1992

The Internalization Paradox and Workers' Compensation

Keith N. Hylton
Boston University School of Law

THE INTERNALIZATION PARADOX AND WORKERS' COMPENSATION

Keith N. Hylton[*]
Steven E. Laymon[**]

TABLE OF CONTENTS

---

* Associate Professor, Northwestern University School of Law.
** Ph.D. candidate, University of Chicago, Political Science Department. This Article is the result of an ongoing effort to gather data on the incentive effects of the tort system.
We thank the American Bar Foundation for financial support.

I. INTRODUCTION

By providing a scientific link between the compensatory and deterrence goals of tort law, the Pigovian theory of externalities[1] has had an enormous influence on modern torts scholarship and tort doctrine.[2]

The theoretical framework and the accompanying goal of internalizing external costs provide a solution to a question that early torts theorists were unequipped to answer: How could a liability rule designed to compensate tort victims provide the proper amount of deterrence? As simple as this question may now seem,[3] it would have been hard to answer rigorously before the externality literature appeared, and one searches in vain for an answer in the early torts theorists such as Holmes.[4] Further, externality theory provides a reason for distinguishing torts and crimes, because it suggests that criminal penalties should be used only in areas where compensatory damages are unlikely to bring about the desired amount of deterrence.[5]

---

1. The literature is enormous, but externality theory is generally traced to Arthur Pigou. *See* A.C. PIGOU, THE ECONOMICS OF WELFARE (4th ed. 1952). For some early contributions, see Francis M. Bator, *The Anatomy of Market Failure*, 72 Q.J. ECON. 351 (1958); James M. Buchanan & W. Craig Stubblebine, *Externality*, 29 ECONOMICA 371 (1962).

The reader may wonder why we refer to the literature as the "Pigovian theory of externalities" rather than simply "theory of externalities." We use the word Pigovian because we want to refer to the state of the externality literature before the contribution of Ronald Coase, *see infra* note 6.

2. The influence is apparent. *See* GUIDO CALABRESI, THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS (1970); *see also* WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW (1987).

3. We say it seems simple because the link between compensation and the appropriate level of deterrence is, at least in simple cases, a settled area of the economic theory of tort law (and in complicated cases one should find agreement on the theoretical approach). For a non-technical, introductory discussion of damages and the appropriate level of deterrence or care-taking, see RICHARD A. POSNER, TORT LAW: CASES AND ECONOMIC ANALYSIS 1-9 (1982).

4. Holmes has little to say in *The Common Law* about the link between full compensation and the appropriate level of deterrence or care-taking, probably because it was not a question that he was interested in answering. In addition, he was writing well before Learned Hand's opinion stating the Hand formula, and before the externality literature in economics. However, the following passage suggests that the intuition underlying the Hand formula and the externality literature was guiding some of his conclusions:

> One who diminishes the value of property by intentional damage knows it belongs to somebody. If he thinks it belongs to himself, he expects whatever harm he may do to come out of his own pocket. It would be odd if he were to get rid of the burden by discovering that it belonged to his neighbor.

OLIVER W. HOLMES, THE COMMON LAW 97 (1881).

5. Externality theory is used to describe the boundary between private and public

In spite of its important contributions, the theory of externalities was criticized by Coase in his article on social cost.[6] The problem for Coase was that the theory, as then understood by many lawyers and economists, suggested that there was no need for a private liability system. The state could control undesirable behavior by setting appropriate taxes that could take the place of liability rules. This was a naive solution to Coase because it ignored the importance of incentives. A private liability system took advantage of individual incentives in a way that a centralized system could not. In the extreme case in which the market operated without friction—the zero transaction costs case—the role of the state could be reduced to the point where it did not even define legal entitlements, it only enforced valid contracts. Parties would set appropriate entitlements and arrange wealth-maximizing contracts between themselves for sharing risks. A system of Pigovian taxes would not necessarily improve upon this and could be considerably worse. Under Coase's theory, the role and significance of the state could not be defined by simply discovering the sources of externality in an economy.[7]

This Article presents an alternative critique of externality theory as a source of normative principles for tort law. We refer to our critique as the "internalization paradox." The argument runs as follows. The Pigovian goal is to internalize all external costs. The simplest way to do this is to require the injurer to compensate the victim. But this is vulnerable to fraudulent claims. How, for example, is the injurer to know whether or not the victim is exaggerating the extent of injury? To turn away fraudulent claims, some administrative mechanism must be devised to determine the validity of claims, espe-

---

enforcement in William M. Landes & Richard A. Posner, *The Private Enforcement of Law*, 4 J. LEGAL STUD. 1 (1975).

6. R.H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1 (1960). The article is known for presenting the "Coase theorem," which states (roughly) that in the absence of obstacles to fully-informed negotiations (transaction costs), bargaining will lead to wealth-maximizing agreements. However, as Coase made clear in the introduction of the article, his primary aim was to present a critique of the theory of externalities.

7. It should be noted that since Adam Smith's discussion of the role of the state in *The Wealth of Nations*, economists generally have accepted the notion that the state may be necessary to provide certain goods (public or collective goods) such as defense, that would otherwise not be provided, or not be provided in sufficient quantity, by the market. The troublesome part of the theory of externalities, as it was presented by Pigou, is that it could be read as (and indeed was read as) providing a justification for an expansive regulatory machinery. Coase's article was not an argument against such a state, but an attempt to discover the assumptions underlying the claim that a highly interventionist state was necessary in order to control the problem of externalities.

cially in cases of unclear causation and novel types of injury. Society has established courts and arbitration procedures for this purpose. But the administrative process is itself costly and subject to error, and this means that some external costs cannot be internalized. The cost to the plaintiff of invoking the administrative machinery generally cannot be internalized.[8] Yet these costs flow from the injurious event as surely as the injury itself. Further, almost any attempt to reduce error by improving the evaluation procedure increases the administrative costs of the system, squeezing out a larger share of potential claims. Efforts to solve the problem of fraudulent claims run head on into the primary reason for setting up a liability system, and the contradiction means that a liability system cannot fulfill its function of internalization.

The basic proposition of this Article is that the full internalization principle is an impossible goal. The friction that is required by the need to determine the validity of claims prevents full internalization.[9] External costs can be partially internalized and improvements in the operational efficiency of a liability system can be sought in this area, but that is as far as we can go. There are many obstacles to full internalization in the tort system, such as tax and transfer systems, and imperfect insurance pricing.[10] The theoretical literature has virtually ignored perhaps the most important obstacle: the litigation process. Efforts to reform the tort system by reforming the litigation process may have the unintended consequence of exacerbating the externalization problem.

---

8. Assume a simple fee shifting rule: all legal costs of successful plaintiffs are shifted to the defendant. This internalizes the legal costs of successful plaintiffs, but what about unsuccessful plaintiffs whose claims are valid? The opposite extreme, shifting all costs to the defendant regardless of the plaintiff's success, would never be adopted as a rule because it opens the door wide to fraudulent claims. Intermediate cost-shifting rules are subject to the same criticisms.

9. While we are not aware of any published works that make the general argument of this Article, there are more detailed studies of the litigation process that suggest a similar proposition. An example is PHILIPPE NONET, ADMINISTRATIVE JUSTICE: ADVOCACY AND CHANGE IN A GOVERNMENT AGENCY (1969), which studies the workers' compensation system in California. Nonet describes the increasing legalization of the workers' compensation system in California, as it transformed itself from primarily a welfare agency into something close to a system of courts. We consider this an illustration of the paradox described in this Article. The capacity to weed out fraudulent claims is an unavoidable component of any compensation system that relies on some "merit" standard, however general.

10. These are two of the three sources of externalization identified by Guido Calabresi in his seminal work on the tort system. *See* CALABRESI, *supra* note 2, at 144-45. The third source of externalization identified by Calabresi was inadequate information. *Id.* at 145.

We examine workers' compensation as an illustration of the internalization paradox. We document the obstacles to full internalization in this area, and find that the litigation process is probably the most important obstacle to the goal of internalization.

This Article is divided into three sections. The first examines the theory of internalization and its implications for tort theory. The second and third examine this theory in light of the history and experience of workers' compensation. We argue that the theory of internalization played an important role in the reform movement that led to the adoption of the workers' compensation statutes. Finally, we evaluate workers' compensation as an internalization mechanism.

II. THE INTERNALIZATION PRINCIPLE AND THE TORT SYSTEM

A. Tort Theory and the Theory of Externalities

1. Externality Theory and the Accident Problem

The implications of the theory of externalities for tort law can be explained using a simple example. Consider the simplest type of accident problem: assume that only the injurer's caretaking matters, so that contributory negligence is not an issue. Assume also that there is only one potential tortfeasor (or potential injurer). The cost of taking care is a cost that the potential tortfeasor bears.[11] For example, the cost of looking both ways at an intersection, which is the income foregone by delaying one's arrival at an important event, is borne by the potential injurer. The same goes for the cost of having brakes fixed. The loss to an accident victim in this example is a contingent cost that the potential injurer bears only if forced to by a liability rule.

According to the theory of externalities, economically inefficient decisions result from a divergence between private and social incentives.[12] Social incentives are determined by comparing the "social

11. In the economics of tort law literature, two types of care have been discussed. One, which is referred to simply as "care," is the instantaneous or momentary level of care. The other, which is referred to as "activity," is the amount a tortfeasor participates in an activity that may generate accidents. *See* Steven Shavell, *Strict Liability Versus Negligence* 9 J. LEGAL STUD. 1 (1980). One of the earlier discussions of this distinction is provided in RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 139-41 (2d ed. 1977).

In this section of the text, care can be understood to mean either instantaneous care or activity. The points made in this section are valid regardless of the definition of care. However, most of the examples discussed involve instantaneous care.

12. WILLIAM J. BAUMOL & WALLACE E. OATES, THE THEORY OF ENVIRONMENTAL POLICY 36-45 (2d ed. 1988) (providing a mathematically rigorous statement of incentive

benefits" and "social costs" of a given action. The social benefits are equal to the sum of all of the private benefits generated by an action, and similarly, the social costs are equal to the sum of private costs. Consider, for example, a homeowner who values having his garage painted at $1,000. Suppose the neighbors benefit by $2,000 from the same action—either because they enjoy looking at the newly painted garage or because it raises their property values slightly. Then the social benefits resulting from the homeowner's decision to have his garage painted sum to $3,000, while the private benefit is only $1,000.

Private incentives are determined by comparing the private benefits with the private costs of a given action. If certain social costs are not borne by the decision-maker, then he will have an excessive incentive to engage in an activity whose net private benefits are positive.[13] On the other hand, if some of the social benefits are not enjoyed by the decision-maker, he will have too little incentive to engage in an activity. So in the example of the homeowner who is deciding whether to have his garage painted, if his private cost is $500, the net private benefit from having the garage painted is $500. Since this is less than the net social benefit of $2,500, the homeowner has too little incentive to paint his garage.[14]

In the absence of a liability rule, the potential tortfeasor has an excessive incentive to engage in activities that may give rise to accidents. For example, when crossing an intersection in a car, the potential tortfeasor would compare the benefit of not stopping to look, which is the money saved by not delaying arrival at an important event, with the cost of not stopping to look. In the absence of a liability rule, the cost of not stopping to look would be zero, so the

---

divergence problem). Social incentives are often referred to as "Pareto Optimality conditions." If, in equilibrium, the Pareto Optimality conditions hold, then the equilibrium is said to be Pareto Optimal. An equilibrium is Pareto Optimal if no one's welfare can be improved without making someone else worse off. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 11-12 (3d ed. 1986) [hereinafter POSNER, ECONOMIC ANALYSIS OF LAW].

13. I refer to activities whose net private benefits are positive because these are the only activities in which the decision-maker will voluntarily engage. Activities whose net-private benefits are negative will obviously be avoided by the decision-maker.

14. This example can be used to provide a simple illustration of Coase's critique of externality theory. *See* Coase, *supra* note 6. We assume that the homeowner does not receive the benefits enjoyed by his neighbors, and so he has too little incentive to paint his garage. But this does not dispose of the issue. What prevents the neighbors from paying or "bribing" the homeowner to paint his garage? The homeowner's lack of incentive can only result from obstacles that prevent the neighbors from offering the appropriate bribes.

overwhelming incentive would be to not stop.

A liability rule corrects the incentive problem in the accident setting by forcing the potential tortfeasor to take into account the possible loss to the accident victim. Because the potential loss suffered by the victim would be internalized, the private incentive to engage in accident-generating activity would then be equivalent to social incentive.

2. Externality Theory and Justifications of Tort Law

Although externality theory has the technical flavor of much of economics, previous commentators have failed to recognize its generality. Virtually any theory of tort law can be stated in a form that is entirely consistent with externality theory.

There are two broad categories of justifications of tort law. One is utilitarian theory.[15] Posner has noted that economic theory is not equivalent to utilitarianism.[16] The utilitarian attempts to maximize happiness, while the economist attempts to maximize society's happiness subject to income constraints. In this sense, an economist may quite accurately be called a "constrained utilitarian." While happiness derived from consuming certain items is the only measure taken into account by utilitarians, the economist measures willingness to pay, and then only when it is backed up by an ability to pay. Thus, in asking whether society's welfare is improved if A engages in activity X, the utilitarian simply asks whether total utility is higher, all other things equal, if A is allowed to do X. The economist, on the other hand, asks whether A's willingness to pay to engage in activity X exceeds the opportunity cost of the resources used. Although the basic goal, welfare maximization, is the same, the utilitarian and economic approaches have different practical implications.

Posner is correct in distinguishing utilitarianism from economics, but the differences are small enough to be pushed into the background, on a general level. The differences Posner points to are important, but as long as they are kept in mind, it does no harm to

---

15. Though preceded by Cesare Beccaria and others, Jeremy Bentham is usually mentioned as the founder of utilitarianism. For an example of Bentham's theoretical work, see JEREMY BENTHAM, THE PRINCIPLES OF MORALS AND LEGISLATION (1948). Today, the field of utilitarian justifications of tort law is dominated by the law and economics literature. *See, e.g.*, LANDES & POSNER, *supra* note 2; STEVEN SHAVELL, ECONOMIC ANALYSIS OF ACCIDENT LAW (1987) [hereinafter SHAVELL, ECONOMIC ANALYSIS].

16. RICHARD A. POSNER, THE ECONOMICS OF JUSTICE 48-87 (1983) [hereinafter POSNER, ECONOMICS OF JUSTICE].

place these two bodies of theory in the same general category. The large scale similarities are far more noticeable than the differences Posner has catalogued.

The economic or constrained utilitarian justification of negligence doctrine is well known,[17] and its consistency with externality theory has been explained.[18] The constrained utilitarian would explain the Hand formula[19] as a rule that guarantees that potential tortfeasors will take care when, and only when, taking care maximizes society's wealth.[20] Thus, if the cost of taking care exceeds the potential loss to the accident victim, society's wealth would be reduced if the potential injurer took care. Tort law, specifically the Hand formula, does not require the tortfeasor to take care in this case. This is consistent with externality theory because the Hand formula, as a liability rule, has the effect of internalizing the external cost of failing to take care to those tortfeasors whose caretaking is required if society's wealth is to be maximized. Of course, the alternative to the Hand test, strict liability, also internalizes the external costs of failing to take care. However, a strict liability rule internalizes the costs of accidents whether or not society's wealth would be maximized by additional caretaking on the injurer's part. Under a system of strict liability, the shifting of accident costs to injurers whose caretaking is not socially desirable has no beneficial effect on society's wealth.

The other broad category of positive theory is made up of Kantian, deontological, or natural law justifications.[21] The arguments were originally stated by a school of German theorists led by Savigny,[22] and recently have been championed by Charles Fried[23]

---

17. The first essentially economic justification of negligence doctrine was provided by Oliver Wendell Holmes in lecture three of *The Common Law*, *see* HOLMES, *supra* note 4. The next economic justification of negligence doctrine was Posner's. *See* Richard A. Posner, *A Theory of Negligence*, 1 J. LEGAL STUD. 28 (1972). Now, of course, there is a large and technically complicated literature. *See, e.g.*, LANDES & POSNER, *supra* note 2; SHAVELL, ECONOMIC ANALYSIS, *supra* note 15.

18. *See supra* note 16 and accompanying text.

19. The Hand formula defines negligence as a failure to take care where the cost of taking care is less than the probability of the accident multiplied by the loss if the accident occurs. United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947).

20. *See, e.g.*, POSNER, ECONOMIC ANALYSIS OF LAW, *supra* note 12, at 147-51.

21. POSNER, ECONOMICS OF JUSTICE, *supra* note 16, at 55.

22. *See* Izhak Englard, *Victor Mataja's Liability for Damages From an Economic Viewpoint: A Centennial to an Ignored Economic Analysis of Tort*, 10 INT'L REV. L. & ECON. 173, 175-77 (1990) (discussing Friedrich Karl von Savigny's influence on European 19th century jurisprudence).

23. *See* CHARLES FRIED, AN ANATOMY OF VALUES: PROBLEMS OF PERSONAL AND

and Ernest Weinrib.[24] The Kantians see the negligence rule as a specific application of the more general Kantian rule concerning the proper amount of freedom permitted by the state: that amount of freedom that is consistent with an equal right on the part of others to enjoy their freedom according to general laws.[25] The more specific rule stated by Charles Fried is that every individual should have the fullest freedom to impose the risk of death on others compatible with an equal right on the part of others to impose the risk of death on the individual, according to general laws.[26]

Kant offered his general verbal formula as a solution to a very old problem known as the "paradox of freedom."[27] The paradox is this: In a lawless state all men would be free. But then the strong would control the weak and many would in effect be robbed of their freedom, so in fact very few would really be free. However, the more laws we introduce to prevent this outcome, the more power we give to the state, which restricts the freedom of the individual. In the limit, very few are free.[28]

The paradox is observed in tort law. The absence of any rule assigning liability or constraining the behavior of injurers leads to the absence of freedom, because we would all then be afraid of being hurt. We would stay inside our homes instead of venturing out. On the other hand, very detailed, rigorously enforced controls on potentially injurious behavior produce the same result. The negligence rule, as envisioned by the Kantian theorists, is a compromise between these

---

SOCIAL CHOICE 177-91 (1970) [hereinafter FRIED, ANATOMY OF VALUES]; CHARLES FRIED, RIGHT AND WRONG 7-17 (1978).

24. *See* Ernest J. Weinrib, *Causation and Wrongdoing*, 63 CHI.-KENT L. REV. 407 (1987); Ernest J. Weinrib, *Law as a Kantian Idea of Reason*, 87 COLUM. L. REV. 472 (1987); Ernest J. Weinrib, *Toward a Moral Theory of Negligence Law*, *in* JUSTICE, RIGHTS, AND TORT LAW 123 (Michael D. Bayles & Bruce Chapman eds., 1983) [hereinafter Weinrib, *Moral Theory*]; Ernest J. Weinrib, *Understanding Tort Law*, 23 VAL. U. L. REV. 485 (1989).

It should be noted that George Fletcher's article on tort law is an important part of the Kantian tort theory literature. *See* George P. Fletcher, *Fairness and Utility in Tort Theory*, 85 HARV. L. REV. 537 (1972).

25. IMMANUEL KANT, THE METAPHYSICAL ELEMENTS OF JUSTICE 35-39 (John Ladd trans., 1965). The provision "according to general laws" is important. Any norm that cannot be applied generally cannot satisfy the Kantian rule. An example of a norm that cannot be applied is a rule that I will not play tennis with a player who is not better than me. If everyone were to follow this rule, no one would ever play tennis. *See, e.g.*, ALAN CARTER, THE PHILOSOPHICAL FOUNDATIONS OF PROPERTY RIGHTS 83 (1989). *See generally* Symposium, *Kantian Legal Theory*, 87 COLUM. L. REV. 421 (1987).

26. FRIED, ANATOMY OF VALUES, *supra* note 23, at 185.

27. *See* 2 KARL R. POPPER, THE OPEN SOCIETY AND ITS ENEMIES 44 (5th ed. 1966).

28. *Id.*

extremes.

Of course, the problem the Kantians were trying to solve is a moral one, not one of simply trying to strike a balance at the right amount of potentially injurious behavior.[29] Indeed, there is no right amount in the aggregate of accident-generating behavior under Kantian theory. The theory is concerned with individual compliance with moral norms, not, unlike the various versions of utilitarianism, the aggregate amount of "happiness." Fried has argued that the moral concern in the area of accident law can be framed as a matter of determining the amount one wishes to draw from a hypothetical aggregate "risk budget."[30] The risk budget is something that is constructed in a hypothetical contract, behind a Rawlsian veil of ignorance. Behind this veil, no one knows what sort of person he or she will be in real life. We must decide then what is the proper rate at which an individual can draw from the aggregate risk budget. Society clearly would not choose zero, since that would mean that no one could get on with the basics of life. Nor would it choose an infinite amount, for the same reason. The amount would be a function of the need. An appropriately pressing need would justify a larger withdrawal from the risk budget. The negligence rule is the formalization of this implicit contract.

The most important assumption made by the Kantian theorists, particularly Fried, who is explicit about this, is that we all have similar tastes for risk.[31] For in the absence of this assumption, there is little reason to think that every member of society would agree to the restrictions imposed by negligence doctrine. The important restriction in this respect is the reasonable man standard,[32] which does not allow the tortfeasor to argue that his individual taste for risk justifies an extraordinarily large draw from society's risk budget. If one or more individuals in the hypothetical contract making stage were aware of this problem and they wanted to impose more risk on others, in exchange allowing others to impose more risk on them, they would presumably seek to have such a clause entered into the contract. But this would be inconsistent with the negligence rule. So the Kantian

29. *See generally* George P. Fletcher, *Law and Morality: A Kantian Perspective*, 87 COLUM. L. REV. 553 (1987).

30. FRIED, ANATOMY OF VALUES, *supra* note 23, at 187. For criticism of Fried's approach from a Kantian perspective, see Weinrib, *Moral Theory*, *supra* note 24, at 131-35.

31. FRIED, ANATOMY OF VALUES, *supra* note 23, at 190.

32. *See* W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 32, at 173-93 (5th ed. 1984).

theory of tort law relies on the assumptions of similar tastes and risk aversion. It is a weakness of the theory, which is incidentally shared by Rawls' theory of justice.[33] Some individuals would prefer to bear more risk and their presence cannot simply be assumed away.[34]

Fried recognizes the problem and addresses it by saying that the reasonableness standard is required to make the negligence rule administrable.[35] Suppose negligence doctrine allowed an injurer to offer as a defense that he simply liked to live on the edge and society ought to get used to it. Besides, he would add, I am willing to let the rest of society impose extra risk on me in return. This would clearly contradict the reasonable man standard, and in Fried's view would make negligence doctrine unadministrable because we would never be able to distinguish injurers who were telling the truth from those who had simply figured out a way to beat the system.

The problem with Fried's response is that it places administrability ahead of the moral problem. Morality requires us to let the injurer who simply likes to live on the edge go free. Administrability does not allow us to do this. So a theory that begins with morality ends up making administrability the overriding concern. The Kantian argument is so thoroughly blended with utilitarianism at this stage that it is impossible to save it as an independent justification.

In spite of the weakness in the Kantian arguments, revealed by the assumption of universal risk aversion, they remain a powerful source of justification in legal theory because they disconnect the administrative concerns from the moral concerns and attempt to place the moral concerns first. As a positive theory of tort law, there are weaknesses in the Kantian approach. However, the weaknesses identified here are not as important when viewed as a normative theory.[36]

Previous writers in this area have seen the theory of externalities as a special case of the utilitarian approach, and therefore incompatible with Kantian theory.[37] However, this view is wrong. The essence

---

33. *See generally* JOHN RAWLS, A THEORY OF JUSTICE (1971).

34. *See* Kenneth J. Arrow, *Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice*, 70 J. PHIL. 245, 251-52 (1973).

35. FRIED, ANATOMY OF VALUES, *supra* note 23, at 190.

36. Viewed as a normative theory, the Kantian approach is not diminished in any sense by the fact that its implications are inconsistent with tort doctrine. Further, the strength of the Kantian approach as a normative theory is that it is capable of defending a moral rule such as one that recognizes the value of lives that are not producing and will never produce anything of value to society.

37. For example, Ernest Weinrib asserts that all approaches suggested by "maximizing

of the externality argument is a notion of violation. External costs are costs that the injurer generates and imposes on others. But that is what the Kantian theory, in Fried's terms, says about those who draw too much from society's risk budget.[38] They draw more than the proper amount, given their needs, and therefore impose an unjustified cost on others.[39] The act of drawing too much from the risk budget can be viewed as an externality that should be internalized just as the utilitarian would hold that a violation of the Hand formula requires loss-shifting.

The popular theoretical understanding of externality theory and tort theories places the externality theory under utilitarianism, as a special type of utilitarian approach. The externality theory is seen as providing a set of tools for the utilitarian theorist. However, the better view of externality theory is that it encompasses utilitarian and Kantian approaches to justify accident law. The concerns of the Kantian theorist can be discussed just as easily within the externality framework as those of the utilitarian.[40]

The differences in the use of externality theory under utilitarian and Kantian approaches are observed in trying to reach specific statements of the compliance requirements. The utilitarian approach yields specific dollars and cents figures for the amount of the external cost that should be shifted. In the Hand formula version, it provides a precise amount of care that should be exercised. In the case of a tort injury, the amount that should be shifted is the injury loss suffered by the victim. That is the external cost suffered by the victim. The only problem in application arises in measuring loss, but at least in theory there is no ambiguity concerning the amount that should be internal-

---

theories" are incompatible with corrective justice, and should therefore be rejected as explanations of negligence law. *See generally* Weinrib, *Moral Theory, supra* note 24.

38. FRIED, ANATOMY OF VALUES, *supra* note 23, at 187-91.

39. *Id.*

40. The corrective justice theory of Jules Coleman offers another perspective on the claim that the externality framework is general enough to encompass Kantian and utilitarian approaches. Coleman's view of corrective justice is that it requires nothing more than the annulment of unjust gains and losses. *See* Jules L. Coleman, *Tort Law and the Demands of Corrective Justice*, 67 IND. L.J. 349, 357 (1992). His theory clearly lends support to the "internalization" suggestions of externality theory.

One of the earliest and perhaps the best example of use of the basic ideas of externality theory to justify deontological norms is JOHN STUART MILL, ON LIBERTY (1859). Mill sought to determine the proper scope of government intervention by starting with two principles: that the state should not attempt to regulate an individual's "self-regarding" behavior in order to make that individual better off, and that the state may interfere or regulate when an individual's behavior affects others directly.

ized. Thus, if the victim suffers an unambiguous $200 loss as a result of an injury, externality theory requires that the injurer compensate the victim precisely that amount.

Under a Kantian approach, externality theory does not clearly tell us how much should be shifted from the victim to the injurer, or how much care should be exercised by the potential injurer. The Kantian/Pigovian approach is incapable of telling us this for several reasons. First, unless we assume that all individuals have similar tastes for risk, the Kantian/Pigovian approach goes no further than simply identifying a violation and counseling us to adopt a rule that will prevent such a violation. It can go no further because if people have different tastes for risk, the amount that should be shifted will depend on the preferences of the individual tortfeasor. A second and more general reason that the Kantian/Pigovian approach is not very useful at a practical level is that it does not yield an explicit formula which allows us to identify when someone has drawn too much from the risk budget. We could use the Hand formula, arguing that it must approximate whatever the Kantian theorist has in mind, but the theory itself does not generate a formula.

The inability of the Kantian approach to generate a mathematical rule such as the Hand formula is a serious weakness because it means that the theory is incapable of telling us what level of internalization is appropriate. For example, it is difficult to use the Kantian approach to criticize the levels of tort compensation. The inability of the theory to provide a rigorous justification of a mathematical internalization rule also means that the theory probably cannot, with a fair degree of rigor, be used to make a principled choice between strict liability and negligence. This problem has been revealed in the literature already. The two most prominent Kantian theorists, Charles Fried and Richard Epstein (in his earlier writings), part ways when it comes to the choice between strict liability and negligence.[41] Fried uses Kantian arguments to provide an elaborate justification of negligence doctrine. Epstein sees the theory as providing a justification for strict liability. Yet the reason for the different positions taken by these theorists is very hard to understand precisely because the theory has not been presented in a suitably rigorous form. This criticism does not apply to utilitarian approaches because the assumptions that would lead one

41. *Compare* Richard A. Epstein, *A Theory of Strict Liability*, 2 J. LEGAL STUD. 151 (1973), *with* FRIED, ANATOMY OF VALUES, *supra* note 23. This point is noted in POSNER, ECONOMICS OF JUSTICE, *supra* note 16, at 55.

theorist to choose strict liability over negligence can be stated clearly.

Because of the weaknesses in the Kantian framework, this Article will adopt the traditional utilitarian/Pigovian approach. However, the point of the discussion in this section is not to uncover weaknesses in the Kantian justifications of tort law, but to demonstrate the generality of the externality framework.

3. The Utilitarian/Pigovian Framework

Much of the economic theory of tort law falls under what could be called a utilitarian/Pigovian framework. There are two branches within this subdivision. One is positive theory, which has been promoted in the work of Richard Posner and William Landes.[42] The other is normative theory, which is exemplified by the work of Guido Calabresi.[43]

Positive theory is concerned largely with demonstrating that the doctrinal rules can be justified on efficiency grounds, an approach which has been referred to as "doctrinal efficiency."[44] The characteristic feature of the doctrinal efficiency approach is that it ignores administrative costs in determining efficient rules.[45] Thus, under the approach of the positive theorists, a rule is efficient if it can be shown to minimize *Total Costs = Accident Costs + Accident Avoidance Costs.*[46] The Hand formula is doctrinally efficient because

42. *See* LANDES & POSNER, *supra* note 2 and the literature cited therein.

43. *See* CALABRESI, *supra* note 2. The normative theory literature is probably larger than the positive theory literature, especially if one counts all of the articles written by economists that offer models of optimal tort systems (e.g., optimal tort damages). Gordon Tullock is responsible for some of the rare contributions to the reform literature that combine theoretical sophistication with knowledge of the legal system. *See* GORDON TULLOCK, THE LOGIC OF THE LAW (1971); GORDON TULLOCK, TRIALS ON TRIAL: THE PURE THEORY OF LEGAL PROCEDURE (1980) [hereinafter TULLOCK, TRIALS ON TRIAL].

44. Keith N. Hylton, *Litigation Costs and the Economic Theory of Tort Law*, 46 U. MIAMI L. REV. 111, 148 (1991) [hereinafter, Hylton, *Economic Theory*].

45. For example, in determining whether the negligence rule is efficient, the doctrinal efficiency approach ignores litigation costs. The Hand formula is consistent with this approach. *Id.* at 147. If litigation costs were taken into account, then the injurer's burden of precaution, B, should be compared to the expected loss, PL, plus the expected cost of litigation, C. For further discussion, see *id*; Keith N. Hylton, *Costly Litigation and Legal Error Under Negligence*, 6 J.L. ECON. & ORG. 433 (1990) [hereinafter Hylton, *Costly Litigation*]. On the influence of litigation costs on welfare under a liability system, see Janusz A. Ordover, *Costly Litigation in the Model of Single Activity Accidents*, 7 J. LEGAL STUD. 243 (1978); Janusz A. Ordover, *On the Consequences of Costly Litigation in the Model of Single Activity Accidents: Some New Results*, 10 J. LEGAL STUD. 269 (1981); A. Mitchell Polinsky & Daniel L. Rubinfeld, *The Welfare Implications of Costly Litigation for the Level of Liability*, 17 J. LEGAL STUD. 151 (1988); Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399 (1973).

46. *See, e.g.*, John P. Brown, *Toward an Economic Theory of Liability*, 2 J. LEGAL

it is the rule that minimizes this sum.

The doctrinal efficiency approach is to be contrasted with a different approach which we refer to as "operational efficiency." Operational efficiency takes administrative and other costs into account in determining an efficient legal rule. Thus, an operationally efficient, rule minimizes *Total Cost = Accident Cost + Accident Avoidance Cost + Administrative Costs.* Calabresi was explicit in stating this as the goal of an operationally efficient tort system.[47]

The distinction between the two approaches is important because a rule that is doctrinally efficient may not be operationally efficient and vice versa. The classic example is the Hand formula for negligence. Brown,[48] Landes and Posner,[49] and others have demonstrated that the Hand formula is doctrinally efficient. But they have not shown it to be operationally efficient. A careful examination reveals that the Hand formula is not operationally efficient, as long as administrative costs are positive.[50] Thus, even the most basic implication of the doctrinal efficiency school is called into question when the goal is shifted to operational efficiency.

Not surprisingly, the doctrinal efficiency approach is predominant in the tort theory literature because most of the literature aims to explain tort doctrine. The literature has a long history beginning probably with Holmes' discussion of tort law in lectures three and four of *The Common Law.*[51] Holmes argued that much of tort doctrine could be explained by using cost benefit arguments. The cost benefit arguments have been expanded upon and brought under the mantle of economic theory by Landes and Posner, and other theorists.[52]

The operational efficiency literature is of fairly recent origin and its best known work is Guido Calabresi's *The Costs of Accidents.*[53] The operational efficiency theorists have generally criticized the tort system, with a view toward reform.[54] They have been less concerned with the doctrines of tort law than with empirical evidence of its performance on deterrence or compensation grounds.

STUD. 323, 324-26 (1973).

47. *See* CALABRESI, *supra* note 2, at 28-29.
48. *See* Brown, *supra* note 46, at 332-47.
49. *See* LANDES & POSNER, *supra* note 2, at 85-103.
50. Hylton, *Costly Litigation*, *supra* note 45, at 444-45.
51. *See* HOLMES, *supra* note 4.
52. *See* LANDES & POSNER, *supra* note 2, at 40-41, 64-69.
53. *See* CALABRESI, *supra* note 2.
54. *Id.* at 1-15; TULLOCK, TRIALS ON TRIAL, *supra* note 43, at 34-48.

Although this Article is in the tradition of Calabresi and other theorists who have examined operational efficiency, we will take no position on whether operational efficiency or doctrinal efficiency should be the approach of tort theorists. One strong argument for doctrinal efficiency is that the administrative costs of the tort system are not fixed and may change at any moment. If operational efficiency is the goal, then the tort rules would have to be changed every time the administrative costs of using the tort system were changed. On the other hand, there is a strong argument for making operational efficiency the main goal. When there is a conflict between operational and doctrinal efficiency, why should we be concerned with doctrinal efficiency? The ultimate goal, after all, is to deter reckless or careless behavior that may cause accidents. Any approach that deviates from this goal threatens to defend or to generate a set of useless rules.

B. Calabresi's Critique of the Tort System: A Reexamination

Since the major work in the vein of this Article is Calabresi's *The Costs of Accidents*, we will use it as a backdrop for some of our arguments concerning tort theory. Calabresi identified three categories of social cost: primary, secondary, and tertiary.[55] Primary costs are accident costs and avoidance costs.[56] The positive theory literature examines this problem only. Recall that the Hand formula is the most basic implication of a simple model which attempts to minimize the sum of accident and avoidance costs.[57] Secondary costs in Calabresi's framework result from inadequate risk spreading.[58] Thus, Calabresi is willing to sacrifice additional primary costs if it will generate a greater saving in secondary costs. For example, Calabresi would be willing to deviate from the Hand rule if its result is to leave accident costs on people who are poor, and therefore unable to purchase insurance in order to shift away the risk of loss. Tertiary costs are the administrative costs of running the tort system.[59] Again Calabresi's argument suggests that we should be willing to sacrifice primary and secondary costs if it will lead to greater savings in tertiary costs. Thus, if a rule of no liability reduces tertiary costs by an amount that more than compensates for the increase in primary and

55. CALABRESI, *supra* note 2, at 26-29.
56. *Id.* at 26-27.
57. *See supra* note 19 and accompanying text.
58. CALABRESI, *supra* note 2, at 27-28.
59. *Id.* at 28.

secondary costs, Calabresi would recommend this rule.[60] Calabresi set out as the goal of tort law the minimization of the sum of primary, secondary, and tertiary costs.[61]

Our approach differs from Calabresi's in two respects. First, we have generally ignored the secondary or risk spreading function of tort rules emphasized by Calabresi. We have done this for two reasons. One is that with insurance available, there seems to be little practical reason for aiming to fashion liability rules as if insurance were not available. As Holmes pointed out more than 100 years ago, the private insurance market is more efficient in reallocating risk than is the court system.[62] Although we recognize the general validity of taking secondary costs into account in considering the welfare arguments for various liability rules, we think it is unlikely that a stable set of recommendations can come from this source. Private arrangements can undermine any attempt to use liability rules to reallocate risk.[63] The second reason, which is specific to this study, is that workers' compensation involves an ongoing contractual relationship between the employer and employee. There is every incentive within that relationship to use the wage contract and other provisions to

60. One important proposition that has come out of the theoretical literature is that this problem is possible. Because private and social incentives to litigate are not the same, plaintiffs may bring suit in areas in which a no liability rule would be socially desirable. For proof of this general proposition, see Steven Shavell, *The Social Versus the Private Incentive to Bring Suit in a Costly Legal System*, 11 J. LEGAL STUD. 333, 334-38 (1982). For refinements of the proposition, see Louis Kaplow, *Private Versus Social Costs in Bringing Suit*, 15 J. LEGAL STUD. 371, 372 (1986) (Shavell's argument shows the externality created by litigation costs and is undisturbed by Menell's argument); Peter S. Menell, *A Note on Private Versus Social Incentives to Sue in a Costly Legal System*, 12 J. LEGAL STUD. 41 (1983); Susan Rose-Ackerman & Mark Geistfeld, *The Divergence Between Social and Private Incentives to Sue: A Comment on Shavell, Menell, and Kaplow*, 16 J. LEGAL STUD. 483, 484 (1987) (explaining that by shifting to the British rule, the Menell-Kaplow result holds in general and is not a special case).

For discussion of the implications of the Coase theorem for Shavell's argument, see Hylton, *Economic Theory*, *supra* note 41, at 120-26 (two sources of inefficiency lie below the claim that suit may be inefficient because the cost of bringing it "exceeds the net benefits from the deterrence provided by the threat of suit," and the suit may be inefficient "because potential plaintiffs are unable to identify those potential defendants who will find it inefficient to bring suit").

61. *See* CALABRESI, *supra* note 2, at 26-29.

62. *See* HOLMES, *supra* note 4, at 96, 199-201.

63. Consider for example the relationship between a heart surgeon and her patient. A liability rule requiring the heart surgeon to pay for any injury to the patient would lead the surgeon to purchase liability insurance. If the patient's demand for her services were not very responsive to price, as might be the case with heart surgeons, the doctor could probably pass the full cost of the insurance policy to the patient through charging higher fees.

spread risk optimally.[64] Hence, we will focus on incentive considerations, restricting our attention to primary and secondary costs.

The second and more important sense in which our approach differs from Calabresi's is the treatment of primary and tertiary costs. It is confusing to put primary and tertiary costs in different categories because they are interdependent. Calabresi's theoretical framework is flawed to the extent that it does not take into account the interdependence of the categories of social cost, especially primary and tertiary costs.[65] The framework suggests that these categories can be examined independently when in fact they cannot. Primary and tertiary costs are interdependent in some obvious senses. One typical component of tertiary costs is the plaintiff's cost of bringing a tort claim. However, if the cost of bringing suit is made sufficiently high, it will choke off all tort claims. There would then be no disincentive to careless behavior, so that primary costs would increase dramatically as well. Alternatively, if the cost to the defendant of defending himself against a claim were trebled, this would increase the cost of careless behavior and reduce injuries to some extent. Thus, our analysis differs from Calabresi's by taking explicitly into account the interaction between administrative and primary costs.

Calabresi based his analysis of the tort system squarely on the theory of externalities.[66] He criticized the tort system for failing to shift accident costs to the important sources of injury.[67] In this sense, his work was more in the spirit of Fleming James than Oliver Wendell Holmes.[68] Calabresi preferred liability rules that performed the function of shifting costs, such as strict liability, and would concentrate accident costs on enterprises that were capable of insuring themselves and passing part of the cost to consumers through charging higher prices for their goods.

Calabresi argued that the tort system, specifically the fault system, was inadequate as a mechanism for internalizing the costs of

64. For example, the workers may accept a lower wage in exchange for employer-financed health or disability insurance. If transaction costs do not prevent bargaining, all of the desired risk spreading will take place through the employment contract.

65. This is not to say that Calabresi himself was guilty of this mistake. The interdependence problem is noted in several areas of his book. *See, e.g.*, CALABRESI, *supra* note 2, at 143-44.

66. *See, e.g., id.* at 68-77, 144-50.

67. *Id.* at 68-70.

68. For discussion of Fleming James's scholarship and its impact on tort law, see George L. Priest, *The Invention of Enterprise Liability: A Critical History of the Intellectual Foundations of Modern Tort Law*, 14 J. LEGAL STUD. 461, 465-83 (1985).

accident-generating activity.[69] He identified three major sources of externalization. One is imperfect insurance pricing.[70] Actuarially-fair insurance prices should equal the discounted expected value of the claim against the insurer, and in theory this equality should hold at all times.[71] The ideal insurance company would require its customers to wear risk meters that would tell it at every moment how much risk was being generated and would vary the price of insurance accordingly. By varying the price, the insurer would not only ensure that the price was actuarially fair, so that it would not lose money on an individual claim, but would also give the insured the appropriate incentive to avoid causing injury. The insured, instead of comparing the change in expected liability caused by an incremental adjustment in risky behavior, would compare the change in the price of insurance caused by the adjustment. In an actuarially fair market, however, these amounts would be the same, so that actuarially fair insurance pricing perfectly internalizes expected external costs. Of course, no insurance arrangement works this well.[72] Insurance prices do not at all times accurately reflect risk, and categories for risk rating purposes are not as precise as they could be drawn. The imperfections have the effect of externalizing accident costs. For example, the costs of particularly bad drivers below twenty years old are externalized under the insurance classifications to all drivers below twenty. As a result, the bad drivers are not given the proper signals for caretaking by the insurance market.

The second reason the tort system fails to internalize external

---

69. CALABRESI, *supra* note 2, at 244.

70. *Id.* at 246-49.

71. *See* SHAVELL, ECONOMIC ANALYSIS, *supra* note 15, at 202-03. This proposition assumes the insurer can monitor the insured's risk creating activities. It also assumes that the insurer incurs no administrative costs. If the insurer cannot costlessly monitor risky activities, then prices will not equal expected claims. On the economics of insurance pricing when the insured party can influence the risk of a claim, see Steven Shavell, *On Moral Hazard and Insurance*, 93 Q.J. ECON. 541 (1979).

72. This should be clear for the reasons stated *supra* note 71. Administrative costs prevent insurers from setting prices at the actuarially fair level. Insurers cannot monitor the risk creating activities of the insured, which gives rise to the problem of "moral hazard" (i.e., of insured parties having inadequate incentives to take care to avoid an accident). Insurers cannot costlessly categorize risks, which gives rise to the "adverse selection" problem (i.e., the problem of low risk parties opting out of the insurance market). For a survey of the literature, see KARL H. BORCH, ECONOMICS OF INSURANCE (Knut K. Aase & Agnar Sandmo eds., 1990); on adverse selection, see Michael Rothschild & Joseph Stiglitz, *Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information*, 90 Q.J. ECON. 629 (1976); Charles Wilson, *A Model of Insurance Markets with Incomplete Information*, 16 J. ECON. THEORY 167, 187-202 (1977).

costs is imperfect information.[73] Calabresi argued that other things being equal, liability should be placed on the party best able to recognize a relationship between caretaking and costs, and to use this information to reduce accidents.[74] In the workers' compensation setting, this argued for the strict liability rules adopted by statute in every state, because in Calabresi's view, these rules shifted the injury costs to the party best able to control them. The worker was too ill-informed on the likelihood of an accident to be able to control accidents through his own behavior. Placing liability on the employer gave the employer the incentive to control the cost or to inform the worker how to control accident costs.

The third reason the tort system fails to internalize costs is the tax and transfer system.[75] If the social security system, for example, compensates someone whose disability was caused by an accident, that person may choose not to seek compensation through the courts. The "collateral source rule"[76] seems to hold this incentive in place, but there are two problems with this answer. One is that someone who has received compensation from a collateral source may simply decide not to seek compensation from the injurer even if an award is likely.[77] Second, some states have passed statutes preventing those who collect from government insurance funds to take advantage of the collateral source rule.[78] Social security disability payments are a

73. *See* CALABRESI, *supra* note 2, at 244-46.

74. *Id.*

75. *Id.* at 246.

76. The collateral source rule holds that the tortfeasor is not entitled to a credit for payments or benefits received by the injured party from independent sources. *See, e.g.*, Smith v. United States, 587 F.2d 1013, 1015 (3rd Cir. 1978); Tebo v. Havlik, 343 N.W.2d 181, 186-87 (Mich. 1984). For a general discussion, see Daena A. Goldsmith, *A Survey of the Collateral Source Rule: The Effects of Tort Reform and Impact of Multistate Litigation*, 53 J. AIR L. & COM. 799 (1988).

77. Why? Because seeking a damage award is time consuming and the victim may prefer not to devote his or her time to it. In other words, for the victim to have an incentive to sue, the award must exceed the cost of litigating and the opportunity cost of the victim's time. Alternatively, the victim may be afraid of some sort of retaliation from the tortfeasor if he brings suit.

78. *See, e.g.*, RICHARD A. EPSTEIN, CASES AND MATERIALS ON TORTS 787-88 (5th ed. 1990); *see also* Jerry J. Phillips, *Comments of the Report of the Governor's Commission on Tort and Liability Insurance Reform*, 53 TENN. L. REV 679 (1986); Linda J. Gobis, Note, Lambert v. Wrensch: *Another Step Toward Abrogation of the Collateral Source Rule in Wisconsin*, 1988 WIS. L. REV. 857; L. Timothy Perrin, Comment, *The Collateral Source Rule in Texas: Its Impending Demise and a Proposed Modification*, 18 TEX. TECH. L. REV. 961 (1987); Julie A. Schafer, Note, *The Constitutionality of Offsetting Collateral Benefits Under Ohio Revised Code Section 2317.45.*, 53 OHIO ST. L.J. 587 (1992).

source of externalization because they essentially allow tortfeasors to shift the costs of accidents to the public at large.

We question below whether insurance pricing presents the obstacle to internalization that Calabresi thought that it did. Our reason is based simply on a reliance or trust in market incentives in an insurance arrangement. The insurance company need not control all risky behavior through pricing. It may make direct requests to the insured to change its behavior in some way, offering in exchange to reduce the price, or threatening not to compensate if an accident should occur, or not to renew the policy. There are many instruments available to the insurer, and we should not expect to see price alone (or, equivalently, risk categorization) used to regulate the behavior of the insured. The insurance company has incentives to use the cheapest method of controlling or regulating the behavior of the insured. Using insurance pricing alone to control risky behavior, we argue, may have harmful effects. We think there needs to be a great deal of evidence to overcome the presumption that the parties to an insurance contract have every incentive to find ways to reduce accident claims that could be avoided at low cost. And in our view, there is simply not enough evidence.

We also argue that Calabresi failed to mention perhaps the most important source of externalization: the litigation process. This is more than just a minor oversight, because the problem stands in the way of the sort of smooth cost shifting that he envisioned under a system which made greater use of strict liability rules. The litigation process is costly, and therefore effectively bars all claims that are not great enough to cover the plaintiff's cost of litigating. In addition, liability rules that make compensation easier to obtain also increase the stakes of litigation, further driving up the costs. The result is that efforts such as those proposed by Calabresi to reduce errors and to improve the likelihood of recovery against certain types of injurers drive up the administrative costs of the system, and may therefore shut out a larger share of claims. Without having to worry about these claims, the tortfeasor has little incentive to avoid accidents.

C. The Internalization Paradox

Our major criticism of Calabresi's argument is a general criticism of the internalization principle emphasized originally by Pigou.[79] The

79. *See* PIGOU, *supra* note 1.

criticism can be stated in the form of a proposition which we call the internalization paradox. The paradox is this: Any effort to internalize accident costs will have to be accompanied by a method of distinguishing fraudulent claims. But any effort to sift out fraudulent claims will be administratively costly, and part of the administrative cost will be borne by victims who seek compensation through the tort system. Thus, victims whose claims are too small to cover the portion of the administrative cost that must be borne by the victim will effectively be barred from the tort system.[80] Further, the administrative costs borne by victims will generally not be shifted in full to injurers, so that those victims who seek compensation through the tort system will not have their costs shifted in full to injurers.

There is more to the paradox. Almost any rule which improves the tort system as an internalization mechanism by reducing the frequency of error, or simply increasing the likelihood of recovery, will probably lead to an increase in the administrative costs of the system, and as a result, a reduction in the number of claims that it can efficiently process. In other words, improvements in internalization are likely to be achieved at the cost of shrinking the tort system by reducing the number of accident victims who effectively have access to it. We provide a largely intuitive argument for this claim below.

The need to separate out fraudulent claims generates administrative costs. By administrative costs, we mean the costs to the plaintiff and the defendant of litigating their claims, and the costs to the state of the resources devoted to operating the court system. The largest component in the cost of litigating is the attorney's fee,[81] which will be positive even if the parties settle the dispute. The cost of litigating also includes fees for experts and court filing fees. The cost also includes the opportunity cost of taking time away from other pursuits to manage a lawsuit, either as a plaintiff or as a defendant. Except for the costs borne by the state or the public in providing a court system, the costs detailed above are private costs. Let C_p represent the private cost to the plaintiff of litigating and let C_d represent the private cost to the defendant of litigating.

---

80. For a mathematical demonstration in a model of strict liability, see Keith N. Hylton, *The Influence of Litigation Costs on Deterrence Under Strict Liability and Under Negligence*, 10 INT'L REV. L. & ECON. 161 (1990) [hereinafter Hylton, *Influence of Litigation Costs*].

81. *See* JAMES S. KAKALIK & NICHOLAS M. PACE, COSTS AND COMPENSATION PAID IN TORT LITIGATION 37-44 (1986) (detailing components of plaintiff's legal expenses); David M. Trubek et al., *The Costs of Ordinary Litigation*, 31 U.C.L.A. L. REV. 72 (1983) (analyzing litigation costs from conducted survey).

Suit will be brought by a rational plaintiff if, and only if, the expected award exceeds the plaintiff's cost of litigating. Let EA_p be the expected award from the perspective of the plaintiff.[82] Let EA_d be the expected award from the defendant's perspective. The net expected award to the plaintiff is $EA_p - C_p$. The expected cost (or expected total liability) to the defendant is $EA_d + C_d$. The plaintiff will have an incentive to bring suit only if the net expected award is positive, i.e., EA_p is greater than C_p.

The first statement of the internalization paradox is easily established. Because litigation is costly, all victims for whom the expected award falls below the plaintiff's cost of litigating $(EA_p - C_p<0)$ will be effectively barred from the courts.

The second statement is a virtually unavoidable result of an administratively costly procedure: unless the plaintiff's litigation costs are shifted in full to the injurer, even those victims who seek compensation through the tort system will not have all of their costs internalized to the injurer. They will be forced to bear the cost of litigating, which is as much a result of the injurer's behavior as the injury that led the victim to bring suit. One possible solution may be to shift litigation expenses to the defendant. But this is not enough because litigation expenses include some virtually unmeasurable costs, such as the time that a plaintiff must put into developing his claim. How does one compensate for the risk of losing a suit that turns out to be successful? There will inevitably be costs that the system will not be able to measure accurately and shift from the victim to the injurer in a tort suit. Further, what about the tort victims who lose their suits? If the claim was fraudulent, then the resulting loss is a desirable result. But what about those victims who lose because of an error on the court's part? They are saddled with legal expenses and the economic loss due to injury that led them to sue in the first place.

The third statement of the paradox is that any effort to reduce courtroom error will increase administrative costs, and therefore shut out a larger number of potential plaintiffs. The typical approach to reducing courtroom error involves a procedural change. For example, the rule against hearsay evidence,[83] or the more basic rule excluding

82. The expected award is equal to the probability of a verdict in favor of the plaintiff multiplied by the judgment.

83. *See, e.g.*, FED. R. EVID. 801.

prejudicial evidence,[84] are methods of reducing the likelihood of courtroom error. To the extent that such procedural changes require more costly methods of presenting a claim, they reduce the number of claims that can reach the court. Restrictions on the kinds of arguments or evidence that a plaintiff can use in court effectively increases the cost of litigating because the plaintiff has an incentive to use only evidence that increases the probability of recovery more than the cost of presenting the evidence.

The fourth and final statement is that an effort to increase the likelihood of recovery, such as shifting from the negligence rule to strict liability, may lead to an increase in administrative costs and a reduction in the number of claims that have access to the system. This should be understood to be a weaker claim than the preceding three. It requires a somewhat more detailed justification.

In litigation, the plaintiff makes expenditures with a view toward maximizing the net award.[85] He therefore trades off increases in the cost of litigating for increases in the expected award. The plaintiff has an incentive to keep spending money in litigation as long as the expenditure increases the expected award more than the cost of litigating. Suppose then that the first dollar of expenditure increases the plaintiff's expected award by $10. The increase in expenditure is, of course, $1. Since the net award increases by $9, the plaintiff will have an incentive to spend the first dollar. It seems reasonable to assume that the amount that additional expenditure on litigation raises the plaintiff's expected award falls as the litigation expenditure increases. This is simply the principle of diminishing returns,[86] which is as likely to apply in the litigation process as in any productive activity. If the principle of diminishing returns holds in this situation, then the plaintiff will continue spending on litigation until the increase in the plaintiff's expected award just equals the increase in the cost of litigating. Let dEA_p/dC_p represent the increase in the expected award that results from one additional dollar of litigation expenditure;

84. *See, e.g.*, FED. R. EVID. 403.

85. *See, e.g.*, John C. Hause, *Indemnity, Settlement, and Litigation, or I'll Be Suing You*, 18 J. LEGAL STUD. 157 (1989).

86. The principle can be stated as follows: If in a productive activity requiring several inputs, one input is increased while the others are held fixed, the resulting increase in output will fall as more and more unrestricted input is added. *See* JACK HIRSHLEIFER, PRICE THEORY AND APPLICATIONS 271-74 (2d ed. 1980); *see also* Mark F. Grady, *Proximate Cause and the Law of Negligence*, 69 IOWA L. REV. 363, 369-70, 424-29 (1984) (explaining a positive economic theory of the role of proximate cause in negligence law).

then the principle of diminishing returns implies that the plaintiff will continue spending on litigation until dEA_p/dC_p is equal to $1.

The defendant will aim to minimize total liability which requires trading off decreases in the expected judgment, EA_d, with increases in the defendant's litigation expenditure, C_d. The principle of diminishing returns should hold in this case also, so that the defendant will keep spending on litigation until the change in the expected award falls to $1.

Almost any effort to improve the tort system as an internalization mechanism will (or should) have the effect of increasing the amount that a slight increase in litigation expenditure adds to the plaintiff's expected award. That is, adoption of a rule that increases the likelihood of recovery should increase the "productivity" of an additional dollar of litigation expenditure on the part of the plaintiff. Thus, after adoption of the new rule, the first dollar of expenditure increases the expected recovery by $20, whereas before the rule the first dollar of expenditure increased expected recovery by only $10. If this "productivity effect" holds at all levels of expenditure, then adoption of a rule that increases the likelihood of recovery will also increase the plaintiff's total expenditure on litigation. A similar argument holds for the defendant's incentive to invest more into litigation. Any increase in the plaintiff's likelihood of recovery will increase the stakes for the defendant, causing him to spend more. If the parties are forced to match each other's litigation expenditures, the result will be a higher level of cost for both parties. Figure 1 illustrates the plaintiff's incentive to invest in litigation before and after the legal innovation.

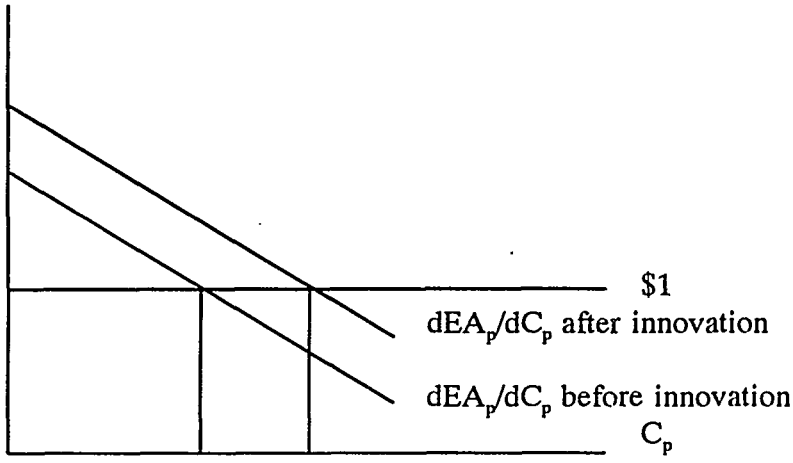

FIGURE 1

However, an increase in the plaintiff's cost of litigating will effectively bar claims that do not satisfy the threshold requirement that the expected award exceed the plaintiff's cost of litigating, i.e., $EA_p > C_p$. This raises the question: Will a larger number of claims be effectively barred if the cost of litigating rises after a change in the liability rule? Only if the expected award does not rise by at least as much as the increase in the plaintiff's cost of litigating. But there is no guarantee that the expected award will rise by at least as much as the increase in the plaintiff's cost of litigating, because the defendant's expenditures are devoted to reducing the expected award. It can be demonstrated that if the plaintiff is unaware of or underestimates the extent of the defendant's response, the plaintiff's additional litigation expenditures may exceed the increase in the expected award.

Our claim is that because of rent-seeking expenditures by both parties, the increase in the expected award following a shift in the liability rule may not cover the plaintiff's additional litigation expenses. A rigorous way of stating the argument is as follows. Let the plaintiff's expected award be a function of the amount invested in litigation by both parties, so that $EA_p = F(C_p, C_d)$. Similarly, let the defendant's expected award be a function of litigation expenditures, $EA_d = G(C_p, C_d)$. Let the increase in the expected award caused by a \$1 increase in C_p, holding C_d fixed, be given by $F_1 > 0$; and let the increase in the expected award caused by a \$1 increase in C_d, holding C_p fixed, be given by $F_2 < 0$. The increase in the expected award

caused by an increase in the plaintiff's litigation expenditures is then: $dEA_p/dC_p = F_1 + F_2(dC_d/dC_p)$. If the plaintiff is unaware of the defendant's response, and $dC_d/dC_p > 0$,[87] he will overestimate the productivity of investment in litigation. Similarly, the decrease in the expected award caused by a small increase in defensive expenditures by the defendant is $dEA_d/dC_d = G_1(dC_p/dC_d) + G_2$. Thus, if the defendant fails to take the plaintiff's response into account, she will overestimate the impact of her defensive expenditures. The result will be a rise in litigation expenditures without an offsetting rise in the expected award. In game theory terminology, we have demonstrated that if parties adopt "Nash" strategies,[88] the increase in the equilibrium level of litigation expenditures in response to an increase in the probability of an award may exceed the increase in the expected award itself.[89]

This completes our argument for the internalization paradox. We do not think that all parts of this proposition hold in all cases, particularly the last claim. We have argued that under certain conditions an effort to improve the tort system as an internalization mechanism will,

---

87. It can be demonstrated that when dC_d/dC_p is positive, the cross partial derivative F_{12} is positive and F_{11} is negative. Both of these assumptions are intuitively plausible. That F_{11} is negative follows from the assumption of diminishing returns. F_{12} being positive is an implication of the diminishing returns assumption. *See supra* note 86 and accompanying text.

88. In this context, "Nash" strategies implies that the litigants determine their level assuming that the other litigant maintains his current level of litigation expenditure. *See generally* ERIC RASMUSEN, GAMES AND INFORMATION: AN INTRODUCTION TO GAME THEORY 32-41 (1989).

89. There are a number of issues that are not appropriate for full consideration in this forum, but should at least be mentioned. First, there is a possibility that there will be several Nash equilibria. However, the basic incentive problem that generates our conclusion remains even in this case. Second, there are other equilibria. One set of equilibria are the Stackelberg equilibria (two of them) in which one party spends less and the other spends more than in the Nash equilibrium. These equilibria are generated by assuming that one party is fully aware of the other's reaction to his increase in litigation expenditures. Another equilibrium is generated by assuming that both parties know each other's responses to increased litigation expenditure. In this equilibrium, the parties accurately assess the private marginal benefit of an increase in litigation expenditure, and therefore spend less than in the Nash equilibrium.

Readers familiar with the Cournot-Nash game will find it strange that the parties spend less on litigation when they are aware of each other's responses, because in the standard story of the game, awareness of responses leads to excessive output. The reason is that consideration of the responses acts as a subsidy to production in the traditional Cournot-Nash game, while it acts as a tax on litigation expenditure in the game discussed in the text. On the details of the Cournot-Nash game, see MICHAEL INTRILIGATOR, MATHEMATICAL OPTIMIZATION AND ECONOMIC THEORY (1971). For an explanation of Stackelberg gaming behavior, see KALMAN J. COHEN & RICHARD M. CYERT, THEORY OF THE FIRM: RESOURCE ALLOCATION IN A MARKET ECONOMY 240-43 (2d ed. 1975).

once litigant behavior is adequately taken into account, have the effect of driving up administrative costs and making courts inaccessible to a larger number of potential plaintiffs.

In the remainder of this Article we examine the United States workers' compensation system in light of the theory presented in this section. In particular, we focus on its success as an internalizing mechanism.

III. THE INTERNALIZATION PRINCIPLE AND WORKERS' COMPENSATION

A. *Emergence of Workers' Compensation*

The long parade of socio-legal reforms that followed on the heels of the new industrial age included as a relatively early development the introduction of workers' compensation arrangements. The origins of the idea of assured compensation for workers injured on the job can be found in Imperial Germany.[90] A series of laws passed between 1883 and 1889 offered disabled workers a solution more reliable than existing remedies established under civil code or common law regimes.[91] The most important were the Sickness Insurance Law of 1883 and the Accident Insurance Law of 1884.[92] The Sickness law required employers and employees to contribute to a fund that would be used to compensate workers who were disabled, as the result of an injury on the job, for thirteen weeks or less.[93] The Accident law established a fund that covered employees who were disabled for longer than thirteen weeks.[94] Only employers were required

90. *See* 1 ARTHUR LARSON, THE LAW OF WORKMEN'S COMPENSATION § 5.10 (1991); Paul R. Gurtler, Comment, *The Workers' Compensation Principle: A Historical Abstract of the Nature of Workers' Compensation*, 9 HAMLINE J. PUB. L. & POL'Y 285, 288-90 (1989).

91. *See* Kenelm E. Digby, *Recent European Legislation with Regard to Compensation for Industrial Accidents*, 17 YALE L.J. 485, 487-88 (1908); Arthur Larson, *The Nature and Origins of Workmen's Compensation*, 37 CORNELL L.Q. 206, 221-31 (1952); J.E. Rhodes II, *The Inception of Workmen's Compensation in the United States*, 11 ME. L. REV. 35, 38-39 (1917).

92. *See* Digby, *supra* note 91, at 488-89; P. Tecumseh Sherman, *Can the German Workmen's Insurance Law Be Adapted to American Conditions?*, 61 U. PA. L. REV. 67, 68-70 (1912) [hereinafter Sherman, *German Workmen's Insurance Law*]; P. Tecumseh Sherman, *The Jurisprudence of the Workmen's Compensation Laws*, 63 U. PA. L. REV. 823, 856-60 (1915) [hereinafter Sherman, *Workmen's Compensation*].

93. Sherman, *German Workmen's Insurance Law*, *supra* note 92, at 69; Sherman, *Workmen's Compensation*, *supra* note 92, at 856-57.

94. Sherman, *German Workmen's Insurance Law*, *supra* note 92, at 69.

to contribute to the Accident fund.[95] Both funds offered no-fault support for disabled workers.[96]

Because the vast majority of worksite injuries were not serious, the Sickness fund turned out to be the most important,[97] paying out on a yearly basis more than twice the amount distributed by the Accident fund.[98] Since in most cases there was no inquiry into fault on the worker's part, the German scheme effectively severed any connection between the risk characteristics of the employer's worksite and the "tax" levied upon the employer and employees at a given worksite. Employer contributions to the Accident fund did increase with the risk of injury.[99] But the Accident fund covered a relatively small part of the total number of accident claims.[100]

While the German scheme served as the spark for worldwide acceptance of the principles of workers' compensation, it was the British Act of 1897 that served as the model for American jurisdictions.[101] This Act made employers in certain industries responsible for compensating employees for injuries "arising out of and in the course of their employment."[102] Unlike the German plan, it did not mandate insurance coverage, although employers were free to purchase insurance. The Act of 1906 expanded coverage to nearly all workers.[103]

In a way similar to the German and British reforms before it, the new U.S. program (or programs) of workers' compensation promised remedies more reliable than the common law tort arrangements.[104] Modelled after the British plan, it sought to shift work

95. Sherman, *Workmen's Compensation*, *supra* note 92, at 857.

96. *Id.* at 858.

97. Sherman, *German Workmen's Insurance Law*, *supra* note 92, at 69.

98. *Id.*

99. *See* Digby, *supra* note 91, at 488; Sherman, *Workmen's Compensation*, *supra* note 92, at 859.

100. Sherman, *German Workmen's Insurance Law*, *supra* note 92, at 69.

101. *See* 1 LARSON, *supra* note 90, § 5.20; Rhodes, *supra* note 91, at 40.

102. *See, e.g.*, Sherman, *German Workmen's Insurance Law*, *supra* note 92, at 68.

103. *Id.*

104. *See* Arthur B. Honnold, *Theory of Workmen's Compensation*, 3 CORNELL L.Q. 264, 264-67 (1918). According to calculations made at the time workers' compensation laws were adopted in the U.S., fifty-two to fifty-three percent of all workers would be ineligible for any compensation under common law arrangements. This was due to the fact that a majority of cases were caused by workers' negligence or by "natural hazards." *See* James H. Boyd, *The Economic and Legal Basis of Compulsory Industrial Insurance for Workmen*, 10 MICH. L. REV. 345, 346-51 (1912).

place accident costs to employers.[105] Businesses were expected to adjust prices in order to pass on the expense to consumers.[106] The private provision of insurance played an important role in the U.S. system, and continues to do so today.[107] The U.S. system preserved a role for government supervision, most obviously in the resolution of disputes between employers and workers. However, in the vast majority of cases, claims are reviewed and benefits are distributed without state intervention. The current British plan no longer allows private or self insurance, and for this reason features a more prominent day-to-day administrative role for public authorities.[108]

The importance of the internalization principle as a theory that guided the employer liability reform movements is evident in the design of the U.S. workers' compensation system. However, there are other pieces of evidence that attest to the importance of the principle. One is the literature of the reform period containing disparaging references to the common law regime of employer liability.[109] Disapproval of the common law regime does not logically compel support for a no-fault workers' compensation statute. One could, for example, argue that an employee should not be in a worse position than a customer of the firm or a stranger.[110] But the basic objection to the common law regime was no different from the objection later made against the principle of contributory negligence: the victim often goes without compensation.[111] This objection does force one to embrace full internalization. If the negligence of the victim, or of others, should not be allowed to prevent a victim from receiving compensation from an injurer, whose behavior is causally connected to the injury, then a regime of strict liability is very difficult to avoid.

The second piece of evidence pointing to the importance of the internalization principle to the reformers is the development of a

105. *See* Boyd, *supra* note 104, at 346.

106. Eugene Wambaugh, *Workmen's Compensation Acts: Their Theory and Their Constitutionality*, 25 HARV. L. REV. 129, 131 (1911).

107. *See* 1 LARSON, *supra* note 90, § 3.10.

108. *Id.*

109. *See* Francis H. Bohlen, *A Problem in the Drafting of Workmen's Compensation Acts (Part I)*, 25 HARV. L. REV. 328, 331 (1912); Boyd, *supra* note 104, at 346-69; Honnold, *supra* note 104, at 264; J.G. Pease, *An English Workman's Remedies for Injuries Received in the Course of His Employment, at Common Law and by Statute*, 15 COLUM. L. REV. 509, 513 (1915); Jeremiah Smith, *Sequel to Workmen's Compensation Acts*, 27 HARV. L. REV. 235, 243 n.20, 256 (1914).

110. Smith, *supra* note 109, at 242.

111. *See* Boyd, *supra* note 104, at 345, 349; Smith, *supra* note 109, at 243.

theory of enterprise liability that justified the reform movement.[112] Enterprise liability theory (or theories) holds that businesses should be liable for injuries to workers in the course of production and to consumers of manufactured goods.[113] According to the theory, the injuries are a part of the cost of production and should be treated as such. The costs of injury will be spread, as are all costs of production, to consumers in the form of higher prices and to shareholders in the form of lower returns. One proposition that is common to any statement of a theory of enterprise liability is that the fault standard, or any standard of conduct, should not be used to deny compensation to accident victims. Enterprise liability emphasizes the insurance function of a liability system, and this function is always in tension with a system of fault standards that have the effect of limiting liability under certain conditions.

Further evidence on the importance of the internalization principle appears in the drafting of the earliest workers' compensation statutes in the United States. In order to avoid equal protection attacks, the early statutes claimed to limit the scope of the no-fault regime to industries that were unusually hazardous.[114] The reason for the limitation is that with respect to unusually hazardous activities, the principle of full internalization had been accepted by courts. It was embodied in the strict liability doctrine of *Rylands v. Fletcher*,[115] and received further support from several nuisance cases.[116]

The public acceptance of the internalization principle urged by proponents of workers' compensation statutes was, at the same time, an emphatic rejection of the doctrine stated by Justice Shaw in the most important opinion on the common law of employer liability, *Farwell v. Boston & Worcester R.R.*[117] Although the opinion is known for its statement of the "fellow-servant rule," that an employer is not liable for the injuries to an employee caused by the negligence of another employee, it is largely concerned with the more general

112. For a clear statement of the theory, see Wambaugh, *supra* note 106, at 129-31.

113. For an important critique of the theory of enterprise liability, see Priest, *supra* note 68.

114. *See* Smith, *supra* note 109, at 248-49 n.41.

115. 3 L.R.-E. & I. App. 330, 338-39 (H.L. 1868). For a discussion of strict liability and unusually hazardous activities, see William K. Jones, *Strict Liability for Hazardous Enterprise*, 92 COLUM. L. REV. 1705 (1992); Virginia E. Nolan & Edmund Ursin, *The Revitalization of Hazardous Activity Strict Liability*, 65 N.C. L. REV. 257 (1987).

116. *See* KEETON ET AL., *supra* note 32, §§ 22-25, at 137-49.

117. 45 Mass. (4 Met.) 49 (1842).

problem of defining the scope of the implicit contract between the employer and employee. Justice Shaw argued against requiring employers to compensate injured workers in all cases because employees would ordinarily be compensated ex ante for unusual risks accepted as part of a job.[118] In Shaw's view, the employment contract is in some respects like an insurance contract, and either party may play the role of the insurer.[119] If the employee accepts higher wages in order to bear the risk of injury then he effectively insures the employer against injury claims that do not arise from the employer's negligence.[120] If the employee demands to be compensated for all workplace injuries, then the employer will be forced to reduce wages by an amount that reflects the expected cost of compensating the employee. Shaw apparently believed, or felt, that courts should assume in the absence of obviously contradictory evidence, that the process of private contract making carried out this insurance function fairly well.

Implicit in Shaw's view of the employment contract is the assumption that employees are as informed as, or, in some cases, better informed than the employer of possible workplace hazards. Under this assumption, the bilateral insurance arrangement would be fair in the sense that the prices of the implicit insurance policies would fully reflect the underlying risks.[121] But if employees are not informed,

---

118. *Id.* at 56-57 ("[The maxim respondeat superior] does not apply to the case of a servant bringing his action against his own employer to recover damages for an injury arising in the course of that employment, where all such risks and perils as the employer and the servant respectively intend to assume and bear may be regulated by the express or implied contract between them, and which, in contemplation of law, must be presumed to be thus regulated." *Id.* at 56).

119. Specifically, if the employer is "risk neutral" (willing to spend as much as $1.00 for a 50-50 chance of winning either $0 or $2.00) and the employee is "risk averse" (unwilling to spend as much as $1.00 for a 50-50 chance of winning either $0 or $2.00), then under a contract in which risk is spread optimally, the employer will bear all of the risk of financial loss resulting from an injury. On the economics of risk-sharing contracts, see A. Mitchell Polinsky, *Risk Sharing Through Breach of Contract Remedies*, 12 J. LEGAL STUD. 427 (1983).

120. In this case the employee bears the residual risk of an accident, where by residual risk, we mean the risk of an accident occurring when the employer is exercising reasonable precaution. For example, the loss to the employee from an injury is $100, and the probability of an injury occurring is 1/100 over a period of one month when the employer complies with the due-care standard, then under an actuarially fair insurance contract the employee would bear the risk of injury in return for receiving $1 more in pay per month than he would otherwise receive. For a discussion of liability insurance and pricing, see Steven Shavell, *On Liability and Insurance*, 13 BELL J. ECON. 120 (1982).

121. The prices would fully reflect underlying risks because employees would not bear the risk of injury if they did not. The employee would either require the employer to com-

and the employer is better informed, then the implicit insurance contract will not work as well as Justice Shaw assumed. The uninformed workers will most likely accept too little in exchange for bearing unusual risks on the job.[122] As a result, the employer will, in effect, shift injury costs to workers, and have less incentive to invest in safety. In this case, a strict liability rule, such as that provided by the workers' compensation statutes, may correct a market failure that would otherwise have resulted in too many workplace injuries, and too many worker injuries going uncompensated.

Adoption of the workers' compensation statutes reflected at bottom an acceptance of a certain view of the relative informational advantages enjoyed by employees and employers. Specifically, it reflected a view that workers are generally not as well informed as are employers of potential workplace hazards. In the remaining sections of this Article, we will not question this assumption; we will take it as valid. The question we examine below is whether the goal of internalization that is justified in part by the assumption that employees are informationally disadvantaged with respect to workplace risks is satisfied by the workers' compensation system.

We are fully aware of the argument that shifting all accident losses to the employer is an undesirable goal in the workers' compensation area because it would leave too little incentive on the part of workers to minimize accident losses.[123] In spite of this argument,

---

pensate him for what we have referred to as "residual risk" injuries or would not accept the employment contract.

122. They will clearly accept too little if they underestimate the risks. Further, they are likely to underestimate the risks for two reasons. One is strategic. If the workers are required to purchase insurance from an employer who is presumably better informed, they are likely to undervalue the risk of loss. The reason is that they would in effect be asked to purchase an item (safety) that has a value that is unknown to them. The rational response is to start with a low bid. *See generally* RASMUSEN, *supra* note 88, at 251-53. The other reason for underestimating the risk is that knowledge of workplace hazards is an experience good. It takes time for workers to learn about the hazards, and they must learn about them on the job. In the learning period they are unlikely to bargain for the right compensation for accepting risks. *See* JOHN R. HICKS, THE THEORY OF WAGES 110-11 (1932); H. Lorne Carmichael, *Reputations for Safety: Market Performance and Policy Remedies*, 4 J. LAB. ECON. 458, 459 (1986). Though in general the empirical results are inconclusive, some empirical studies suggest that workers do underestimate job hazards. For example, Viscusi and Moore, ignoring the problem of worker incentives for safety, found that after taking compensating wage differentials into account, the level of benefits failed to compensate workers for their losses. *See* W. Kip Viscusi & Michael J. Moore, *Workers' Compensation: Wage Effects, Benefit Inadequacies, and the Value of Health Losses*, 69 REV. ECON. & STAT. 249 (1987).

123. This is a general problem with any strict liability scheme. *See* Brown, *supra* note 46, at 338-43. For a discussion of this problem in the workers' compensation context, see

we think an examination of the extent of cost shifting and of externalization under workers' compensation is interesting for several reasons. Shifting accident costs to employers seems to have been an important goal of the legislation. The workers' compensation system has gone further than, perhaps, any other legal regime in trying to shift accident losses from one party to another. An examination of the extent of accident cost externalization in this area should reveal the practical limits on our ability to shift such losses. Further, if employees are relatively uninformed about workplace hazards, then shifting costs to the relatively informed party, the employer, should minimize the total costs of accidents by giving employers an incentive to take care and to inform employees about worksite hazards.[124]

B. An Evaluation of Obstacles to Internalization

1. The Role of Private Insurance

Six states—Ohio, Nevada, North Dakota, Washington, West Virginia, and Wyoming—do not permit private carriers to offer workers' compensation insurance; these six states, instead, operate state managed insurance funds.[125] Each of the remaining forty-four

---

Richard J. Butler & John D. Worrall, *Work Injury Compensation and the Duration of Nonwork Spells*, 95 ECON. J. 714 (1985); Richard J. Butler & John D. Worrall, *Workers' Compensation Benefit and Injury Claims Rates in the Seventies*, 65 REV. ECON. & STAT. 580 (1983); James R. Chelius, *The Influence of Workers' Compensation on Safety Incentives*, 35 INDUS. & LAB. REL. REV. 235 (1982); James R. Chelius & Karen Kavanaugh, *Workers' Compensation and the Level of Occupational Injuries*, 35 J. RISK & INS. 315 (1982); Bruce D. Meyer et al., *Workers' Compensation and Injury Duration: Evidence from a Natural Experiment* (1990) (unpublished manuscript, on file with the *Hofstra Law Review*); John D. Worrall & David Appel, *The Wage Replacement Rate and Benefit Utilization In Workers' Compensation Insurance*, 49 J. RISK & INS. 361 (1982). These papers generally demonstrate that increases in benefit levels are followed by increases in the frequency and duration of injury claims. The most thorough recent study of the incentive effects of workers' compensation is provided in MICHAEL J. MOORE & W. KIP VISCUSI, COMPENSATION MECHANISMS FOR JOB RISKS: WAGES, WORKERS' COMPENSATION, AND PRODUCT LIABILITY (1990).

124. The notion that informational asymmetry might justify strict employer liability was first examined rigorously in Oliver E. Williamson et al., *Externalities, Insurance, and Disability Analysis*, 34 ECONOMICA 235 (1967). The theory has been developed further in Peter Diamond, *Insurance Theoretic Aspects of Workers' Compensation*, *in* NATURAL RESOURCES, UNCERTAINTY, AND GENERAL EQUILIBRIUM SYSTEMS 67 (Alan S. Blinder & Philip Friedman eds., 1977) and Samuel Rea, Jr., *Workmen's Compensation and Occupational Safety Under Imperfect Information*, 70 AM. ECON. REV. 80 (1981).

125. It should be noted that other states (eighteen, to be precise) also offer businesses the option of buying workers' compensation insurance from a state managed fund. However, these states allow their funds to stand alongside private funds and permit private insurers to compete for employers' business. Puerto Rico and the U.S. Virgin Islands also operate exclusive government managed funds. *See* U.S. DEP'T OF LABOR, STATE WORKERS' COMPEN-

states (as well as the District of Columbia and Guam) establish specific statutory guidelines to police the insurance industry, and to regulate the pricing of workers' compensation insurance. For the present we will set aside the idiosyncrasies of each state's approach to oversight. Regulatory intervention generally seeks two goals: to assure that prices charged by insurance companies remain at "reasonable" levels and to help insure solvency among insurance funds.[126] Additionally, states act to provide for continued payments in the event that failure occurs.[127] Regulatory oversight extends, as well, to cover those businesses that choose to self-insure. Self-insurance is an option available only to the largest employers.[128]

The insurance industry is peculiar in that it is largely exempt from federal antitrust legislation.[129] To a great extent, the federal government has ceded to the states the right to regulate insurance pricing and industry practice.[130] Two approaches are presently employed to regulate pricing: prior approval and policed open competition.[131] Only a small handful of states now allow open competition in the area of workers' compensation.[132] Because workers' compensation is viewed as a quasi-social insurance, states have not been as willing to open up the system to the potential uncertainties of open competition as they have been with other types of insurance.[133] In-

---

SATION LAWS tbl. 1 (1992) [hereinafter STATE COMPENSATION].

126. C. Arthur Williams, Jr., *Workers' Compensation Insurance Rates*, *in* CURRENT ISSUES IN WORKERS' COMPENSATION 209, 225 (James Chelius ed., 1986).

127. States are especially likely to be concerned with workers' compensation insurance arrangements, since the program is seen as a form of quasi-social insurance. *See id.* at 225.

128. JEFFREY V. NACKLEY, PRIMER ON WORKERS' COMPENSATION 6 (2d ed. 1989).

129. DOUGLAS CADDY, LEGISLATIVE TRENDS IN INSURANCE REGULATION 3 (1986). This means that insurers are free to exchange information about risk and to establish uniform rates through bureaus. Insurers are only brought under scrutiny when they engage in coercive or intimidating practices, when they fail to be properly regulated by state authorities, or when two large national firms seek merger. *See id.* at 43.

130. In response to the Supreme Court's holding that the business of insurance was a part of interstate commerce subject to the Sherman Antitrust Act, United States v. South-Eastern Underwriters Ass'n, 322 U.S. 533 (1944), Congress reasserted the preeminence of state regulation of insurance by enacting the McCarran-Ferguson Insurance Regulation Act, 59 Stat. 33 (1945) (codified as amended at 15 U.S.C. §§ 1011-1015 (1988)). The McCarran-Ferguson Act exempted the business of insurance from antitrust scrutiny under certain circumstances, and declared that "the continued regulation and taxation by the several States of the business of insurance is in the public interest." 15 U.S.C. § 1011 (1988).

131. Williams, *supra* note 126, at 216-17. Texas is unique in that rates are determined by a state board. *Id.* Still, this could be considered an extreme variant of the prior approval arrangement.

132. The open competition states include Arkansas, Georgia, Illinois, Kentucky, Michigan, Minnesota, Oregon, and Rhode Island. *See id.* at 217.

133. *See id.* at 225. Twenty-one states allow open competition in property and liability

deed, it is improper to describe any state as practicing "open" competition in workers' compensation, since regulatory oversight is still in place in each state. To begin with, all states have some control over which carriers are allowed to operate in the state.[134] Oregon and Arkansas require insurers to file all rate plans with state commissioners.[135] Georgia and Minnesota have excess profit statutes which prohibit insurers from reaping profits beyond an established threshold. Florida specifically outlaws excessive profits in workers' compensation insurance.[136] With the exception of Minnesota, all of the states practicing open competition in workers' compensation allow rate bureaus to establish "advisory rates."[137] These target rates serve to shape pricing, although companies are in no way required to match them.[138] Michigan, Minnesota, and Oregon each operate state insurance funds which compete with private carriers for business.[139] This option provides employers with a low-cost alternative. While it is generally true that state funds fail to offer services comparable to larger private carriers (such as loss control programs, year-end rebates, and medical expert testimony), state funds encourage price competition. An inexpensive option will probably remain attractive to smaller firms in particular; smaller firms are often poorly situated to take great advantage of loss control/accident prevention programs offered by larger insurers.

An overwhelming majority of states demand that insurers seek prior approval for all rate plans.[140] In most of these states, rate plans are prepared by rate bureaus, which present rates to the state insurance commission for final approval.[141] Once the state's insurance commission approves a rate schedule, companies can petition for permission to deviate from the rate plan.[142] This arrangement—the

---

insurance, while only eight have opened up competition in the workers' compensation field. *See id.* at 223.

134. U.S. DEP'T OF LABOR, STATE WORKERS' COMPENSATION: ADMINISTRATION PROFILES (1990) [hereinafter ADMINISTRATION PROFILES]. As a minimum, companies must comply with fair practice standards which guide all business activities in the state. In most states, carriers must meet additional minimum requirements.

135. Williams, *supra* note 126, at 222.

136. *Id.* at 232.

137. *Id.* at 218-19.

138. *Id.* at 217.

139. *See* ADMINISTRATION PROFILES, *supra* note 134, at 186, 194, 317.

140. *Id.*; *see also* Williams, *supra* note 126, at 218-19.

141. Williams, *supra* note 126, at 216-17.

142. *Id.* at 220. California, Massachusetts, Missouri, New Jersey, North Carolina, Pennsylvania, and Wisconsin are among the states that do not allow deviations.

allowance of modest deviation from approved rates—introduces opportunities for limited competition into the system. Insurers, perhaps with justification, complain that prior approval arrangements interfere with the efficient management of business.[143] Because rates must be approved in advance, and remain relatively fixed, insurers are not situated to respond with immediacy to shifting loss experience.[144] It should also be pointed out that prior approval arrangements involve higher levels of state intervention, and, as a result, carry greater administrative costs.

There are three basic approaches to setting prices in workers' compensation insurance: retrospective rating, class rating, and experience rating.[145] Actually, it should be recognized that class rating and experience rating are often employed side-by-side within the insurer's rate fixing calculation.[146] Experience rating is, in fact, a "way of modifying these class rates to take into account the specific firm's injury history or experience."[147] This interaction will be described in a moment, but first we will introduce the basic elements of retrospective rating. Retrospectively rated companies pay a premium that is determined by their loss performance over the previous year. That is, at the conclusion of the policy period, the company pays the insurer an amount equal to its total losses plus, of course, an additional sum sufficient to provide the insurer with payment for service.[148] Retrospective rating is unique among the three systems to be discussed because it alone assesses premiums not on the basis of expected losses over the policy period ahead, but according to the actual losses incurred in the period just passed. The insurer renders compensation to injured workers, carries the administrative expenses attached to the dispensation of benefits, and, at the conclusion of the policy period, prepares a bill for the full cost of these services—inserting, of course, a profit margin. Retrospective rating, in a sense, provides a sort of "quasi-self-insurance."[149] There is one important qualification: the insurer often requires employers to pay a

143. *Id.* at 224-25.

144. CADDY, *supra* note 129, at 49-50.

145. Williams, *supra* note 126, at 210-14.

146. *Id.* at 213.

147. Richard B. Victor, *Experience Rating and Workplace Safety, in* WORKERS' COMPENSATION BENEFITS: ADEQUACY, EQUITY, AND EFFICIENCY 71, 72 (John D. Worrall & David Appel eds., 1985) [hereinafter Victor, *Experience Rating*].

148. Williams, *supra* note 126, at 213.

149. *Id.* at 214.

minimum fee, although in exchange, the employer is given the promise that,however costly losses may be, the employer's fee will not exceed a maximum.[150] Accordingly, in a year with very low losses—that is, with only a few accidents resulting in fairly small claims—the employer pays a premium that, perhaps, actually exceeds losses. Yet in a less fortunate year—filled with many accidents and costly claims—the employer may actually reduce the cost of his losses. This is a benefit not available to the self-insured.

The two more frequently used approaches to price setting are class rating and experience rating.[151] Under a class rated system, all employers whose businesses fall within a particular class of industry pay an identical rate.[152] In prior approval states, these rates must be approved before the insurer can use them to calculate premiums.[153] An individual employer's premium is determined by the size of his payroll: generally, employers pay a specific industry-fixed rate per $100 of payroll.[154] For example, an employer with a $1,000,000 payroll might pay $1,200 per year for insurance (an unlikely $0.12 per $100 of payroll), while another employer, in the same industry, with a payroll of $10,000,000, would pay $12,000 a year in premiums. Different rates are developed for each industry according to the incidence (and cost) of work related injury.[155] In practice, insurance companies take into account the various types of labor undertaken under the same roof. For example, a construction company will often employ a clerical staff as a well as crane operators.[156] Clerical work, of course, is usually much less risky than operating a crane at a construction site. Accordingly, insurers will calculate premiums not only according to industry type, but also according to job description. Imagine that a construction company has a total payroll of $200,000 for construction site workers and $50,000 for clerical office employees.[157] Assume that the class rate for construction work is $2.50 per $100 of payroll, and the class rate for clerical work is $0.25 per $100 of payroll. The company's premium would be *2,000($2.50) + 500($0.25)*, or $5,125.

150. *Id.* at 213.
151. *Id.* at 211.
152. *Id.* at 210-11.
153. *Id.* at 222.
154. *Id.* at 210-11.
155. *Id.* at 215-16.
156. *Id.* at 210-11.
157. *See id.* at 211.

2. Experience Rating

This would be a suitable arrangement if all competing businesses within a particular industry were equal in their acceptance of risk and commitment to safety (and equally lucky). If this were true, class rating would provide a useful mechanism for establishing prices that accurately reflect each company's loss-experience. Obviously, some employers enjoy better safety records than others operating within the same industry. Insurance companies need to put into place some procedure for identifying these differences and adjusting prices. Under experience rating, an employer's class premium is modified to reflect the firm's specific injury experience. As a standard, the employer's loss experience over the preceding three year period is reviewed and compared to the "average" loss profile expected for a company engaged in that particular type of industry. For example, if an employer's accident losses were only half of the industry's expected average, his premium would be trimmed back to half that paid by the "average" employer in that industry.

However, insurers recognize that through the intervention of a variety of factors (including chance), accident losses from year to year may vary widely. That is, the loss experience observed over any particular three year period may not accurately reflect an employer's proper loss expectation. Accordingly, insurers introduce into the calculation a "credibility factor."[158] The smaller the firm, the more important the use of this credibility factor. The loss experience of larger firms reveal a statistical regularity from year to year.[159] Smaller firms, especially those engaged in less dangerous industries, are especially susceptible to the play of chance. For example, a bakery which employs ten people may not experience any significant losses due to workplace accidents over a five or seven year period.[160] However, without explanation, three or four workers may be seriously injured over the course of a two year period (say two are badly burned in a gas-oven explosion and two others are injured the following year in separate kitchen accidents). Following this atypical sequence of inju-

158. *Id.* at 212.

159. *Id.* at 212-13; *see also* RICHARD B. VICTOR, WORKERS' COMPENSATION AND WORKPLACE SAFETY: THE NATURE OF EMPLOYER FINANCIAL INCENTIVES ix (1982); Victor, *Experience Rating*, *supra* note 147, at 74.

160. This is a simplified example presented merely to provide illustration. It is perhaps likely that in reality this firm would pay a class premium of less than $2,500 a year and, as we explain below, would therefore not be eligible for experience rating.

ries, a straight experience based calculation focusing on a review of the past three years would result in a large jump in the bakery's insurance fee. The introduction of a credibility factor would yield a more appropriate calculation of expected loss for the year ahead. It is important to remember that experience rating schemes emphasize a forward-looking calculation, that is, insurers wish to predict losses and costs in the year ahead, not to penalize for losses in the years passed. Imagine that a moderate-sized employer has a credibility rating of twenty percent, and over the course of the period under review the company experiences only half the losses expected for an "average" company of its size engaged in that particular industry.[161] Instead of reducing the company's premiums for the year ahead by one half, the insurer would reduce the fee by an amount equal to (50%) x (20%), or ten percent. That is, the experience component (50% of expected losses) would be multiplied by the credibility factor (20%) to yield the percentage of adjustment in the employer's fees (in this case, a downward adjustment). As one observer explains, "[i]n practice, insurers assign no credibility to the experience of employers with average class premiums of less than $2,500."[162] Above that, threshold credibility increases gradually from one to one hundred percent.[163] Of course, as this suggests, the premiums of very large firms are almost wholly determined by their experience rating, without the meaningful intervention of a credibility component.[164] The effect is not unlike that arrived at through retrospective rate calculations: the employer pays the "full" expense of shop floor accident costs. The difference, of course, is that the retrospective rate fixing arrangement sets prices with reference to past experience, while experience rating shapes prices according to expected losses in the year ahead. This means that only large firms carry the "full" burden of their losses arising from work-related accidents. Small businesses may pay less (or, in some cases, more) than the full cost of their annual losses because class rates fail to accurately reflect the specific dimensions of their year to year loss experiences.[165]

---

161. *See* Williams, *supra* note 126, at 212.

162. *Id.*

163. *Id.*

164. Victor, *Experience Rating*, *supra* note 147, at 72-74.

165. Empirical studies demonstrate that the incentive effects of experience rating are negligible for small firms. *See* John W. Ruser, *Workers' Compensation Insurance, Experience-Rating, and Occupational Injuries*, 16 RAND J. ECON. 487, 493 (1985); Louise B. Russell, *Safety Incentives in Workmen's Compensation Insurance*, 4 J. HUM. RESOURCES 361,

The vast majority of shop floor injury claims are processed, and benefits are paid, without the intervention of the state accident board.[166] It is difficult (and may be impossible) to specify exactly what percentage of claims are paid without challenge. There are differences from company to company and between states, and different types of claims are challenged at different rates. It seems the more costly the claim or the more ambiguous the disability, the more likely it is that an employer will file a challenge.[167] One can, however, offer a fairly reliable thumbnail estimate. Nationally, about eleven percent of all work place accident claims are challenged.[168] This means that eighty-nine percent of all work place accident claims are processed—and, in day-to-day reality, probably somewhat more—and reimbursement is provided, without the involvement of state arbitrators or hearing officers.[169] In short, administrative costs are carried by the insurer and passed on to the employer in the form of the service fee charged to every customer as a part of his or her premium. Private insurance arrangements influence the internalization of loss by shaping premiums (primarily for large businesses) through the use of retrospective-rating and experience-rating to more accurately capture an employer's true loss experience,[170] and by shifting most

363 (1974).

166. *See* ADMINISTRATION PROFILES, *supra* note 134, at 481-88 (providing nationwide figures of contested case information).

167. For a discussion of this conclusion, see LESLIE I. BODEN, REDUCING LITIGATION: EVIDENCE FROM WISCONSIN 34-36 (1988). Boden offers other possibilities, but extends recognition to the plausibility of this conclusion. *See also* LINDA DARLING-HAMMOND & THOMAS J. KNIESNER, THE LAW AND ECONOMICS OF WORKERS' COMPENSATION 32-36 (1980); David E. Bloom, *Is Arbitration Really Compatible with Bargaining?*, 20 INDUS. REL. 233, 233-36 (1981); Karen Roberts, *Predicting Disputes in Workers' Compensation*, 59 J. RISK & INS. 252, 255 (1992).

168. *See* ADMINISTRATION PROFILES, *supra* note 134, at 481-88. Annually, among states providing data, about 2.4 million claims were filed, while nearly 266,000 hearings were requested. Another report offers a figure of nearly twelve percent. This figure represents an estimate drawn from a sample of states. *See* JOHN H. LEWIS, ILLINOIS WORKERS' COMPENSATION SYSTEM: A REPORT TO THE GOVERNOR 82 (1989) (providing National Council on Compensation Insurance data).

169. Many shop floor accident victims are promptly treated by company physicians and sent home with pay for a few days to recover. Such incidents, however, do not appear on state industrial accident board reports if, as is likely, no formal claim is filed.

170. For a discussion of this relationship, see THE REPORT OF THE NATIONAL COMMISSION ON STATE WORKMEN'S COMPENSATION LAWS 93-98 (July 1972) [hereinafter NATIONAL COMMISSION REPORT]. The Commission's findings call into question certain common expectations attached to experience-rating. The Commission found unconvincing evidence that experience-rating shapes company safety performance. However, the Commission did place enough confidence in experience-rating that it recommended that experience-rating be more

administrative costs to employers through the service fee charged by insurers.

3. Obstacles to Internalization in the Insurance System

In evaluating obstacles in the insurance system to the internalization goal of workers' compensation, two general areas of the insurance market should be examined. One is the regulation of competition and pricing in the insurance industry. The other is the pricing itself; whether, for reasons unrelated to regulation, it accurately reflects risk.

a. Regulation of Competition and Pricing

Regulation of competition and pricing in the insurance industry can be subdivided into two categories: structural regulation, or laws requiring "structural compliance"; and operational regulation, or laws requiring "operational compliance." Structural regulations must be satisfied by any insurer before it can operate lawfully. In this category fall most types of minimum capital requirement. Exclusive state provision of insurance is another type of structural compliance requirement, since it eliminates private provision of workers' compensation insurance. Excess profits statutes, like those in place in Minnesota and Georgia, may also be considered structural compliance regulations, though this is admittedly on the borderline between operational and structural regulation. Laws requiring operational compliance are those that apply to the day-to-day operations of an insurer. Prior approval is one type of operational compliance requirement. The provision of a state run insurance program, as an alternative to private insurance, is also a form of operational compliance, since the state fund will set price standards that put pressure on private insurers.

Structural and operational compliance requirements affect externalization of accident costs largely through limiting the ability of insurers to match prices with costs. In a world of perfect insurance pricing, the price per dollar of coverage of a policy would be equal at all times to the probability of an accident (which is the expected cost per dollar of coverage).[171] Competition in the insurance market tends to drive the price per dollar of coverage toward the probability of an accident, and in this sense, drives prices toward costs. If prices

widely used. *See also* Victor, *Experience Rating*, *supra* note 147, at 83-85.

171. *See, e.g.*, SHAVELL, ECONOMIC ANALYSIS, *supra* note 15, at 203.

are not equal to costs, then externalization results because insured parties do not bear the full costs of their accident-causing behavior.

Consider the externalization effects of the following types of structural regulation: banning private insurers, minimum capital requirements, and excess profits statutes. The outright banning of private insurers could lead to the worst form of externalization because a state run monopoly insurer would not have competitors to push its prices in line with its costs. There are two reasons for this: the obvious one is that competition forces firms to lower their prices; the less obvious reason is that it is hard to identify the administrative costs of running an insurance firm in the absence of competition.[172] Further, without the profit motive, there is little incentive for a public insurer to align prices with costs. It may seem intuitive that the profit motive would lead a monopoly insurer to charge the highest price possible.[173] But there are reasons to think that a monopolist in the insurance market would not simply charge the highest price possible. First, the best pricing strategy for such a firm would be a two-part price consisting of a flat charge to all customers and a price per dollar of coverage set at roughly the competitive rate.[174] Second, any profit-oriented insurer would be concerned about the incentive effects of a flat monopoly price charged to all customers. In order to give customers the proper incentives to avoid losses, the price would have to vary with the level of risk created by the insured, and this, other things equal, will put a constraint on the firm's power to charge the highest price possible.

Minimum capital requirements present a barrier to entry and therefore restrict competition, which eliminates some of the pressure that drives price toward cost in each line of insurance. Of course, they also eliminate undercapitalized, fly-by-night insurers, so it is impossible to say that the net effect on the welfare of insurance consumers is negative. We can only say unambiguously that capital requirements are not a necessary part of a consumer protection plan. Consumers can be equally well protected by certification procedures

172. In the absence of competition, monopoly rents tend to be transferred into "costs" as unions and other constituencies bargain against shareholders for shares of the rent. *See* DONALD DEWEY, MONOPOLY IN ECONOMICS AND LAW 249-51 (1959).

173. More precisely, the monopoly will charge the "monopoly price." For an introductory discussion of the theory of monopoly, see GEORGE J. STIGLER, THE THEORY OF PRICE 210-14 (1952).

174. Walter Y. Oi, *A Disneyland Dilemma: Two Part Tariffs for a Mickey Mouse Monopoly*, 85 Q.J. ECON. 77 (1971).

or any information gathering agency that allows consumers to obtain reliable information on the reputations of insurers.[175]

The excess profits statutes affect externalization through a more subtle route. If the rate of return is defined as the return on capital, such statutes create an incentive for insurers to use capital intensive methods of operation.[176] To the extent that these are not cost-minimizing, the overall operating costs of regulated firms will be greater. These costs will be passed on to the consumer to some extent in the form of higher prices. Thus, regulation of this form may widen that portion of the wedge between price and cost that is attributable to administrative costs.

Operational compliance requirements have similar effects on the pricing of insurance. Prior approval limits the ability of each insurer to change rates immediately to reflect new information, positive or negative, on potential claims. The result is a set of rates that are smoother over time and that less closely track underlying costs than would be observed in a truly free or open competition regime. To see why, consider the following simple example. Suppose unit cost is $1 in period 1 and $10 in period 2; and suppose that the insurer will be able to win approval for a higher price only after one half of the second period has passed. Suppose further that the interest rate is zero. Then, assuming a uniform rate of service, to break even, the insurer's first period price would have to be: *$1(2/3) + $10(1/3) = $4.*

The existence of a state run program has an effect on prices which may result in further externalization of accident costs. In the presence of a guaranty fund designed to cover the claims of bankrupt insurers, a state run insurance provider may give some insurers an

175. Note that the same argument may be made about licensing. The purpose is always to protect consumers who cannot determine product quality easily. Further, consumers have too little incentive to gather information on product quality because such information is a public good. Once one consumer incurs the cost of gathering information, the information can be made available to all consumers almost costlessly; and given this incentive structure, no rational consumer would go out and collect information. But consumers can be equally protected by a certification process and the public good problem can be solved by having the state provide information. None of this requires the state to limit entry. The economic case against licensing was first put forward by John Stuart Mill in the final chapter of *On Liberty*. *See* MILL, *supra* note 40; *see also* MILTON FRIEDMAN, CAPITALISM AND FREEDOM 137-60 (1962) (elaborating on the argument against licensing).

176. This is the familiar "Averch-Johnson" effect of the industrial organization literature. For the original discussion of the effect, see Harvey Averch & Leland L. Johnson, *Behavior of the Firm Under Regulatory Constraint*, 52 AM. ECON. REV. 1052 (1962).

"excessive" incentive to underprice in order to compete with the state run insurer. This will tend to push price below cost and result in a larger number of claims than would be observed in the absence of such pressures. The reason the incentive to underprice may be excessive is that the guaranty fund eliminates much of the downside risk for the insurer. At the same time, the state insurance provider may force the private insurer to cut prices drastically in order to hold on to its customers. The end result is something analogous to the problem caused by federal deposit insurance in the banking industry:[177] private firms have too great an incentive to engage in risky activities (risky insurance contracts in this case, risky lending in the banking case). The state run insurer, by competing at subsidized rates which are difficult for private firms to match, provides a further push in the direction of excessively risky activity.

The likely effect of all of this regulation on accident cost externalization is unclear. Licensing and minimum capital requirements limit entry and therefore lead to supracompetitive prices, which implies that accident costs will be overinternalized. By making prices rigid, prior approval should lead to overinternalization followed by underinternalization when costs are expected to rise, and conversely when costs are expected to fall. Of course, if because of political pressure, insurance commissions force firms to set price below cost, externalization will result.[178] Guaranty funds and state provided insurance both tend to encourage firms to set prices below cost, which results in underinternalization.

b. Externalization and Experience Rating

If there is a common argument concerning experience rating, it is that there is too little of it.[179] The typical complaint is that the insurance industry, for some reason, does not use experience rating as much as it should. As a result, prices do not reflect costs as closely as they would in a regime of liberal experience rating. If experience rating were perfect and costless, this position would be unassailable.

---

177. *See* EDWARD J. KANE, THE S & L INSURANCE MESS: HOW DID IT HAPPEN? 2-9 (1989) (describing adverse incentive effects of federal deposit insurance leading to excessively risky behavior).

178. On the politicization of state regulation of workers' compensation prices, see Richard J. Butler & David Appel, *Benefit Increases in Workers' Compensation*, 56 S. ECON. J. 594 (1990).

179. *See, e.g.*, NATIONAL COMMISSION REPORT, *supra* note 170, at 98.

It would be difficult to understand why insurance companies use experience rating as little as they do.

But experience rating is not costless. There are two types of cost that are fairly important. One is the administrative cost of running such a system—which is significant. The other involves the set of costs that are generated by the adverse incentives introduced by an experience rating system. The adverse incentive problem is one that has received far too little attention in the literature to date. The problem is that once an experience rating system is put into place, the insured has every incentive to minimize reported claims in order to keep rates from increasing. For example, a company would have an incentive under experience rating to set up an internal administrative board to review claims before they reach the insurer and suppress or in some other way pay off those that could be settled cheaply. Terence Ison argues that it is claims control, not accident prevention, that is promoted by shifting accident cost responsibility to employers.[180] Prevention is costly. Investments must be made in new equipment, training programs, and supervision. If the number of injury claims can be reduced without the investment associated with the adoption of these measures, then the firm can enjoy the benefits offered by reduced premiums without burdening itself with any significant new costs. Accordingly, firms may try to discourage injured workers from reporting injuries, they may pressure supervisors to resist or denounce workers' claims, or they may formally challenge more claims, hoping to reduce the amount of the award.[181]

One would imagine that in most work sites, the largest number of claims would fall into this "claims control" category. The problem with this is that the insurer would then, in effect, be denied information on the true claims experience of the company, and its rates would therefore be inadequate. In situations where the costs of monitoring are extremely high, it is possible that the insurer would prefer some alternative pricing arrangement which did not have the perverse incentive effect of experience rating.

Our view is that in light of the underlying economics, there is too little evidence to support the conclusion that, in the absence of government regulation, the inability of insurers to set prices that accurately reflect underlying risks is a cause of large scale externalization

180. Terence Ison, *The Significance of Experience Rating*, 24 OSGOODE HALL L.J. 723, 725-26, 729 (1986).

181. *Id.* at 725-26.

in the insurance market. Note that this was one of the major claims of Calabresi in his critique of the tort system.[182] The contract between the insured and the insurer provides incentives for both to aim to increase precaution whenever it is less costly than paying the claims for accidents. Pricing is one method of providing an incentive for the insured to adopt all cost-justified precautions, but it is not the only method. The insurer can ask or demand that the insured invest more in safety enhancement. The insurer will use one of its instruments, price or direct bargaining backed up by threats to deny or to not renew coverage, to encourage the insured to increase precaution.

The view that imperfect insurance pricing inevitably leads to a dilution of incentives to take care is based on an inaccurate, though widespread, view of the function of insurance firms. The traditional notion is that the insurance firm simply pools risks. But the insurance firm is often a source of information for the insured.[183] To consider the simplest example, suppose the insurance firm is better at calculating the risk of an accident occurring than is the insured. Then enhanced precaution might be the result of simply passing on information on the relationship between precaution and accidents to the insured. There are many other examples in which the insurer stands between two contracting parties, such as manufacturer and a shipper. Obviously, the manufacturer and shipper could in many cases arrange their own insurance between themselves. Yet we still see these parties purchasing insurance. The likely reason is that the insurance provides information that the parties are unwilling to trust each other to provide. It is quite inaccurate in this setting to conclude that insurance tends to dilute incentives to take care. The information provision role may dominate the other functions in a way that leads to a net enhancement in incentives for parties to make precautionary investments.

4. Legal Challenges and the Externalization of Costs

a. Costs Related to Legal Challenges

When claims are challenged, the state industrial accident board (or a similar body) retains authority for resolving disputes and setting benefit levels.[184] If initial resolution efforts fail, most states allow

182. CALABRESI, *supra* note 2, at 239-308.

183. *See* Göran Skogh, *The Transactions Cost Theory of Insurance: Contracting Impediments and Costs*, 56 J. RISK & INS. 726 (1989).

184. *See* ADMINISTRATION PROFILES, *supra* note 134, at 462-80 (providing a full listing of such bodies).

claims to be appealed to the courts.[185] When parties to a dispute turn to the arbitration process (or later to the courts), they are likely to seek legal representation. As indicated earlier, about eleven percent of all workers' compensation claims are appealed to at least the first stage of the resolution process.[186] An even larger number of cases—about twenty-five percent of the whole—feature the involvement of attorneys.[187] Why do disabled workers secure the services of attorneys even in cases where challenges are not filed? The answer, it seems, is that workers' compensation laws too often seem an impenetrable tangle to disabled workers.[188]

What do legal challenges translate into in terms of costs? Exact figures are difficult, if not impossible to isolate. However, a sample of figures from several states may provide a hint. In Illinois, for example, among cases featuring attorney involvement, workers pay, on average, $1,975 in legal fees per case.[189] Legal fees should be understood to combine attorneys' fees and medicolegal expenses connected with medical evaluation and testimony. Employers can be imagined to pay a similar amount, or perhaps a little more. All in all, litigation costs for all cases in Illinois amounted to $178,000,000 in 1988, or a little more than fourteen percent of the total sum of benefits distributed to workers.[190] In neighboring Wisconsin, workers pay an average of $3,370 in legal fees per case, while employers pay a considerably greater $8,010 in fees.[191] Altogether, litigation costs in Wisconsin constitute about twenty-five percent of the cost of the average claim.[192] In Texas, a state often characterized as unusually litigious,[193] workers on average spend $3,661 per case on litigation costs, while employers spend about $1,500 more.[194] Again, as was the case with costs in Wisconsin, litigation expenses in Texas make

185. *Id.* at 481-88.

186. *Id.*; *see also supra* note 168 and accompanying text.

187. *See* LEWIS, *supra* note 168, at 82.

188. *Id.* at 105-06.

189. *Id.* at 83.

190. *Id.*

191. SARA R. PEASE, PERFORMANCE INDICATORS FOR PERMANENT DISABILITY: LOW-BACK INJURIES IN WISCONSIN 19 (1987). Figures are for permanent disability cases.

192. *Id.*

193. PETER S. BARTH ET AL., WORKERS' COMPENSATION IN TEXAS: ADMINISTRATIVE INVENTORY 65 (1989).

194. SARA R. PEASE, PERFORMANCE INDICATORS FOR PERMANENT DISABILITY: LOW-BACK INJURIES IN TEXAS 21-22 (1988).

up twenty-five percent of the cost of the average claim.[195] Finally, we turn to New Jersey. New Jersey is unusual in that employers are usually required to bear half the cost of workers' legal fees.[196] Perhaps as a result, nearly one hundred percent of all disabled workers contesting claims seek legal representation.[197] Still, friction costs directly linked to litigation per case are lower in New Jersey than in Texas and Wisconsin, and are about the same as costs in Illinois—workers' legal fees average $1,559 per case (with employers paying half of that total), while employers, on average, carry separate legal costs of $2,375.[198] Legal costs make up eighteen percent of the cost of the average claim.[199]

Large or costly claims are particularly likely to be challenged.[200] There are a variety of standard appeals available to employers (and their insurers). To begin with, injuries must be clearly linked to a worker's employment.[201] This seems straightforward enough: either the worker was or was not injured on the shop floor. But it is not so simple. Lawyers, responding to a recent survey prepared by the authors, point out that it is often difficult to prove that compensation is due when accidents occur without a witness present.[202] Employers will generally contend that the injury occurred outside of the work place, and that the employee's claim is spurious. Employers will also question whether or not a worker was serving in a work-related capacity at the time of injury. Workers who were engaged in some activity outside of their normal duties at the time of injury may find their claims challenged, the assumption being that the worker had chosen to engage in the activity without the knowledge, authorization, or supervision of his or her employer. While compensa-

---

195. *Id.* at 21.

196. SARA R. PEASE, PERFORMANCE INDICATORS FOR PERMANENT DISABILITY: LOW-BACK INJURIES IN NEW JERSEY 20 (1987) [hereinafter PEASE, NEW JERSEY].

197. *See* ADMINISTRATION PROFILES, *supra* note 134, at 255. In many cases, workers involved in disputed claims must by law secure the services of an attorney. *See also* PEASE, NEW JERSEY, *supra* note 196, at 18.

198. PEASE, NEW JERSEY, *supra* note 196, at 19-20.

199. *Id.* at 19.

200. On the determinants of the likelihood of a challenge, see Philip S. Borba & David Appel, *The Propensity of Permanently Disabled Workers to Hire Lawyers*, 40 INDUS. & LAB. REL. REV. 418 (1987); Roberts, *supra* note 167, at 253.

201. Joseph H. King, Jr., *The Exclusiveness of an Employee's Workers' Compensation Remedy Against His Employer*, 55 TENN. L. REV. 405, 406, 419 (1988).

202. Information drawn from responses and personal communications between authors and Illinois attorneys. Because the number of respondents was low, this reflects not a conclusion, but the opinion of experienced attorneys with years of practice in the Illinois system.

tion is due even in cases where the worker plays a contributing role in his or her injury, compensation is not due if a working relationship is not established, or if the injury occurs outside of such a relationship.[203]

Workers will find particular types of disabilities difficult to link with certainty to their employment in the work place. Claims associated with occupational diseases or mental illnesses which develop slowly, or which can also be linked with causes outside of the work place, are often successfully challenged by employers.[204] Such medical conditions include: lung cancer, hearing loss, blindness, heart disease, radiation sickness, lead poisoning, asbestosis, and stress disorders. Claims of this type are frequently targeted for appeal for three reasons: (1) they are often extremely costly (since they generally involve permanent disability), (2) they involve afflictions that can be caused by a variety of environmental factors completely unrelated to employment, and (3) the extended period of latency associated with some of these afflictions can cause workers to delay filing a claim.[205]

This final point is significant because most states have filing deadlines that require workers to put forward a claim within a specified period of time following an accident.[206] Failure to do so can result in the dismissal of a worker's claim for compensation.[207] Most states require that workers notify their employer of their disability within thirty days of the accident, and that an official claim be filed with the state industrial accident board within a year or two.[208] In cases where exposure to harmful chemical or radioactive contamination produces illness years later, workers will find that their claims very likely will be challenged by their employers. In many states, industrial accident board officials are willing to waive filing date requirements and will allow workers' claims to stand.[209] Such judgments are made on a case-by-case basis, however. As a result, workers filing claims for illnesses caused by a worksite exposure several

203. King, *supra* note 201, at 418-19.

204. DARLING-HAMMOND & KNIESNER, *supra* note 167, at 33.

205. *Id.* at 33-35.

206. U.S. CHAMBER OF COMMERCE, 1990 ANALYSIS OF WORKERS COMPENSATION LAWS 31-34 (1990) [hereinafter 1990 ANALYSIS].

207. *Id.*

208. *Id.*

209. The rules regarding filing deadlines are in place not to impose undue restrictions on employees, but to allow employers a reasonable opportunity to investigate the facts surrounding an injury. *See, e.g.*, LEWIS, *supra* note 168, at 72-73.

years earlier will proceed under uncertain conditions. In their uncertainty, they may well be disposed to make-do with less than generous settlements offered by their employers.

Sometimes employers are willing to challenge claims even though they hold out little hope that their challenges will ultimately prevail. Nearly every state statutorily condemns spurious challenges by employers (and insurers); yet, because of the inherent difficulty in enforcing a good faith requirement, some employers will have incentives to continue to pursue appeals with little regard for the strength of their cases. Since workers must provide for their own legal costs, repeated appeals may encourage workers to seek an arranged settlement.

In most states the size of workers' attorneys' fees are regulated in some fashion. Nineteen states require that attorneys representing injured workers secure approval of all fees charged, either by the industrial accident board or by a court.[210] Many other states determine by statute or industrial accident board policy the specific allowable fee.[211] In Illinois, for example, attorneys' fees are limited to twenty percent of the total award.[212] Nevertheless, legal costs can become an obstacle to the pursuit of compensation. In the first place, since attorneys' fees are often capped, lawyers may decline to challenge an appeal unless there is a high probability that the challenge will be successful. If success is not assured, the attorney has an incentive to recommend accepting an offered settlement. Second, while attorneys' fees are capped, fees for expert medical testimony are not. Obviously, workers require such testimony if their cases are to be presented before (to draw on the Illinois example) the Industrial Commission or a circuit court judge.

Figures illustrating the considerable cost of attorneys' fees, of

---

210. STATE COMPENSATION, *supra* note 125, at tbl. 18. These states (or jurisdictions) include: Alaska, the District of Columbia, Florida, Georgia, Hawaii, Louisiana, Maine, Minnesota, Mississippi, New Jersey, New Mexico, New York, North Carolina, South Carolina, Tennessee, Washington, West Virginia, Wisconsin, and Wyoming.

211. It should be noted that as a rule, states do not impose limits on the fees charged by attorneys representing employers or insurers. *See id.* This differential treatment seems to be two-edged. On one side, injured workers are protected from attorneys who may otherwise charge inappropriate fees. But on the other side, there is no limit to the expense businesses can devote to overturning a claim: employers and insurers can arrange to secure costly (and perhaps more effective) legal representation.

212. *Id.* It should be stressed that this amount is *not additional* to the award, but drawn from it. *See* Illinois Workers' Compensation Act § 16, ILL. ANN. STAT. ch. 48, para. 138.16 (Smith-Hurd 1986).

course, only provide a part of the picture; they say nothing about the expenses associated with the state's administrative and adjudicative instruments. Whenever a case is appealed, state officials—industrial accident board arbitrators, administrative law jurists, and state court judges—must step forward to render a judgment on the disputed claim. One can gain a rough sense of the cost of administrative expenses if the annual expenditures of each state's industrial accident board is divided by the number of cases handled each year. Of course, in most cases this would not include the costs carried by the court system, where some claims eventually are resolved. However, the number of cases which find their way into the courts is relatively small (although the number is a bit greater in certain states).[213] All in all, therefore, we can safely propose that the annual expenditures carried by a state industrial accident board offers a rough estimate of the public costs of administration and adjudication.[214]

Let us first look at each of the four states examined in the preceding section. In Illinois, the system hands down about 58,500 decisions per year.[215] The annual budget for the Industrial Commission is $10,985,800.[216] This means that the state spends on average about $188 per disputed case. In Wisconsin, appeals are first heard by administrative law judges within the Workers' Compensation Division of the Department of Industry, Labor, and Human Relations ("WCD"). The WCD's operations are supported by an annual budget of a little less than $4,300,000.[217] Given that the WCD renders 6,853 decisions per year, the average cost to the state per disputed

213. *See* ADMINISTRATION PROFILES, *supra* note 134.

214. Of course, the budget in most cases also includes moneys for the administration of special funds--second injury funds and the like--under the control of the industrial accident board. This should not be a problem if our goal is to identify the costs of administration and adjudication. Indeed, these are clearly *administrative* costs, and often the distribution of benefits from such funds requires an *adjudicative* decision. Most budgets also, however, contain moneys directed toward the acquisition of information about the frequency and cause of work place injuries, and other services not directly associated with the processing of accident claims.

215. THE INDUSTRIAL COMMISSION OF ILLINOIS, ANNUAL REPORT FOR FISCAL YEAR 1990 11 (1990).

216. *See* ADMINISTRATION PROFILES, *supra* note 134, at 116. It should be noted that the comparability of these figures is open to review. These numbers are taken from state reports and it is possible that states with different approaches to case resolution offer different understandings of when a decision is taken. That is to say, some states may include among the total number of decisions, resolutions negotiated by arbitrators, while others only list final decisions handed down by the chief appeals panel.

217. *See id.* at 421.

claim is about $627.[218] We should make mention of Texas, since we have previously used it as an example. Texas law was revised in 1991 and new figures giving a solid sense of how challenges are processed are not yet available. If we look at numbers for 1989 and 1990, we find that the old system carried an annual budget of nearly $59,000,000.[219] The system settled around 71,500 cases, indicating an administrative cost per case of approximately $825.[220] In New Jersey about 48,700 disputed cases are resolved annually.[221] The New Jersey annual budget is approximately $6,020,000, resulting in a cost per case of about $123.[222] Nationally, Ohio's Bureau of Workers' Compensation has the largest budget of any state: $81,400,000.[223] It resolves around 200,525 cases per year, meaning that each case costs the state about $405.[224] However, it must be recognized that this figure includes within it the various costs attached to the payment of benefits to disabled workers. Ohio, one may remember, requires all businesses to insure with the state's exclusive insurance fund.[225] New York State's Workers' Compensation Board has the next greatest annual budget—about $77,748,800—and may offer a more representative example.[226] It hands down about 433,000 decisions per year, resulting in a seemingly efficient average cost per claim of about $180.[227] The system with the smallest annual budget is Delaware's. The Delaware Industrial Accident Board operates on a budget of $420,600.[228] It hands down 829 decisions per year, for an average cost per claim of about $507.[229] Among all the states, Alabama's WCD may be among the most efficient. Like

218. *Id.*

219. *Id.* at 369.

220. TEXAS INDUSTRIAL ACCIDENT BOARD, 1989 ANNUAL REPORT 50 (1989). In this case decisions are understood to be compromise settlement agreements ("CSAs"). Eighty-five percent of disputed cases are resolved at some point in the appeals process through CSAs. The administrative cost is reached by dividing the budget by the number of cases. *See* BARTH, *supra* note 193, at 107. This is an unusual feature of the old Texas system. Whether this standard will continue under the new system remains to be seen.

221. Correspondence with Mark E. Litowitz, Director and Chief Judge, State of New Jersey Department of Labor, Division of Workers' Compensation (June 18, 1992) (on file with the *Hofstra Law Review*).

222. *Id.*

223. *See* ADMINISTRATION PROFILES, *supra* note 134, at 299.

224. *Id.*

225. *See* STATE COMPENSATION, *supra* note 125, at tbl. 1.

226. *Id.* at 276.

227. *Id.*

228. *Id.* at 62.

229. *Id.*

Delaware, Alabama devotes an unusually small share of resources toward workers' compensation. With an annual budget of $631,000, the Alabama WCD. resolves about 4,703 cases per year, resulting in an average cost of $134 per claim.[230]

b. Externalization Due to Litigation

To this point, we have detailed the factors that generate litigation in the workers' compensation program. We have said little, however, about the connection between litigation and the externalization of accident costs. There are three ways in which the costliness of litigation leads to externalization under the workers' compensation system. The first is claims that are dropped, never filed, or settled for less than full compensation because the anticipated award is less than the anticipated litigation cost. In these cases, the losses due to injury are borne by the employees, and therefore are obviously not shifted to the employer. If these are losses that should be borne by the employer, which is the assumption behind the workers' compensation system, then it is appropriate to refer to them as externalized losses.

The second way in which litigation results in externalization of losses is that the litigation costs of plaintiffs who pursue challenged claims are generally not shifted to the employer. If the claim is valid, the cost of pursuing it should be considered part of the loss resulting from the initial injury.[231] In order to compensate the victim, this cost should be shifted to the employer.

The third sense in which the costliness of litigation generates accident cost externalization is that the administrative costs of litigating claims—the incremental costs of running the administrative machinery—are externalized to the public.

The first two sources of externalization, dropped claims and the costs of litigated claims, stand in the way of shifting the injury losses associated with workplace accidents. Because these losses are not shifted to the employer, the employer will have less incentive to invest in safety than in a regime in which all losses were fully shifted. It is impossible to measure the impact of these sources of externalization because there is very little worthwhile data in this area. To gain some idea of the possible importance of these obstacles to inter-

230. *Id.* This may be a deceptive figure, however, since many claims in Alabama find their way into the courts. Court costs are not accounted for under the W.C.D. budget.

231. *See, e.g.*, Richard L. Abel, *The Real Torts Crisis—Too Few Claims*, 48 OHIO ST. L.J. 443 (1987); Hylton, *Influence of Litigation Costs*, *supra* note 80.

nalization, consider the following simulation.

Suppose the losses that occur as the result of workplace injuries are distributed randomly and suppose the probability distribution is given by the first order exponential function $H(z) = 1 - exp\{-tz\}$, where $1/t$ is equal to the expected value of the random variable z. The shape of the distribution of losses is shown in Figure 2(a). Studies of accident claims report distributions that are similar to that shown in the figure.[232] If the cost of litigating is fixed and equal to $300, all claims within the shaded area of Figure 2(a) will be dropped. Because these losses are not shifted to the employer, these claims make up part of what we refer to as the externalized injury losses. The other source of externalized injury losses consists of the litigation costs of plaintiffs who pursue challenged claims.

Define the ratio of externalized losses to the expected loss as follows: *Ratio = (Expected externalized loss) / (Expected loss)* where the expected externalized loss is given by the following expression:

Prob(Claim dropped) x (Expected loss, given claim dropped) + Prob(Claim not dropped) x (Expected loss, given claim not dropped).

The expected loss, given that the plaintiff's claim is not dropped is the plaintiff's cost of litigating, provided that the plaintiff's injury is fully compensated. If the injury is not fully compensated, then the expected loss, given that the claim is not dropped, is equal to the plaintiff's litigation cost plus the extent to which the loss is under compensated. We will assume for simplicity that the loss is fully compensated in this section. Thus, if V is the plaintiff's loss (which is a random variable), $E(V)$ the expected loss, and Cv the plaintiff's cost of litigating, the expected externalized loss is: $Prob(Claim\ dropped)\ x\ [E(V|V<Cv)] + Prob(Claim\ not\ dropped)\ x\ (Cv)$ where $E(V|V<Cv)$ is the expected value of V given that V is less than Cv.

Under the assumptions just stated, the claim will be dropped if $V < Cv$. Note that the expression for the expected externalized loss is: $H(Cv)E(V|V<Cv) + [1-H(Cv)]Cv$. Substituting the first order exponential function, it is straightforward to show that $Ratio = 1 - exp\{-Cv/E(V)\}$. We will use this in the simulation as a measure of the degree to which losses are externalized because of costly litiga-

232. *See, e.g.*, ALFRED F. CONARD ET AL., AUTOMOBILE ACCIDENT COSTS AND PAYMENTS: STUDIES IN THE ECONOMICS OF INJURY REPARATION 137-52 (1964). Of course, the reported distributions are "censored" because they capture only those claims that were not dropped. But censoring of this sort should not greatly change the shape of the distribution from its uncensored version.

tion. The index of externalization is reported in Figure 2(b) for different levels of the ratio of plaintiff's litigation expenses to the average injury loss. The table in Figure 2(b) shows that when the cost to the plaintiff of litigating is one-half of the average injury loss, roughly forty percent of injury losses will be externalized. When the plaintiff's cost is twice the average injury loss, eighty-six percent of injury losses will be externalized.

The simulation demonstrates that litigation costs can be a significant obstacle to loss-shifting, even if the plaintiff's costs are less than half of the average injury loss. Of course, the real average injury loss is never observed because we see only the claims that are filed and not dropped. The losses that are either never filed or dropped after being contested are likely to be small. If we were to find an accurate measure of average injury loss in a given category of injury, it is quite plausible that it might be roughly equal to the plaintiff's average cost of litigating the claim.

Let us consider a concrete example. Recall that the plaintiff's litigation expenses amounted to twenty-five percent of the average claim in Texas and in Wisconsin. The average claim is an upward-biased measure of the average injury loss because it does not reflect claims that are dropped. If the bias is small enough to be ignored, we can use the average claim to approximate the average injury loss, and in this case Figure 2(b) would imply that twenty-two percent of injury losses are externalized. We can therefore take twenty-two percent as a lower bound on the extent of injury loss externalization that results solely from the litigation of contested claims in Texas and in Wisconsin. If the bias is so substantial that the average claim is twice the average injury loss, then Figure 2(b) implies that forty percent of injury losses would be externalized. Finally, it should be recognized that externalized losses should be increased by the state's average cost of processing a dispute, since this is cost imposed on the public by each contested claim. This is fairly high in both Texas and Wisconsin, on the order of $500 to $1,000. However, there is not enough information on the state's average injury loss to say how taking this figure into account would affect the estimate of the ratio of externalization.

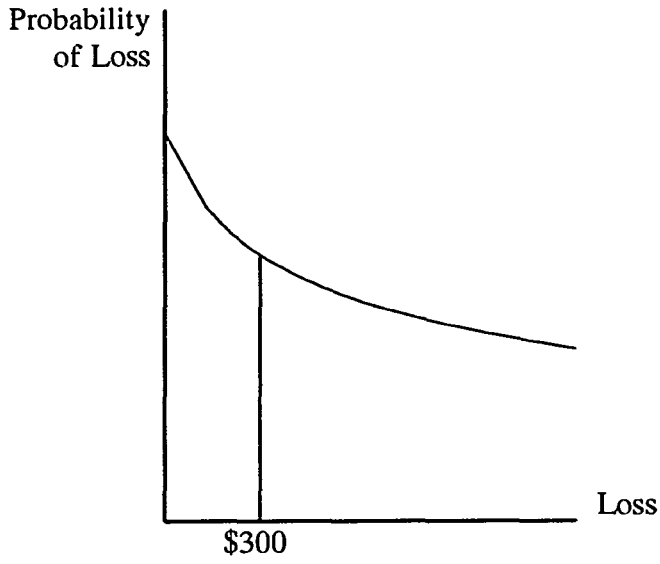

FIGURE 2(a)

Simulation of Loss Externalization

$$Ratio = 1 - e^{\frac{-Cv}{E(V)}}$$

*Cv/E(V)**	
O	Ratio = 0.00
1/4	Ratio = 0.22
1/2	Ratio = 0.39**
3/4	Ratio = 0.53
1	Ratio = 0.63
1.5	Ratio = 0.78
2	Ratio = 0.86
2.5	Ratio = 0.92
3	Ratio = 0.95

* $E(V)$ = average injury loss, Cv = plaintiff's litigation costs

** If cost of litigation for plaintiff is ½ average damage claim, then 39% of injury losses will be externalized.

FIGURE 2(b)

5. Special Funds and Externalization

States have created a handful of special state managed funds designed to move the costliest claims to the edges of the system, i.e., into the hands of the state and out of the hands of business and the insurance industry.[233] These reforms were adopted primarily as a way of assuring protection to particular classes of disabled workers while reducing costs to business.

Several types of special funds can be identified. Second injury funds were introduced earlier than any of the other types of special funds, and are in place in every jurisdiction in the United States.[234] They were designed to encourage employers to hire disabled workers.[235] If a workplace accident serves to worsen an already existing disability, the employer is only required to pay the cost of the injury which would have resulted had the worker not been suffering from a pre-existing condition.[236] Additional expenses are paid out of the state second-injury fund.[237] For example, if a worker already suffered from blindness in one eye, and lost sight in his other eye following an on-the-job accident, the employer would only be held responsible for the award associated with partial blindness. The state would make up the difference. In the particular case of second-injury funds, two ideas are being given expression: (1) the state should provide incentives to ensure that employers hire the disabled, and (2) employers should not be held liable for expenses that do not directly result from work place accidents.

Other types of state managed funds include: (1) funds to provide for the expenses associated with extended hospital care, (2) funds to support rehabilitation expenses, (3) funds to provide cost-of-living adjustments to permanently disabled workers, (4) funds to provide continuing benefits to workers when an insurer (or a self insuring employer) fails financially and can no longer provide support, and (5) funds to pay benefits to workers injured while employed in particular high risk industries.[238] One difficulty with all of these funds is the

233. Lloyd W. Larson & John F. Burton, Jr., *Special Funds in Workers' Compensation, in* WORKERS' COMPENSATION BENEFITS: ADEQUACY, EQUITY, AND EFFICIENCY 117, 118-21 (John D. Worrall & David Appel eds., 1985).

234. *Id.* at 122-35.

235. *Id.* at 123.

236. *Id.*

237. *Id.*

238. *Id.* at 121-22.

following: too often, arrangements in place for financing special state funds are insufficient to the task. Many states demand that employers or insurers pay into these funds some fixed amount in cases where an employee without dependents dies in a work-related accident. The amount contributed by employers/insurers per case varies from state to state. In Alabama, employers/insurers need to contribute only $100 in each no-dependency death case, while in Texas the figure can be as great as $85,680.[239] Arrangements which link employer/insurer contributions to no-dependency death cases carry with them an important failing: few industrial accidents result in death, and even fewer involve workers without dependents.[240] In short, it is difficult to finance a second-injury fund (or any other type of special fund) exclusively with contributions drawn from no-dependency death cases. This is especially true in states like Alabama, which demand such a small payment in such cases.[241] Because such arrangements are so obviously ill-suited to provide sufficient funds, additional funding strategies are put into place. The result is a cut-and-paste funding strategy replete with holes. Employees can find that while employers have been relieved of the responsibility to provide compensation, the state is unprepared to deliver benefits in their place.

If workers' compensation arrangements are designed to help internalize accident costs, state managed special funds too often serve to overturn this effort. In the first place, there is no guarantee that employers responsible for workers' injuries are actually being required to contribute to the state funds. Some states suspend employers' contributions whenever the fund rises above a particular statutorily fixed level.[242] In seeking to supplement the meager sums collected in no-dependency death cases, some states collect from each employer (or insurer) a sum determined by a fixed percentage of all benefit payments made.[243] For example, Montana requires each employer to contribute to the state fund an amount equal to five percent of all

239. U.S. CHAMBER OF COMMERCE, 1989 ANALYSIS OF WORKERS' COMPENSATION LAWS 38-41 (1989) [hereinafter 1989 ANALYSIS].

240. Larson & Burton, *supra* note 233, at 129.

241. *Id.* at 135.

242. *Id.* at 152-53. For example, Iowa only forces employers to contribute to its second-injury fund if the fund falls below $50,000. Because state authorities invest the monies they collect in these special funds, it may be that returns from investments often work to boost the resource total above the minimum, reducing the need to collect payments from employers. In any event, the practice of pursuing returns from investments, while practical, would seem to violate simple notions of enterprise liability.

243. *Id.*

benefits paid by the employer during the previous year.[244] Or, as an example, an employer who paid out $1,000 in compensation benefits in 1990 would be required to contribute $50 to the state fund. Many states adopt this approach, with certain differences from program to program.[245] In some states, assessments are made only on certain types of disability payments. Other states place an assessment on insurance premiums. Louisiana, for one, demands from insurers an amount equal to one percent of all premiums collected.[246] Of course, insurers pass on this cost (or at least some portion of it) to employers. However, some types of funds—principally those devoted to the payment of benefits to workers injured while employed in specific high risk industries covered by special arrangements—draw a great majority of their contributions from employers which have no connection whatsoever with the industry within which the injuries occur. For example, Alaska has a special fund designed to provide workers injured in the fishing industry with compensation.[247] In some cases, the fund is not created exclusively with contributions from boat owners, loggers, or processing factories, but instead draws contributions from all types of businesses working in the state. As a result, disabled workers who draw benefits from the fund will not be deriving their compensation exclusively, or even primarily from the industry "responsible" for their injuries.

Similar concerns can be attached to funds created by states to provide insurance to employers unable to secure affordable protection through the market and to funds designed to pick up benefit payments to disabled workers when an insurer or a self-insured employer

244. *See* ADMINISTRATION PROFILES, *supra* note 134, at 215; *see also* Larson & Burton, *supra* note 233, at 153.

245. Larson & Burton, *supra* note 233, at 152-53. Financing arrangements similar to those of Montana are in place in Alaska, Connecticut, the District of Columbia, Idaho, Indiana, Maryland, Minnesota, Missouri, New Mexico, Oklahoma, Pennsylvania, and South Dakota. *Id.*

246. *Id.* at 152. Arizona, Delaware, Florida, Georgia, Hawaii, Kentucky, Louisiana (as mentioned), Nebraska, Nevada, New Hampshire, New Jersey, New York, Oregon, Rhode Island, South Carolina, Virginia, and West Virginia all finance their funds this way.

247. *See* ADMINISTRATION PROFILES, *supra* note 134, at 9. Other states offer similar funds for specific, high risk industries. Ohio also offers a marine workers' fund. Michigan extends protection to workers employed in logging. Ohio and West Virginia offer special protection to coal miners. Several states offer special protection for workers suffering from diseases associated with exposure to asbestos. In each instance the motivation is clear: the state is intervening to assure that important industries which might not be able to get affordable insurance are assured protection. In each of the preceding examples, the industry covered enjoys an important status, and the work is extremely hazardous. *Id.* at 192, 293, 409.

is brought down by financial failure.[248] With reference to state-provided, low-cost insurance protection, the state steps forward to offer less costly options to employers whose safety records prevent private insurers from extending affordable premiums. If the premiums being offered by the private insurance market are appropriate to the manufacturer's loss experience, the cheaper fees offered by the state will necessarily fail to shift to the employer all accident costs. The gap will be filled by reserves in the state fund, which is financed by contributions from all employers/insurers, not simply those who practice high-risk operations. Moreover, all administrative costs attached to the distribution of benefits will be carried by the state. The concerns attached to state funds created to deliver benefit payments to disabled workers when an insurer or a self-insured employer is brought down by financial failure are even more straightforward. In this case, all costs are shifted to the state's special fund, which is, of course, maintained by contributions from other, presumably more financially responsible, employers and insurers.

One of the main ideas behind these funds is to allow particular classes of employers to share the burden of meeting certain types of costly compensation—not to relieve them entirely of that burden. However, in some cases employers may be entirely relieved of contributing. Moreover, since the administrative costs of dispensing these benefits often are carried by the state, these expenses—ordinarily shouldered by insurers and passed on to employers through service fees—are wholly externalized. Another peculiar twist in the story is that these special funds were created, at least in part, out of a desire to avoid unnecessary delays (and associated costs) linked to repeated employer challenges. Many of these state managed funds are designed to provide benefits to workers bringing forward unusually costly claims. By creating special funds to provide for the payment of benefits for specific types of costly disabilities, states hope to reduce the number of employer challenges. Unfortunately, in order to do so, the state must create new administrative instruments to oversee the state managed funds and distribute benefits, and this, of course, carries a significant cost.

To sum up, special funds generally lead to two types of accident cost externalization. The first is the externalization of injury losses, which results when the injury losses are not borne entirely by the

248. Most states have such standards. *See* Larson & Burton, *supra* note 233, at 118-22.

employer. The second is the generation of additional administrative costs, which are borne by the public. These are essentially the same types of externalization generated by the costliness of litigation. However, the first type of externalization is more worrisome in the case of the special funds because the losses are shifted to the industry or general public. In the case of litigation, most of the injury losses that are not shifted to the employer are borne by the employee. In the case of the special funds, some of the injury losses are shifted away from both parties. This is a more worrisome type of externalization because it removes all incentives from the responsible parties to take injury losses into account.

A further problem with externalization caused by special funds is noticed once competition in the insurance industry is taken into account. The guaranty funds are similar to federal deposit insurance in the banking sector in the sense that they eliminate one class of creditor—customers of the insurance company—that will be seeking to have their claims honored if the insurance company goes bankrupt. To the extent that this reduces the downside risk of competition, insurance firms are given an incentive to underprice their policies and to take on additional risk. Reputational concerns will prevent some insurers, notably larger firms, from taking full advantage of this incentive to take on additional risk, but small insurers will be vulnerable to these incentives.

6. Undervaluation and Undercompensation

a. Compensable Losses

The gap between what a worker is actually due—as measured by his or her losses—and what a worker actually receives in benefits is difficult to describe with certainty. What should be counted as a compensable loss? Surely forfeited wages should be included. It seems evident that one should count as compensable losses all medical expenses directly associated with the on-the-job injury. Beyond this, the issue becomes less easily resolved. How should an arbitrator measure lost earning capacity when calculating permanent partial disability benefits? The "whole man" approach—whereby a disability is evaluated according to the resultant diminishment from total well-being—is imprecise.[249] Additionally, should one count among

249. Nevada provides a good example of this approach. In Nevada, a worker suffering from a hand injury, for example, might be determined to be suffering from an impairment of ten percent (or perhaps less, if the worker is right-handed and the injury is to his left hand).

losses the series of raises a typical worker could expect through the course of his or her working life as he or she is promoted to better positions within the company? What of the loss of employer provided fringe-benefits? What obligation does an employer owe a child who has lost his father to a work related accident? Should the child be assured the opportunity to complete a college education? Equally difficult to measure are the costs of emotional injury and mental anguish. A second set of concerns is generated by the problem of measuring undercompensation. One can find in the workers' compensation field isolated examples where individual cases result in awards which surpass the judgments handed down in tort cases involving similar injuries. Should this lead one to conclude that in these particular cases the worker is over-compensated?

State laws distinguish between *duration* of disability and *degree* of disability. It seems convenient to think of disability awards as the product of a calculation based on the following two by two diagram:

	Duration	
Temporary	**Permanent**	
Temporary Total	Permanent Total	Total
Temporary Partial	Permanent Partial	Partial

Degree

FIGURE 3

Not placed on this table are benefits paid to surviving spouses who have lost partners due to work place accidents. A fair number of states offer lifetime benefits (generally 66 2/3% of worker's wage) to surviving spouses, and seek to offer some type of payment for the educational costs of surviving children.[250] In most states, benefits stop when a surviving spouse remarries (although some states allow

---

That is, compared to the "whole man," this worker has been diminished by ten percent. Nevada calculates permanent partial disability benefits in the following fashion: for each one percent of impairment, the worker is provided with a monthly payment equal to 0.6% of his monthly wage. The payments are limited to a five year period. In this case, therefore, the worker would be provided with six percent of his ordinary monthly wage for a period of five years (that is, 10 x 0.6%). *See* STATE COMPENSATION, *supra* note 125, at tbl. 8.

250. *See id.* at tbl. 12.

support for a child's educational expenses).[251] Because lifetime benefits of this sort are potentially costly, employers work to encourage the introduction of limits. Pressure from business directly encouraged the 1977 revisions of amendments adopted in Illinois in 1975 which substantially increased death benefits.[252] There are limitations in place which reduce the level of death benefits a surviving spouse can collect. Every state imposes a limitation on burial expenses.[253] This limit is usually stated as a specific dollar amount, but some states fix the maximum allowance through more complicated means.[254] As a standard, however, burial allowances generally range from around $1,000 to as high as $6,000.[255] With regard to benefits paid to surviving spouses, there is some diversity in the ways limitations are set.[256] Many states do not set limits.[257] Among those which do offer limits, some states benefits off after a specified number of weeks,[258] while others impose fixed dollar limits,[259] and still others combine these approaches.[260]

Returning to Figure 3, a temporary total ("TT") disability finding would result from an injury of a type which is sufficiently extensive as to keep a worker off the shop floor entirely—and thus prohibit him or her from earning wages—for a limited duration. On the other hand, as a general standard, workers who demonstrate that a work related injury has permanently "rendered them unable to engage in substantially remunerative employment" are qualified to receive permanent total ("PT") disability benefits.[261] In most cases, benefits in PT cases end when unexpected recovery allows claimants to return to

---

251. *See id.*

252. *See, e.g.*, John Elmer & James String, *Showdown Ahead on Workmen's Aid,* CHI. TRIB., Apr. 11, 1976, § 2, at 16.

253. *See* STATE COMPENSATION, *supra* note 125, at tbl. 13.

254. *See* 1990 ANALYSIS, *supra* note 206, at 24-25.

255. *See* STATE COMPENSATION, *supra* note 125, at tbl. 13.

256. Nearly every state stops benefits when the surviving spouse remarries (in some cases the state offers a final lump sum settlement). *Id.* at tbl. 12. Also, it should be pointed out that most states set different limits for spouses without children, and for those with children. *See id.*

257. Arizona, Arkansas, Colorado, Connecticut, Delaware, Iowa, Kentucky, Louisiana, Maine, Nebraska, Nevada, New Hampshire, Ohio, Oregon, Pennsylvania, Rhode Island, South Dakota, Texas, and West Virginia do not. *Id.*

258. Michigan, Mississippi, New Mexico, North Carolina, South Carolina, Utah, Wisconsin, and Wyoming. *Id.*

259. California, Florida, Kansas, North Dakota, and Tennessee. *Id.*

260. Georgia, Illinois, Indiana, and Mississippi. *Id.*

261. JEFFREY V. NACKLEY, PRIMER ON WORKERS' COMPENSATION 42 (1987).

the work force, or the disabled worker dies. In some jurisdictions, workers who are totally blinded or lose two limbs are eligible by statute to receive PT benefits, whether they return to some form of employment or not.[262]

The category of temporary partial disabilities—although making an appearance in our two-by-two table—infrequently receives mention in the literature. It may well describe the largest number of work related injuries. The reason temporary partial disabilities are so frequently ignored by analysts is that in many cases, temporarily partially injured workers do not miss any work and, as a result, do not surrender wages. Consider an example. An assembly line worker, normally responsible for lifting sections of metal trim for placement on car doors, strains her back on the job. The worker visits the shop physician, who gives her a pain killer and a muscle relaxant and, with the agreement of the worker's supervisor, assigns her to perform lighter tasks for the remainder of the week. The following Monday, the worker is back at her normal station without missing a day of work. No wages were lost and, as a result, the employer does not need to offer any compensation payments (although the cost of treatment was, of course, carried by the plant or the worker's health plan).

The last category, permanent partial ("PP") disabilities, is complicated because there are two approaches to the delivery of benefits: scheduled and non-scheduled.[263] Both scheduled and non-scheduled awards are provided to permanently partially disabled workers. Scheduled awards are generally offered to workers who suffer specific permanent partial disabilities, such as the loss of a limb, or the loss of hearing in an ear.[264] The worker is provided with a standardized benefit package, drawn from a statutorily fixed benefit schedule.[265] Scheduled awards are often presented in a single lump-sum payment.[266] Non-scheduled awards, on the other hand, allow a disabled worker to collect an on-going weekly benefit which is calculated with reference to the worker's weekly wage at the time of injury.[267] Non-scheduled awards are provided in cases of temporary disability, in cases of permanent total disability, and for injuries which result in

262. *Id.*
263. *Id.* at 45-48; STATE COMPENSATION, *supra* note 125, at tbl. 9(a).
264. NACKLEY, *supra* note 261, at 45.
265. *Id.* at 47.
266. *Id.* at 162.
267. *See* STATE COMPENSATION, *supra* note 125, at tbl. 8.

permanent partial disabilities that fail to fall under the several categories of impairment established by statute.[268]

Permanent Partial awards are often paid whether or not an injured worker actually details a prospect of lost wages.[269] However, in many cases, it would seem that state officials are attempting to calculate the average expected wage loss associated with a particular type of impairment.[270] State industrial accident boards want to provide compensation, and close each case, without the delay that would be associated with a long-term review of an injured worker's actual post-injury wage profile. Accordingly, in each case a medical evaluation provides an estimation of the degree of impairment, with an understood assumption that impairment levels carry certain expectations about work place disability. Scheduled awards for the loss of a limb, loss of an eye, hearing loss, or disfigurement, come the closest to actual "impairment" awards. In most states, victims of occupational injuries can secure payment of scheduled benefits while at the same time receiving other types of compensation benefits (i.e., temporary total disability benefits) or while continuing to earn a wage.[271] In addition to awards for the actual loss of the limb, many states also provide for injured workers to receive artificial limbs (and training in their use) as a part of their medical award.[272]

There is a great variety in the dollar amounts provided for in states' scheduled awards. As a standard, scheduled awards are calculated according to the following formula. Each type of disability included under the schedule[273] is linked with a statutorily fixed number of weeks. The benefit is determined by multiplying some portion of the injured worker's weekly wage (in most states two-thirds of the weekly wage) by the specific number of weeks demanded by the law. As an example, an injured worker in Alabama who makes $180 a week could expect to collect $26,640 for the loss of an arm.[274] There are, in nearly every state, maximum award lev-

268. *See id.*

269. NACKLEY, *supra* note 261, at 45.

270. *Id.* at 47.

271. *Id.* at tbl. 9(b); *see also* NACKLEY, *supra* note 261, at 48. This is true in Illinois. *See* Illinois Workers' Compensation Act § 8(f), ILL. ANN. STAT. ch. 48, para. 138.8(f) (Smith-Hurd 1986).

272. NACKLEY, *supra* note 261, at 51.

273. In most states scheduled injuries include the loss of an arm, a hand, a thumb, a finger, a leg, a foot, a toe, or an eye, and the loss of hearing in one or both ears. *See* STATE COMPENSATION, *supra* note 125, at tbl. 9(a).

274. *Id.* at tbls. 8, 9(a). This is calculated by multiplying $120 (two-thirds of the

els set in law for each type of scheduled injury. In Alabama, for example, the maximum award for the loss of an arm is $48,840.[275]

Measured by after-tax income, temporarily disabled workers are not, as a group, severely under compensated. As a general accounting, about seventy-five percent of all temporarily disabled claimants receiving benefits under a "typical" benefit scheme retain eighty to one-hundred percent of their regular after-tax income.[276] Many states place statutorily fixed caps on the benefits available to temporarily disabled workers, but, in most cases, these limits are set at relatively generous levels.[277]

More serious are limits placed on permanent disability payments. It is difficult to evaluate the adequacy of PT awards. Such an evaluation would require setting a price on the loss of a limb. We will focus, instead, on limits set on PT benefits.

Most states which cap PT disability payments situate the limitation through a statutory restriction on the number of weeks a claimant can collect benefits.[278] For example, suppose that the maximum weekly disability payment allowed in Texas is $238.[279] Texas law

---

worker's weekly wage) by 222 weeks (the number of weeks indicated in the schedule).

275. *Id.* at tbl. 9(a). This is calculated on the basis of income figures for 1990.

276. KAREN R. DEVOL, INCOME REPLACEMENT FOR SHORT-TERM DISABILITY: THE ROLE OF WORKERS' COMPENSATION xi-xiii (1985). DeVol's "typical" system was the Massachusetts system. Looking at other figures she provides, it seems that the Massachusetts system is, indeed, representative. *Id.* at 52-57.

277. *See* STATE COMPENSATION, *supra* note 125, at tbl. 6. New Mexico limits temporary disability payments (and permanent total disability payments) to 700 weeks; Indiana, South Carolina, and Virginia limit payments to 500 weeks (Indiana and South Carolina's arrangements also limit PT disability benefits); Arkansas and Mississippi limit to 450 weeks (Mississippi's limit applies to PT benefits as well). Texas limits temporary benefits to 104 weeks (and PT benefits to 401 weeks); Missouri and New Jersey to 400 weeks; Florida to 260 weeks; Puerto Rico and Utah to 312 weeks; Oklahoma to 300 weeks; Massachusetts to 156 weeks; and West Virginia to 208 weeks. Idaho and Ohio offer reduced benefits after the passage of a statutorily fixed period of time. Minnesota has a more complicated arrangement which cuts off benefits ninety days after maximum improvement is made, or ninety days after occupational retraining is completed. In most cases (but probably not all), workers exhausting temporary benefits would be found to qualify for PT benefits--except, of course, in states where temporary and PT benefits end at the same statutorily fixed point. *See also* 1989 ANALYSIS, *supra* note 239, at 18-29.

278. In addition to those states listed earlier which limit the duration of both PT and temporary benefits (New Mexico, Indiana, South Carolina, Mississippi, and Texas), two states limit only the duration of PT benefits: Tennessee and Wyoming. *See* 1989 ANALYSIS, *supra* note 239, at 18-29.

279. This was the average weekly wage in the state, and the ceiling on weekly benefits, in 1989. *See id.*

limits the duration of PT benefits to 401 weeks.[280] Accordingly, a disabled Texas worker can only expect to collect $95,438 in weekly disability payments from his employer ($238 x 401 weeks). However, this amount does not represent the total amount of compensation possible. Workers could still collect payments for medical treatments and rehabilitation services, and in cases where the worker has suffered the loss of a limb or an eye, for example, permanent partial disability benefits will also be awarded. Limitations on the number of weeks a disabled worker can draw benefits can be found in about half the states in the nation.[281]

While temporary disability arrangements are intended to pay the immediate expenses of medical care and sustain the worker until he or she returns to work, permanent total disability benefits, as the description suggests, are designed to provide continuing support to a disabled worker whose injuries are so severe that a return to gainful employment is precluded. Benefit limitations can be burdensome, from the perspective of a permanently disabled worker. In New Mexico, for example, which limits PT benefits to 700 weeks, a worker will exhaust his or her benefits in a little more than thirteen years.[282] For a worker disabled early in his working life, say at age thirty, weekly disability assistance can come to an end while the claimant is in his mid-forties. The disabled worker's only alternative in such cases remains a reliance upon Social Security Disability Insurance ("SSDI") and similar support programs which offer less generous benefit arrangements.

b. Inflation

Without inflation protection, permanent total disability benefits will increasingly fall short of replacing lost income.[283] Only a little more than a dozen states offer automatic cost-of-living increases.[284]

In two celebrated reports published in 1972 and 1974, federal reviewers recommended that states adopt measures designed to allow an annual cost of living adjustment to be paid to injured workers receiving benefits under permanent total disability programs.[285] At

280. *Id.*
281. *See id.*
282. *See id.* at 19.
283. *See* Larson & Burton, *supra* note 233, at 135-45.
284. *See id.* at 156.
285. *See* NATIONAL COMMISSION REPORT, *supra* note 170, at 64-65; WHITE PAPER ON

present, only sixteen states have put in place measures offering automatic cost of living increases.[286] Two federally administered compensation programs also offer automatic cost of living increases, as does the U.S. Virgin Islands.[287]

There is some variety among the sixteen states which offer automatic cost of living adjustments to those receiving long-term disability benefits. Some of the sixteen offer cost of living adjustments only to those receiving permanent total disability benefits.[288] Others offer adjustments to several classes of beneficiaries: to workers receiving temporary total disability benefits, to those receiving permanent total disability benefits, and to surviving spouses receiving death benefits.[289] California alone offers adjustments only to those receiving temporary total disability benefits—but only if the disability persists for more than two years.[290] Nearly all of these states adjust benefits upward once a year.[291]

There seem to be two main reasons why additional states have not adopted automatic cost of living measures. In the first place, there are incentive effect concerns. If benefit levels are made to reflect increases in the cost of living, some argue, injured workers will lack incentives to follow rehabilitation programs and return to work.[292] Many injured workers might not enjoy automatic cost of living adjustments on the job; to offer greater inflation protection in a benefit program might discourage an injured worker from returning to work. On the other hand, any move which serves to increase benefit levels might encourage workers to exercise less care in their daily efforts. If disability benefits can nearly equal wages (and medical costs are assured), a worker may more willingly risk partaking in potentially

---

WORKERS' COMPENSATION: A REPORT ON THE NEED FOR REFORM OF STATE WORKERS' COMPENSATION 4-5 (1974); *see also* ALLAN H. HUNT, INFLATION PROTECTION FOR WORKERS' COMPENSATION CLAIMANTS IN MICHIGAN 2-3 (1981).

286. These sixteen states are: California, Connecticut, the District of Columbia, Hawaii, Idaho, Illinois, Maine, Maryland, Michigan, Minnesota, Montana, New Hampshire, South Dakota, Vermont, Virginia, and Washington. *See* NACKLEY, *supra* note 128, at 18-20.

287. *Id.* at 117-23.

288. *Id.* at 18-20.

289. *Id.*

290. *Id.* at 18.

291. *Id.* at 18-20. The Virgin Islands is one notable exception; it adjusts benefits every two years. New Hampshire is another exception; it adjusts benefits every three years.

A few other states offer other qualifications. For example, in Virginia those eligible for Social Security benefits do not receive cost of living adjustments. In the District of Columbia, cost of living adjustments are capped at five percent.

292. DARLING-HAMMOND & KNIESNER, *supra* note 167, at 20.

dangerous work place activities.

Much more important as a reason inflation adjustments are not more widely offered is the cost associated with offering such protections. It has been estimated that a five percent annual cost of living adjustment would require an increase of between 18.5 and 25 percent in indemnity payments by employers.[293] Indeed, it is likely that insurers will demand that employers' premiums be sufficient to cover the cost of a greater than expected cost of living increase. The financial burden of maintaining cost of living adjustments over a period of many years would undoubtedly be considerable.

c. Offsets

Also of some importance is the consideration of contributions to income replacement made by other benefit programs. In the case of temporary benefits, SSDI may play a small role. There is a five month waiting period before SSDI benefits can begin, and in some cases benefits may be delayed for even longer.[294] Still, temporary disabilities can be of a long duration, extending well beyond two years in some cases.[295] SSDI benefits may prove a more important concern for workers suffering from permanent disability. While many states impose upon disabled workers receiving SSDI some type of offsetting reduction in compensation payments, a surprising number of jurisdictions do not impose such limits.[296] The combination of SSDI and workers' compensation benefits can often boost income replacement levels to near one hundred percent—at least for temporarily disabled workers.[297] Two approaches to the calculation of offsetting

293. KAREN R. DEVOL, INCOME REPLACEMENT FOR LONG-TERM DISABILITY: THE ROLE OF WORKERS' COMPENSATION AND SSDI 66 (1986) [hereinafter DEVOL, LONG-TERM DISABILITY]; *see also* HUNT, *supra* note 285, at 46 (citing a National Council on Compensation Insurance figure which estimates a 41.6 percent increase in employer costs given a six percent annual cost of living increase).

294. DEVOL, LONG-TERM DISABILITY, *supra* note 293, at 49.

295. See, for example, the case of James J. Scholl, whose multiple injuries entitled him to 137 weeks of temporary disability benefits. 2 ILLINOIS INDUSTRIAL COMMISSION DECISIONS AND OPINIONS 1988 645-46 (1989).

296. *See* STATE COMPENSATION, *supra* note 125, at tbl. 17. Alabama, Arizona, Connecticut, Delaware, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kentucky, Mississippi, Missouri, Nebraska, Nevada, New Hampshire, Oklahoma, Pennsylvania, Puerto Rico, Rhode Island, South Carolina (unless the injured worker is a member of the State National Guard), South Dakota, Tennessee, Texas, Vermont, Virginia, and West Virginia seem to require no offsetting reductions.

297. DEVOL, LONG-TERM DISABILITY, *supra* note 293, at 48. In some cases, DeVol establishes that income replacement levels can *exceed* one-hundred percent. Federal law

reductions can be identified. In some states, a dollar for dollar offset is introduced, whereby the injured worker surrenders one dollar of workers' compensation for each dollar of SSDI support provided. Other states use a formula which positions workers' compensation benefits at fifty percent of SSDI levels.[298] As an example, imagine that a disabled worker is collecting $250 a week in SSDI benefits. The state would allow the employer to reduce the worker's weekly workers' compensation payment to $125. Of course, the employer would still be required to pay all medical costs and rehabilitation fees.

d. Summary

If the theory underlying workers' compensation is that worker losses should be shifted or internalized to the employer, then work place injury losses that are not compensable should be considered external losses—and, at the least, non-compensable losses provide an easy-to-identify upper limit on the amount of externalization under the workers' compensation system. We have identified three potential sources of externalization in this section: the definition of compensable loss, inflation, and offsetting reductions in employer payments.

The extent of externalization through undervaluation of compensable loss is difficult to determine because it is impossible to measure accurately the value of a limb or the ability to work. However, if any undervaluation occurs, it is through the maximum permanent disability award levels set in the state statutes. The maximum award levels weigh heaviest against high wage workers in industries in which injuries occur frequently.

Similarly, evaluating externalization that results from inflation also raises the problem of putting a price on total disability. However, one solution to the incentive problem generated by indexation to consumer prices is to index total disability payments to wage levels. This would remove any windfall that might result because price inflation exceeded wage inflation, and at the same time preserve some of the purchasing power of the stream of disability payments. This approach would also be consistent with the theory underlying the

---

requires a reduction in SSDI benefits if the combination of SSDI and workers' compensation exceeds 80% of the pre-injury wage. However, as indicated earlier, the tax free nature of both workers' compensation benefits and SSDI means that disabled workers can often replace a greater percentage of their pre-injury take home pay.

298. *See* STATE COMPENSATION, *supra* note 125, at tbl. 17.

workers' compensation statutes: it is the injury to a worker's earning power—injury to the worker as a worker—not injury to a worker's consuming power that the statute seeks to provide compensation for and to internalize to the employer.

Perhaps the clearest source of externalization is the offsetting reduction. Offsetting reductions externalize losses by allowing the employer/insurer to shift them to the taxpaying public.

7. Sources of Externalization and Potential Solutions

We have noted that there are a number of problems in determining whether any externalization of accident costs occurs under the workers' compensation system. If employees are informed about the risks of injury on the job, any losses that are not required to be compensated under the current system will nevertheless be compensated; either informally, i.e., outside of the procedures set up by the compensation statutes, or implicitly in wage payments that compensate employees for accepting additional risk.[299] Of course, if employees were fully informed, the compensation statutes themselves would provide no benefits that would not have been provided in ordinary employment agreements.

We have therefore set out under the assumption that employees are not fully informed about workplace hazards. Under this assumption, we can defend workers' compensation as a mechanism for internalizing accident costs to the party that is best-informed and most likely to respond to incentives for safety, and that will generally be the employer. Loss shifting provides the employer with incentives to make all safety enhancing investments as long as the cost of investing in safety is less than the value of the injury losses prevented.

We have identified several areas in which externalization of injury losses is likely to occur under the workers' compensation system. Two of the potential areas were originally identified by Calabresi in his critique of the tort system: insurance pricing, and the tax and transfer system. Two new categories have been identified in this Article: undercompensation and the litigation process. In the case of workers' compensation, externalization through the tax and transfer system occurs largely through the operation of the special funds and as a result of the state statutes that allow employers to make reductions in compensation payments to offset SSDI payments. Externaliza-

299. *See, e.g.*, RICHARD B. VICTOR ET AL., WORKERS' COMPENSATION AND WORKPLACE SAFETY: SOME LESSONS FROM ECONOMIC THEORY 30-31, 58-59 (1982).

tion through undercompensation occurs through the statutorily fixed maximum permanent disability awards, and the failure in most states to index the maximum awards to some measure of inflation. Externalization through the litigation process occurs because it is costly to bring suit, and bringing suit is required if the claim is contested by the employer. The cost of litigation borne by the employee is like a tax that eats away part of the compensation for injury losses. In cases involving legitimate claims, the injured employee's cost of litigating should be considered part of the injury suffered by the employee.

Although it is impossible to say whether insurance pricing is an important source of externalization, we think the presumption should be that it is not. First, the externalization effects of the regulation of pricing and competition are generally unclear.[300] Second, market imperfections, such as the inability of insurers to accurately identify risk categories, are not clearly important sources of externalization. Market incentives should be sufficient in forcing insurance companies to make every effort to price policies in ways that reflect the likelihood of a claim being made. Experience rating is not perfect, but there are reasons to think that it will never be (even in the absence of price regulation) and that perfection should not even be the goal of risk-based pricing. Rigorous experience rating may lead employers to put pressure on employees to not report injuries. That experience rating is not as widespread as theoretically possible may simply verify that at some point its costs begin to outweigh its benefits.

Externalization that results from the operation of the special funds and through undercompensation are potentially serious problems. But they are fairly easy problems to solve. The externalization that occurs through the existence of special funds can be corrected either by eliminating the special funds or by requiring the states that operate the funds to levy fees that reflect the employer's charges against the special funds. This would probably happen anyway if the funds were run by private insurers.

The most troubling source of externalization is the litigation process. Although empirical work needs to be done in this area, our model simulation suggests that litigation could cause a substantial share of externalization of injury losses. The troubling part of the problem is that there is no apparent solution. If anything, the problem

300. Of course, in states in which insurance commissions hold prices below cost, accident costs will be externalized. But we have argued that the general result of regulation is ambiguous.

is likely to grow worse over time as injuries that are more difficult to trace to the worksite enter the system.

IV. CONCLUSION

This Article has offered an alternative critique of externality theory as a source of normative principles for tort law, and has examined workers' compensation in light of the critique. The (justifiably) better known critique is that provided by Ronald Coase in his article that stated the Coase theorem. The gist of Coase's argument is that the theory of externalities is biased in favor of finding market failures. Coase demonstrated that the important question is whether transaction costs prevent parties from reaching agreements that minimize the social costs of accidents.

The alternative critique provided here is an argument that we refer to as the internalization paradox. It recognizes an inherent tension between the goal of full internalization and that of assessing the validity of compensation claims. Assessing validity is costly and this has several implications. Some claims will not be brought because the victim's cost of proving validity will exceed the value of the claim. In addition, any effort to reduce errors in the assessment process will close out a larger number of claims. Few systems in existence make any effort to fully shift the costs of assessing validity, and it is difficult to see how a system could do this with a high degree of success.

There are several potential sources of externalization in the workers' compensation systems, and we have devoted several pages to identifying them. It seems that the most troubling, because of its potential size and because it seems to be unsolvable, is the litigation of contested claims.