* (Elgar, Aldershot, UK) pp. 139–146.
- Kitch, E.W. (1977), "The nature and function of the patent system", *Journal of Law and Economics* 20:265–290.
- Klein, B., and K.B. Leffler (1981), "The role of market forces in assuring contractual performance", *Journal of Political Economy* 89:615–641.
- Klein, B., R.G. Crawford and A.A. Alchian (1978), "Vertical integration, appropriable rents, and the competitive contracting process", *Journal of Law and Economics* 21:297–326.
- Klemperer, P. (1990), "How broad should the scope of patent protection be?" *Rand Journal of Economics* 21:113–130.
- Klerman, D. (1996), "Settling multidefendant lawsuits: the advantage of conditional setoff rules", *Journal of Legal Studies* 25:445–462.
- Klevorick, A.K., M. Rothschild and C. Winship (1984), "Information processing and jury decisionmaking", *Journal of Public Economics* 23:245–278.
- Klitgaard, R.E. (1988), *Controlling Corruption* (University of California Press, Berkeley).
- Kobayashi, B.H., and J.R. Lott (1996), "In defense of criminal defense expenditures and plea bargaining", *International Review of Law and Economics* 16:397–416.
- Kofman, F., and J. Lawarree (1993), "Collusion in hierarchical agency", *Econometrica* 61:629–656.
- Kolstad, C.D., T. Ulen and G.V. Johnson (1990), "Ex post liability for harm vs. ex ante safety regulation: substitutes or complements?" *American Economic Review* 80:888–901.
- Kornhauser, L.A. (1982), "An economic analysis of the choice between enterprise and personal liability for accidents", *California Law Review* 70:1345–1392.
- Kornhauser, L.A. (1983), "Reliance, reputation, and breach of contract", *Journal of Law and Economics* 26:691–706.
- Kornhauser, L.A. (1992a), "Modeling collegial courts I: path-dependence", *International Review of Law and Economics* 12:169–185.
- Kornhauser, L.A. (1992b), "Modeling collegial courts II: legal doctrine", *Journal of Law, Economics, & Organization* 8:441–470.
- Kornhauser, L.A., and R.L. Revesz (1994), "Multidefendant settlements: the impact of joint and several liability", *Journal of Legal Studies* 23:41–76.
- Kraakman, R.H. (1984), "Corporate liability strategies and the costs of legal controls", *Yale Law Journal* 93:857–898.
- Kraakman, R.H. (1986), "Gatekeepers: the anatomy of a third-party enforcement strategy", *Journal of Law, Economics, & Organization* 2:53–104.
- Kremer, M. (1998), "Patent buyouts: a mechanism for encouraging innovation", *Quarterly Journal of Economics* 113:1137–1167.
- Kronman, A.T. (1978a), "Mistake, disclosure, information, and the law of contracts", *Journal of Legal Studies* 7:1–34.
- Kronman, A.T. (1978b), "Specific performance", *University of Chicago Law Review* 45:351–382.
- Landes, E.M. (1982), "Insurance, liability, and accidents: a theoretical and empirical investigation of the effect of no-fault accidents", *Journal of Law and Economics* 25:49–65.
- Landes, W.M. (1971), "An economic analysis of the courts", *Journal of Law and Economics* 14:61–107.
- Landes, W.M. (1993), "Sequential versus unitary trials: an economic analysis", *Journal of Legal Studies* 22:99–134.
- Landes, W.M., and R.A. Posner (1975), "The private enforcement of law", *Journal of Legal Studies* 4:1–46.
- Landes, W.M., and R.A. Posner (1976), "Legal precedent: a theoretical and empirical analysis", *Journal of Law and Economics* 19:249–307.
- Landes, W.M., and R.A. Posner (1978), "Salvors, finders, good samaritans, and other rescuers: an economic study of law and altruism", *Journal of Legal Studies* 7:83–128.

- Landes, W.M., and R.A. Posner (1979), "Adjudication as a private good", *Journal of Legal Studies* 8:235–284.
- Landes, W.M., and R.A. Posner (1981), "An economic theory of intentional torts", *International Review of Law and Economics* 1:127–154.
- Landes, W.M., and R.A. Posner (1987a), *The Economic Structure of Tort Law* (Harvard University Press, Cambridge, MA).
- Landes, W.M., and R.A. Posner (1987b), "Trademark law: an economic perspective", *Journal of Law and Economics* 30:265–309.
- Landes, W.M., and R.A. Posner (1989), "An economic analysis of copyright law", *Journal of Legal Studies* 18:325–363.
- Landsberger, M., and I. Meilijson (1982), "Incentive generating state dependent penalty system: the case of income tax evasion", *Journal of Public Economics* 19:333–352.
- Langbein, J.H. (1985), "The German advantage in civil procedure", *University of Chicago Law Review* 52:823–866.
- Leland, H.E. (1992), "Insider trading: should it be prohibited?" *Journal of Political Economy* 100: 859–887.
- Levitt, S.D. (1996), "The effect of prison population size on crime rates: evidence from prison overcrowding litigation", *Quarterly Journal of Economics* 111:319–351.
- Levitt, S.D. (1997a), "Using electoral cycles in police hiring to estimate the effect of police on crime", *American Economic Review* 87:270–290.
- Levitt, S.D. (1997b), "Incentive compatibility constraints as an explanation for the use of prison sentences instead of fines", *International Review of Law and Economics* 17:179–192.
- Levitt, S.D. (1998a), "Juvenile crime and punishment", *Journal of Political Economy* 106:1156–1185.
- Levitt, S.D. (1998b), "Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error?" *Economic Inquiry* 36:353–372.
- Lewis, T., and M. Poitevin (1997), "Disclosure of information in regulatory proceedings", *Journal of Law, Economics, & Organization* 13:50–73.
- Libecap, G.D. (1986), "Property rights in economic history: implications for research", *Explorations in Economic History* 23:227–252.
- Libecap, G.D. (1998), "Common property", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 1 (Macmillan, London) pp. 317–324.
- Loewenstein, G., S. Issacharoff, C. Camerer and L. Babcock (1993), "Self-serving assessments of fairness and pretrial bargaining", *Journal of Legal Studies* 22:135–159.
- Lott, J.R. (1992), "Do we punish high income criminals too heavily?" *Economic Inquiry* 30:583–608.
- Lott, J.R., and D.B. Mustard (1997), "Crime, deterrence, and right-to-carry concealed handguns", *Journal of Legal Studies* 26:1–68.
- Lueck, D. (1998), "First possession", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 2 (Macmillan, London) pp. 132–144.
- Macey, J.R., and G.P. Miller (1991), "The plaintiffs' attorney's role in class action and derivative litigation: economic analysis and recommendations for reform", *University of Chicago Law Review* 58:1–118.
- Machlup, F. (1958), "An economic review of the patent system", Study of the subcommittee on patents, trademarks, and copyrights, Committee on the Judiciary, United States Senate, study no. 15 (United States Government Printing Office, Washington, D.C.).
- MacLeod, W.B., and J.M. Malcomson (1993), "Investments, holdup, and the form of market contracts", *American Economic Review* 83:811–837.
- Magat, W.A., and W.K. Viscusi (1992), *Informational Approaches to Regulation* (MIT Press, Cambridge, MA).
- Malik, A.S. (1990), "Avoidance, screening and optimum enforcement", *Rand Journal of Economics* 21:341–353.
- Malik, A.S. (1993), "Self-reporting and the design of policies for regulating stochastic pollution", *Journal of Environmental Economics and Management* 24:241–257.

- Maskin, E., and J. Moore (1999), "Implementation and renegotiation", *Review of Economic Studies* 66:39–56.
- Mathios, A.D. (2000), "The impact of mandatory disclosure laws on product choices: an analysis of the salad dressing market", *Journal of Law and Economics* 43:651–677.
- McMillan, J. (1994), "Selling spectrum rights", *Journal of Economic Perspectives* 8(3):145–162.
- Menell, P. (1983), "A note on private versus social incentives to sue in a costly legal system", *Journal of Legal Studies* 12:41–52.
- Menell, P. (2000), "Intellectual property: general theories", in: B. Bouckaert and G. De Geest, eds., *Encyclopedia of Law and Economics*, Vol. II (Edward Elgar, Cheltenham, UK) pp. 129–188.
- Meurer, M.J. (1992), "The gains from faith in an unfaithful agent: settlement conflicts between defendants and liability insurers", *Journal of Law, Economics, & Organization* 8:502–522.
- Miceli, T.J. (1993), "Optimal deterrence of nuisance suits by repeat defendants", *International Review of Law and Economics* 13:135–144.
- Miceli, T.J. (1996), "Plea bargaining and deterrence: an institutional approach", *European Journal of Law and Economics* 3:249–264.
- Miceli, T.J. (1997), *Economics of the Law: Torts, Contracts, Property, Litigation* (Oxford University Press, New York).
- Miceli, T.J., and K. Segerson (1996), *Compensation for Regulatory Takings: An Economic Analysis with Applications* (JAI Press, Greenwich, CT).
- Milgrom, P.R. (1981), "Good news and bad news: representation theorems and applications", *Bell Journal of Economics* 12:380–391.
- Milgrom, P.R., and J. Roberts (1986), "Relying on the information of interested parties", *Rand Journal of Economics* 17:18–32.
- Miller, G.P. (1986), "An economic analysis of rule 68", *Journal of Legal Studies* 15:93–125.
- Miller, G.P. (1987), "Some agency problems in settlement", *Journal of Legal Studies* 16:189–215.
- Miller, G.P. (1998), "Class actions", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 1 (Macmillan, London) pp. 257–262.
- Mnookin, R.H. (1993), "Why negotiations fail: an exploration of barriers to the resolution of conflict", *Ohio State Journal of Dispute Resolution* 8:235–256.
- Mnookin, R.H., and R. Wilson (1998), "A model of efficient discovery", *Games and Economic Behavior* 25:219–250.
- Montesquieu, C. (1748), *The Spirit of the Laws*, 1977 reprint edition (University of California Press, Berkeley).
- Mookherjee, D. (1997), "The economics of enforcement", in: A. Bose, M. Rakshit, and A. Sinha, eds., *Issues in Economic Theory and Public Policy: Essays in Honor of Professor Tapas Majumdar* (Oxford University Press, Delhi) pp. 202–249.
- Mookherjee, D., and I.P.L. Png (1992), "Monitoring vis-à-vis investigation in enforcement of law", *American Economic Review* 82:556–565.
- Mookherjee, D., and I.P.L. Png (1994), "Marginal deterrence in enforcement of law", *Journal of Political Economy* 102:1039–1066.
- Mookherjee, D., and I.P.L. Png (1995), "Corruptible law enforcers: how should they be compensated?", *Economic Journal* 105:145–159.
- Moore, M.J., and W.K. Viscusi (1990), *Compensation Mechanisms for Job Risks: Wages, Workers' Compensation, and Product Liability* (Princeton University Press, Princeton).
- Murphy, R.S., and E.A. O'Hara (1997), "Mistake of federal criminal law: a study of coalitions and costly information", *Supreme Court Economic Review* 5:217–278.
- Nagin, D. (1978), "General deterrence: a review of the empirical evidence", in: A. Blumstein, J. Cohen and D. Nagin, eds., *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates* (National Academy of Sciences, Washington, D.C.) pp. 95–139.
- Netter, J.M. (1998), "Adverse possession", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 1 (Macmillan, London) pp. 18–21.

- Newman, H.A., and D.W. Wright (1990), "Strict liability in a principal-agent model", *International Review of Law and Economics* 10:219–231.
- Newman, P., ed. (1998), *New Palgrave Dictionary of Economics and the Law*, Vols. 1–3 (Macmillan, London).
- Noldeke, G., and K.M. Schmidt (1995), "Option contracts and renegotiation: a solution to the holdup problem", *Rand Journal of Economics* 26:163–179.
- Ordover, J.A. (1978), "Costly litigation in the model of single activity accidents", *Journal of Legal Studies* 7:243–261.
- Osborne, E. (1999), "Who should be worried about asymmetric information in litigation?" *International Review of Law and Economics* 19:399–409.
- Ostrom, B.J., and N.B. Kauder (1996), *Examining the Work of State Courts 1994* (National Center for State Courts, Williamsburg, VA).
- Ostrom, E. (1998), "Self-governance of common-pool resources", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 3 (Macmillan, London) pp. 424–433.
- Pigou, A.C. (1932), *The Economics of Welfare* (Macmillan, New York).
- Pitchford, R. (1995), "How liable should a lender be? The case of judgment-proof firms and environmental risk", *American Economic Review* 85:1171–1186.
- Png, I.P.L. (1986), "Optimal subsidies and damages in the presence of judicial error", *International Review of Law and Economics* 6:101–105.
- Polinsky, A.M. (1980a), "Private versus public enforcement of fines", *Journal of Legal Studies* 9:105–127.
- Polinsky, A.M. (1980b), "On the choice between property rules and liability rules", *Economic Inquiry* 18:233–246.
- Polinsky, A.M. (1983), "Risk sharing through breach of contract remedies", *Journal of Legal Studies* 12:427–444.
- Polinsky, A.M. (1989), *An Introduction to Law and Economics*, 2nd edition (Little, Brown and Company, Boston).
- Polinsky, A.M. (1997), "Are punitive damages really insignificant, predictable, and rational?" *Journal of Legal Studies* 26:663–677.
- Polinsky, A.M., and Y.-K. Che (1991), "Decoupling liability: optimal incentives for care and litigation", *Rand Journal of Economics* 22:562–570.
- Polinsky, A.M., and D.L. Rubinfeld (1988a), "The welfare implications of costly litigation for the level of liability", *Journal of Legal Studies*, 17:151–164.
- Polinsky, A.M., and D.L. Rubinfeld (1988b), "The deterrent effects of settlements and trials", *International Review of Law and Economics* 8:109–116.
- Polinsky, A.M., and D.L. Rubinfeld (1991), "A model of optimal fines for repeat offenders", *Journal of Public Economics* 46:291–306.
- Polinsky, A.M., and D.L. Rubinfeld (1993), "Sanctioning frivolous suits: an economic analysis", *Georgetown Law Journal* 82:397–435.
- Polinsky, A.M., and D.L. Rubinfeld (1996), "Optimal awards and penalties when the probability of prevailing varies among plaintiffs", *Rand Journal of Economics* 27:269–280.
- Polinsky, A.M., and S. Shavell (1979), "The optimal tradeoff between the probability and magnitude of fines", *American Economic Review* 69:880–891.
- Polinsky, A.M., and S. Shavell (1981), "Contribution and claim reduction among antitrust defendants: an economic analysis", *Stanford Law Review* 33:447–471.
- Polinsky, A.M., and S. Shavell (1984), "The optimal use of fines and imprisonment", *Journal of Public Economics* 24:89–99.
- Polinsky, A.M., and S. Shavell (1992), "Enforcement costs and the optimal magnitude and probability of fines", *Journal of Law and Economics* 35:133–148.
- Polinsky, A.M., and S. Shavell (1993), "Should employees be subject to fines and imprisonment given existence of corporate liability?" *International Review of Law and Economics* 13:239–257.

- Polinsky, A.M., and S. Shavell (1998a), "Punitive damages: an economic analysis", *Harvard Law Review* 111:869–962.
- Polinsky, A.M., and S. Shavell (1998b), "On offense history and the theory of deterrence", *International Review of Law and Economics* 18:305–324.
- Polinsky, A.M., and S. Shavell (1999), "On the disutility and discounting of imprisonment and the theory of deterrence", *Journal of Legal Studies* 28:1–16.
- Polinsky, A.M., and S. Shavell (2000), "The economic theory of public enforcement of law", *Journal of Economic Literature* 38:45–76.
- Polinsky, A.M., and S. Shavell (2001), "Corruption and optimal law enforcement", *Journal of Public Economics* 81:1–24.
- Posner, E.A. (1995), "Contract law in the welfare state: a defense of the unconscionability doctrine, usury laws, and related limitations on the freedom to contract", *Journal of Legal Studies* 24:283–319.
- Posner, E.A. (1997), "Altruism, status, and trust in the law of gifts and gratuitous promises", *Wisconsin Law Review* 1997:567–609.
- Posner, R.A. (1972), *Economic Analysis of Law*, 1st edition (Little, Brown and Company, Boston).
- Posner, R.A. (1973), "An economic approach to legal procedure and judicial administration", *Journal of Legal Studies* 2:399–458.
- Posner, R.A. (1975), "The economic approach to law", *Texas Law Review* 53:757–782.
- Posner, R.A. (1977a), *Economic Analysis of Law*, 2nd edition (Little, Brown and Company, Boston).
- Posner, R.A. (1977b), "Gratuitous promises in economics and law", *Journal of Legal Studies* 6:411–426.
- Posner, R.A. (1985), "An economic theory of the criminal law", *Columbia Law Review* 85:1193–1231.
- Posner, R.A. (1986), "The summary jury trial and other methods of alternative dispute resolution: some cautionary observations", *University of Chicago Law Review* 53:366–393.
- Posner, R.A. (1993a), "What do judges and justices maximize? (The same thing everybody else does)", *Supreme Court Economic Review* 3:1–41.
- Posner, R.A. (1993b), "Blackmail, privacy, and freedom of contract", *University of Pennsylvania Law Review* 141:1817–1847.
- Posner, R.A. (1998), *Economic Analysis of Law*, 5th edition (Aspen Law and Business, New York).
- Posner, R.A. (1999), "An economic approach to the law of evidence", *Stanford Law Review* 51:1477–1546.
- Posner, R.A., and A.M. Rosenfield (1977), "Impossibility and related doctrines in contract law: an economic analysis", *Journal of Legal Studies* 6:83–118.
- Pound, R. (1959), *An Introduction to the Philosophy of Law* (Yale University Press, New Haven).
- Priest, G.L. (1977), "The common law process and the selection of efficient rules", *Journal of Legal Studies* 6:65–82.
- Priest, G.L. (1981), "A theory of the consumer product warranty", *Yale Law Journal* 90:1297–1352.
- Priest, G.L. (1988), "Products liability law and the accident rate", in: R.E. Litan and C. Winston, eds., *Liability: Perspectives and Policy* (Brookings Institution, Washington, D.C.) pp. 184–222.
- Priest, G.L., and B. Klein (1984), "The selection of disputes for litigation", *Journal of Legal Studies* 13:1–55.
- Pyle, D.J. (1983), *The Economics of Crime and Law Enforcement* (Macmillan, London).
- Rabin, M. (1998), "Psychology and economics", *Journal of Economic Literature* 36:11–46.
- Raiffa, H. (1968), *Decision Analysis* (Addison-Wesley, Reading, MA).
- Ramseyer, J.M. (1998), "Judicial independence", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 2 (Macmillan, London) pp. 383–387.
- Ramseyer, J.M., and M. Nakazato (1989), "The rational litigant: settlement amounts and verdict rates in Japan", *Journal of Legal Studies* 18:263–290.
- Rasmusen, E. (1994), "Judicial legitimacy as a repeated game", *Journal of Law, Economics, & Organization* 10:63–83.
- Reinganum, J.F. (1988), "Plea bargaining and prosecutorial discretion", *American Economic Review* 78:713–728.

- Reinganum, J.F. (1989), "The timing of innovation: research, development, and diffusion", in: R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization*, Vol. 1 (North-Holland, Amsterdam) pp. 849–908.
- Reinganum, J.F. (2000), "Sentencing guidelines, judicial discretion, and plea bargaining", *Rand Journal of Economics* 31:62–81.
- Reinganum, J.F., and L.L. Wilde (1986), "Settlement, litigation, and the allocation of litigation costs", *Rand Journal of Economics* 17:557–566.
- Roberts, M.J., and M. Spence (1976), "Effluent charges and licenses under uncertainty", *Journal of Public Economics* 5:193–208.
- Rogerson, W.P. (1984), "Efficient reliance and damage measures for breach of contract", *Rand Journal of Economics* 15:39–53.
- Rogerson, W.P. (1992), "Contractual solutions to the hold-up problem", *Review of Economic Studies* 59:777–794.
- Rose-Ackerman, S. (1978), *Corruption: A Study in Political Economy* (Academic Press, New York).
- Rose-Ackerman, S. (1985), "Inalienability and the theory of property rights", *Columbia Law Review* 85:931–969.
- Rose-Ackerman, S. (1999), *Corruption and Government: Causes, Consequences and Reform* (Cambridge University Press, New York).
- Rose-Ackerman, S., and M. Geistfeld (1987), "The divergence between social and private incentives to sue: a comment on Shavell, Menell and Kaplow", *Journal of Legal Studies* 16:483–491.
- Rosenberg, D. (1984), "The causal connection in mass exposure cases: a 'public law' vision of the tort system", *Harvard Law Review* 97:849–929.
- Rosenberg, D., and S. Shavell (1985), "A model in which suits are brought for their nuisance value", *International Review of Law and Economics* 5:3–13.
- Rubin, P.H. (1977), "Why is the common law efficient?" *Journal of Legal Studies* 6:51–63.
- Rubin, P.H. (1993), *Tort Reform by Contract* (AEI Press, Washington, D.C.).
- Rubinfeld, D.L., and D.E.M. Sappington (1987), "Efficient awards and standards of proof in judicial proceedings", *Rand Journal of Economics* 18:308–315.
- Rubinfeld, D.L., and S. Scotchmer (1993), "Contingent fees for attorneys: an economic analysis", *Rand Journal of Economics* 24:343–356.
- Rubinfeld, D.L., and S. Scotchmer (1998), "Contingent fees", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 1 (Macmillan, London) pp. 415–420.
- Rubinstein, A. (1979), "An optimal conviction policy for offenses that may have been committed by accident", in: S.J. Brams, A. Schotter and G. Schwodiauer, eds., *Applied Game Theory* (Physica-Verlag, Wurzburg) pp. 406–413.
- Sah, R.K. (1991), "Social osmosis and patterns of crime", *Journal of Political Economy* 99:1272–1295.
- Sanchirico, C.W. (2000), "Taxes versus legal rules as instruments for equity: a more equitable view", *Journal of Legal Studies* 29:797–820.
- Scherer, F.M., and D. Ross (1990), *Industrial Market Structure and Economic Performance*, 3rd edition (Houghton Mifflin, Boston).
- Schmitz, P.W. (2000), "On the joint use of liability and safety regulation", *International Review of Law and Economics* 20:371–382.
- Schrag, J.L. (1999), "Managerial judges: an economic analysis of the judicial management of legal discovery", *Rand Journal of Economics* 30:305–323.
- Schrag, J.L., and S. Scotchmer (1994), "Crime and prejudice: the use of character evidence in criminal trials", *Journal of Law, Economics, & Organization* 10:319–342.
- Schwartz, A. (1979), "The case for specific performance", *Yale Law Journal* 89:271–306.
- Schwartz, A. (1992), "Relational contracts in the courts: an analysis of incomplete agreements and judicial strategies", *Journal of Legal Studies* 21:271–318.
- Schwartz, A., and L.L. Wilde (1979), "Intervening in markets on the basis of imperfect information: a legal and economic analysis", *University of Pennsylvania Law Review* 127:630–682.

- Schwartz, E.P., and W.F. Schwartz (1996), "The challenge of peremptory challenges", *Journal of Law, Economics, & Organization* 12:325–360.
- Schweizer, U. (1989), "Litigation and settlement under two-sided incomplete information", *Review of Economic Studies* 56:163–178.
- Scotchmer, S. (1996), "Protecting early innovators: should second-generation products be patentable?" *Rand Journal of Economics* 27:322–331.
- Scotchmer, S. (1999), "On the optimality of the patent renewal system", *Rand Journal of Economics* 30:181–196.
- Scott, K.E. (1998), "Insider trading", in: P. Newman, ed., *The New Palgrave Dictionary of Economics and the Law*, Vol. 2 (Macmillan, London) pp. 326–330.
- Segal, I., and M.D. Whinston (2001), "The Mirrlees approach to mechanism design with renegotiation (with applications to hold-up and risk sharing)", *Econometrica*, forthcoming.
- Sen, A., and B. Williams (1982), *Utilitarianism and Beyond* (Cambridge University Press, Cambridge, UK).
- Sethi, R., and E. Somanathan (1996), "The evolution of social norms in common property resource use", *American Economic Review* 86:766–788.
- Shavell, S. (1980a), "An analysis of causation and the scope of liability in the law of torts", *Journal of Legal Studies* 9:463–516.
- Shavell, S. (1980b), "Damage measures for breach of contract", *Bell Journal of Economics* 11:466–490.
- Shavell, S. (1980c), "Strict liability versus negligence", *Journal of Legal Studies* 9:1–25.
- Shavell, S. (1981), "A note on efficiency vs. distributional equity in legal rulemaking: should distributional equity matter given optimal income taxation?" *American Economic Association Papers and Proceedings* 71:414–418.
- Shavell, S. (1982a), "On liability and insurance", *Bell Journal of Economics* 13:120–132.
- Shavell, S. (1982b), "The social versus the private incentive to bring suit in a costly legal system", *Journal of Legal Studies* 11:333–339.
- Shavell, S. (1982c), "Suit, settlement, and trial: a theoretical analysis under alternative methods for the allocation of legal costs", *Journal of Legal Studies* 11:55–81.
- Shavell, S. (1984a), "A model of the optimal use of liability and safety regulation", *Rand Journal of Economics* 15:271–280.
- Shavell, S. (1984b), "The design of contracts and remedies for breach", *Quarterly Journal of Economics* 99:121–148.
- Shavell, S. (1984c), "Liability for harm versus regulation of safety", *Journal of Legal Studies* 13:357–374.
- Shavell, S. (1985a), "Criminal law and the optimal use of nonmonetary sanctions as a deterrent", *Columbia Law Review* 85:1232–1262.
- Shavell, S. (1985b), "Uncertainty over causation and the determination of civil liability", *Journal of Law and Economics* 28:587–609.
- Shavell, S. (1986), "The judgment proof problem", *International Review of Law and Economics* 6:45–58.
- Shavell, S. (1987a), *Economic Analysis of Accident Law* (Harvard University Press, Cambridge, MA).
- Shavell, S. (1987b), "The optimal use of nonmonetary sanctions as a deterrent", *American Economic Review* 77:584–592.
- Shavell, S. (1987c), "A model of optimal incapacitation", *American Economic Review* 77:107–110.
- Shavell, S. (1988), "Legal advice about contemplated acts: the decision to obtain advice, its social desirability, and protection of confidentiality", *Journal of Legal Studies* 17:123–150.
- Shavell, S. (1989a), "Sharing of information prior to settlement or litigation", *Rand Journal of Economics* 20:183–195.
- Shavell, S. (1989b), "Optimal sanctions and the incentive to provide evidence to legal tribunals", *International Review of Law and Economics* 9:3–11.
- Shavell, S. (1990), "Deterrence and the punishment of attempts", *Journal of Legal Studies* 19:435–466.
- Shavell, S. (1991a), "An economic analysis of altruism and deferred gifts", *Journal of Legal Studies* 20:401–421.

- Shavell, S. (1991b), "Specific versus general enforcement of law", *Journal of Political Economy* 99:1088–1108.
- Shavell, S. (1991c), "Individual precautions to prevent theft: private versus socially optimal behavior", *International Review of Law and Economics* 11:123–132.
- Shavell, S. (1992), "A note on marginal deterrence", *International Review of Law and Economics* 12:345–355.
- Shavell, S. (1993a), "The optimal structure of law enforcement", *Journal of Law and Economics* 36:255–287.
- Shavell, S. (1993b), "Suit versus settlement when parties seek nonmonetary judgments", *Journal of Legal Studies* 22:1–13.
- Shavell, S. (1993c), "An economic analysis of threats and their illegality: blackmail, extortion, and robbery", *University of Pennsylvania Law Review* 141:1877–1903.
- Shavell, S. (1994), "Acquisition and disclosure of information prior to sale", *Rand Journal of Economics* 25:20–36.
- Shavell, S. (1995a), "Alternative dispute resolution: an economic analysis", *Journal of Legal Studies* 24:1–28.
- Shavell, S. (1995b), "The appeals process as a means of error correction", *Journal of Legal Studies* 24:379–426.
- Shavell, S. (1996), "Any frequency of plaintiff victory at trial is possible", *Journal of Legal Studies* 25:493–501.
- Shavell, S. (1997), "The fundamental divergence between the private and the social motive to use the legal system", *Journal of Legal Studies* 26:575–612.
- Shavell, S. (1999), "The level of litigation: private versus social optimality", *International Review of Law and Economics* 19:99–115.
- Shavell, S. (2000), "On the social function and the regulation of liability insurance", *Geneva Papers on Risk and Insurance, Issues and Practice* 25:166–179.
- Shavell, S. (2002), *Principles of Economic Analysis of Law* (Harvard University Press, Cambridge, MA) forthcoming.
- Shavell, S., and T. van Ypersele (2001), "Rewards versus intellectual property rights", *Journal of Law and Economics* 44:525–547.
- Shepherd, G.B. (1999), "An empirical study of the economics of pretrial discovery", *International Review of Law and Economics* 19:245–263.
- Shin, H.S. (1998), "Adversarial and inquisitorial procedures in arbitration", *Rand Journal of Economics* 29:378–405.
- Shleifer, A. (1998), "State versus private ownership", *Journal of Economic Perspectives* 12(4):133–150.
- Shleifer, A., and R.W. Vishny (1993), "Corruption", *Quarterly Journal of Economics* 108:599–617.
- Sieg, H. (2000), "Estimating a bargaining model with asymmetric information: evidence from medical malpractice disputes", *Journal of Political Economy* 108:1006–1021.
- Silver, C. (2000), "Class actions-representative proceedings", in: B. Bouckaert and G. De Geest, eds., *Encyclopedia of Law and Economics*, Vol. V (Edward Elgar, Cheltenham, UK) pp. 194–240.
- Sloan, F.A., B.A. Reilly and C.M. Schenzler (1994), "Tort liability versus other approaches for deterring careless driving", *International Review of Law and Economics* 14:53–71.
- Sobel, J. (1985), "Disclosure of evidence and resolution of disputes: who should bear the burden of proof?" in: A.E. Roth, ed., *Game-Theoretic Models of Bargaining* (Cambridge University Press, Cambridge, UK) pp. 341–361.
- Sobel, J. (1989), "An analysis of discovery rules", *Law and Contemporary Problems* 52:133–159.
- Spence, M. (1977), "Consumer misperceptions, product failure, and product liability", *Review of Economic Studies* 44:561–572.
- Spier, K.E. (1992a), "The dynamics of pretrial negotiation", *Review of Economic Studies* 59:93–108.
- Spier, K.E. (1992b), "Incomplete contracts and signalling", *Rand Journal of Economics* 23:432–443.

- Spier, K.E. (1994a), "Pretrial bargaining and the design of fee-shifting rules", *Rand Journal of Economics* 25:197–214.
- Spier, K.E. (1994b), "Settlement bargaining and the design of damage awards", *Journal of Law, Economics, & Organization* 10:84–95.
- Spier, K.E. (1997), "A note on the divergence between the private and the social motive to settle under a negligence rule", *Journal of Legal Studies* 26:613–621.
- Spier, K.E., and M.D. Whinston (1995), "On the efficiency of privately stipulated damages for breach of contract: entry barriers, reliance, and renegotiation", *Rand Journal of Economics* 26:180–202.
- Spitzer, M., and E. Talley (2000), "Judicial auditing", *Journal of Legal Studies* 29:649–683.
- Stigler, G.J. (1970), "The optimum enforcement of laws", *Journal of Political Economy* 78:526–536.
- Stole, L.A. (1992), "The economics of liquidated damages clauses in contractual environments with private information", *Journal of Law, Economics, & Organization* 8:582–606.
- Sykes, A.O. (1981), "An efficiency analysis of vicarious liability under the law of agency", *Yale Law Journal* 91:168–206.
- Sykes, A.O. (1990), "The doctrine of commercial impracticability in a second-best world", *Journal of Legal Studies* 19:43–94.
- Sykes, A.O. (1994), "'Bad faith' refusal to settle by liability insurers: some implications of the judgment-proof problem", *Journal of Legal Studies* 23:77–110.
- Tauchen, H.V., A.D. Witte and H. Griesinger (1994), "Criminal deterrence: revisiting the issue with a birth cohort", *Review of Economics and Statistics* 76:399–412.
- Tirole, J. (1986), "Hierarchies and bureaucracies: on the role of collusion in organizations", *Journal of Law, Economics, & Organization* 2:181–214.
- Tirole, J. (1988), *The Theory of Industrial Organization* (MIT Press, Cambridge, MA).
- Trebilcock, M.J. (1993), *The Limits of Freedom of Contract* (Harvard University Press, Cambridge, MA).
- Ulen, T. (1984), "The efficiency of specific performance: toward a unified theory of contract remedies", *Michigan Law Review* 83:341–403.
- Umbeck, J.R. (1981), *A Theory of Property Rights with Application to the California Gold Rush* (Iowa State University Press, Ames).
- van Wijck, P., and B. van Velthoven (2000), "An economic analysis of the American and the continental rule for allocating legal costs", *European Journal of Law and Economics* 9:115–125.
- Viscusi, W.K. (1986a), "The determinants of the disposition of product liability claims and compensation for bodily injury", *Journal of Legal Studies* 15:321–346.
- Viscusi, W.K. (1986b), "The risks and rewards of criminal activity: a comprehensive test of criminal deterrence, Part I", *Journal of Labor Economics* 4:317–340.
- Viscusi, W.K. (1988), "Product liability litigation with risk aversion", *Journal of Legal Studies* 17: 101–121.
- Viscusi, W.K. (1991), *Reforming Products Liability* (Harvard University Press, Cambridge, MA).
- Viscusi, W.K., and W. Evans (1990), "Utility functions that depend on health status: estimates and economic implications", *American Economic Review* 80:353–374.
- Viscusi, W.K., and W.A. Magat (1987), *Learning About Risk: Consumer and Worker Responses to Hazard Information* (Harvard University Press, Cambridge, MA).
- Viscusi, W.K., J.M. Vernon and J.E. Harrington Jr (1995), *Economics of Regulation and Antitrust*, 2nd edition (MIT Press, Cambridge, MA).
- Waldfoegel, J. (1995a), "Are fines and prison terms used efficiently? Evidence on federal fraud offenders", *Journal of Law and Economics* 38:107–139.
- Waldfoegel, J. (1995b), "The selection hypothesis and the relationship between trial and plaintiff victory", *Journal of Political Economy* 103:229–260.
- Weitzman, M.L. (1974), "Prices vs. quantities", *Review of Economic Studies* 41:477–491.
- Wilde, L.L. (1992), "Criminal choice, nonmonetary sanctions, and marginal deterrence: a normative analysis", *International Review of Law and Economics* 12:333–344.

- Wilkins, D.B. (1992), "Who should regulate lawyers?" *Harvard Law Review* 105:799–887.
- Williamson, O.E. (1975), *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization* (Free Press, New York).
- Williamson, O.E. (1985), *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting* (Free Press, New York).
- Wils, W.P.J. (1993), "Who should bear the costs of failed negotiations? A functional inquiry into precontractual liability", *Journal des Economistes et des Etudes Humaines* 4:93–134.
- Wilson, J.Q., and R.J. Herrnstein (1985), *Crime and Human Nature* (Simon and Schuster, New York).
- Witte, A.D. (1980), "Estimating the economic model of crime with individual data", *The Quarterly Journal of Economics* 94:57–84.
- Wittman, D. (1977), "Prior regulation versus post liability: the choice between input and output monitoring", *Journal of Legal Studies* 6:193–212.
- Wittman, D. (1981), "Optimal pricing of sequential inputs: last clear chance, mitigation of damages, and related doctrines in the law", *Journal of Legal Studies* 10:65–91.
- Wittman, D. (1985), "Is the selection of cases for trial biased?" *Journal of Legal Studies* 14:185–214.
- Wright, B.D. (1983), "The economics of invention incentives: patents, prizes, and research contracts", *American Economic Review* 73:691–707.