ECONOMIC ANALYSIS OF ACCIDENT LAW

Steven Shavell

Discussion Paper No. 19

7/86

Program in Law and Economics
Harvard Law School
Cambridge, MA 02138

The Program in Law and Economics is supported by a grant from the John M. Olin Foundation.
Note. The material in this Discussion Paper is excerpted from a draft of a book entitled ECONOMIC ANALYSIS OF ACCIDENT LAW. The book will be published by Harvard University Press in 1987. The material included here is the Table of contents, Ch. 1, Ch. 2, Appendix to Chapter 2, Ch. 4, Ch. 9 - Ch. 11, and Ch. 13. Footnotes are omitted.

(A limited number of copies of the entire manuscript are available. If you would like one, please write to Steven Shavell.)
forthcoming, 1987

Harvard University Press

ECONOMIC ANALYSIS OF ACCIDENT LAW

by

Steven Shavell
Table of Contents

Chapter One. Introduction
   A. Use of models to answer two types of theoretical questions
   B. Organization and plan of the book

Chapter Two. Liability and Deterrence: Basic Theory
   A. Levels of care the only determinant of risk
   B. Levels of care and levels of activity the determinants of risk

Appendix to Chapter Two

Chapter Three. Liability of Firms
   A. Victims are strangers to firms
   B. Victims are the customers of firms

Appendix to Chapter Three

Chapter Four. Factors Bearing on the Negligence Determination
   A. Differences among parties
   B. Prior precautions
   C. Uncertainty, error, and misperception
   D. Miscellaneous comments

Appendix to Chapter Four
Chapter Five. Causation and the Scope of Liability

A. Accidents caused by injurers
B. Coincidental accidents
C. Uncertainty over causation

Appendix to Chapter Five

Chapter Six. The Magnitude of Liability: "Damages"

A. The level of losses
B. The probability of losses
C. Courts' uncertainty about the level of losses
D. Pecuniary versus nonpecuniary losses
E. Economic losses
F. Particularistic elements in the computation of liability
G. Victims' opportunities to mitigate losses
H. Liability in excess of losses (punitive damages)

Appendix to Chapter Six

Chapter Seven. Miscellaneous

A. Multiple injurers
B. Injurers' inability to pay for losses
C. Vicarious liability
D. Limited liability of shareholders for the losses caused by corporations

Appendix to Chapter Seven
Chapter Eight. The Allocation of Risk and the Theory of Insurance
A. Risk aversion and the allocation of risk
B. The theory of insurance

Appendix to Chapter Eight

Chapter Nine. Liability, Risk-bearing, and Insurance: Basic Theory
A. The socially ideal solution to the accident problem
B. The accident problem in the absence of liability
C. The accident problem given liability alone
D. The accident problem given liability and insurance

Appendix to Chapter Nine

Chapter Ten. Liability, Risk-bearing and Insurance: Extensions to the Basic Theory
A. Nonpecuniary losses and insurance
B. Divergence between awards that are optimal for purposes of compensation and awards that are optimal for purposes of deterrence
C. Victims' collateral insurance benefits and injurers' liability
D. Injurers' inability to pay for losses and liability insurance
E. Structure of a system of pure accident insurance

Appendix to Chapter Ten
Chapter Eleven. Liability and Administrative Costs

A. Factors determining administrative costs
B. Social versus private incentives to make use of the liability system given its costs

Appendix to Chapter Eleven

Chapter Twelve. Liability Versus Other Approaches to the Control of Risk

A. The different approaches
B. Factors bearing on the appeal of *ex ante* versus *ex post* approaches
C. Factors bearing on the desirability of privately initiated versus state initiated approaches
D. Factors bearing on the appeal of nonmonetary, criminal sanctions
E. Concluding observations

Appendix to Chapter Twelve

Chapter Thirteen. Critical Comments on the Analysis

A. Predictive and normative analysis
B. Purpose and future of tort liability

Bibliography
Chapter One

INTRODUCTION

The subject of this book is the law governing liability for accidents. By "liability," I refer of course to the legal obligation of a party who causes harm to make a payment to the victim of the harm, and by "accidents," I mean harmful outcomes that neither injurers nor victims wished to occur -- although outcomes whose likelihood or severity they might have affected.¹ The body of law determining liability for accidents is included in what is known to us in Anglo-American legal systems as tort law,² and I shall use this term and accident and liability law synonymously.³

A. Use of models to answer two types of theoretical questions

§A.1 Two types of theoretical questions will be analyzed in the book: "predictive" questions, asking about the effects of employing legal rules; and "normative" questions, seeking to evaluate the goodness of legal rules. The questions will be examined using stylized models of parties' behavior and of the working of the legal system.

§A.2 The predictive questions will have definite answers in the models because a complete set of assumptions about parties' situation, the nature and risk of accidents, and the legal system will always be specified. It will be stated, for instance, whether injurers have enough money to pay fully
judgments rendered against them, whether they own liability insurance, whether victims' losses would be purely financial or might include physical injury, whether liability rules are applied perfectly or subject to error, and so forth. Given such assumptions, it will be possible to ascertain the effects of liability rules.

In particular, parties' behavior will be assumed to be determined by the theory of expected utility. According to the theory, if a party is to choose among different actions, and each would result in some consequence with certainty, then he would simply select the action leading to the consequence having the greatest "utility" to him.\textsuperscript{4} However, a party will typically face uncertainty over the consequences of his actions. In such cases, a party will evaluate a potential action in terms of its "expected" utility. The expected utility of an action is obtained by multiplying the probability of each outcome that could follow from the action by the utility of the outcome, and then by adding these values over all the possible outcomes. Thus, if an action would result in an outcome producing utility of 100 with probability 90% and in another outcome yielding utility of 200 with probability 10%, the expected utility associated with the action would be 90%\times100 plus 10%\times200 or 110. Parties will be assumed to take the action with the highest expected utility.\textsuperscript{5}

§A.3 Given the answers to predictive questions, normative questions can be answered, once a social welfare criterion is stipulated. If, for example, the social welfare goal is
posed to be minimization of the sum of accident losses and the costs of accident prevention, one legal rule would be said to be better than a second if the first rule resulted in a lower sum of accident losses plus prevention costs. The answer to normative questions obviously depends on the measure of social welfare under consideration. If the measure just mentioned were altered to take into account compensation of victims, say, the comparison of legal rules could well be different.

§A.4 The advantage of studying models is that it allows predictive and normative questions to be answered in an unambiguous way. Practicality, however, requires that the models be kept simple; while there is no conceptual bar to introducing in them all manner of complications, admitting even a few tends to make the models difficult to solve or to interpret. Thus, the use of the models to understand reality in some of its complexity must be inexact, and rough judgments about the fit of the models have to be made.

B. Organization and plan of the book.

§B.1 Each of the chapters of the book (other than this one and the conclusion) contains several types of section. The first type discusses the models. Sections of this type provide the core of the analysis and often include numerical examples (which are set off so as not to interrupt the flow of the text). The reader should be careful to bear in mind that the
predictive and normative statements made in these sections apply to the models only.

The second type of section is comprised of comments and of interpretation of the analysis. Sections of this type are usually labeled "remarks."

The third type of section contains summary descriptions of the actual law. These descriptions focus on Anglo-American law, but sometimes mention the French, German, and Soviet legal systems.

The fourth type of section notes relevant literature.

§B.2 The claims made about the models in the chapters are demonstrated in mathematical appendices using the standard methods of microeconomic theory. The appendices are essentially self-contained.

§B.3 The plan of the book is straightforward. The book begins with an analysis of "deterrence," that is, of the effects of liability rules on parties' behavior and therefore on the occurrence of accidents. Here and throughout, the two major forms of liability are considered: negligence and strict liability. Under the negligence rule (I omit details), an injurer must compensate a victim only if he, the injurer, was at fault, which is to say, only if his behavior was subpar; whereas under strict liability, an injurer must compensate a victim regardless of whether he was at fault.

To understand deterrence under liability rules, it is helpful initially to study parties' behavior in the absence of insurance, and it is convenient to consider measures of social
welfare that depend on total accident losses and other aggregates, but not on the particular distribution of losses between victims and injurers. These assumptions are made in Chapter 2 on the basic theory of deterrence, and they are maintained in Chapters 3 through 7, extending the basic theory.

Then, in Chapters 8 through 10, insurance is incorporated in the analysis, and measures of social welfare are examined that do take into account the distribution of accident losses (or, more exactly, compensation of "risk averse" parties and the allocation of risk). In these chapters, both accident insurance covering victims directly against losses and liability insurance covering injurers against liability are considered. Insurance is of interest, it need hardly be emphasized, because its ownership is widespread. This makes insurance an important determinant not only of the ultimate bearing of accident losses, but also of parties' financial reasons to avoid doing or suffering harm.

The subject of Chapter 11 is the "administrative costs" of the liability system, namely, the legal and other costs borne by litigants and the public costs associated with operation of the courts. After administrative costs under different liability rules are discussed, a contrast is drawn between the private and the social incentives to make use of the liability system given its administrative costs.

In Chapter 12, liability is compared to other methods for controlling risk: safety regulation, the injunction, fines for harm done, corrective taxes based on anticipated harm, and use
of criminal sanctions. This examination is intended to place liability in perspective and to enable us to appreciate its distinctive aspects.

Finally, in Chapter 13, concluding comments are offered on the analysis of the book. The value of the analysis for predictive and normative ends, the importance of omitted factors, and the purpose and future of tort law are briefly discussed.

Readers who wish to gain a quick understanding of deterrence theory, of the role of insurance, and of administrative costs may wish to look initially only at Chapters 2, 9, and 11.
Chapter Two

LIABILITY AND DETERRENCE: BASIC THEORY

A model of accidents involving two types of parties, injurers and victims is considered here. We might think, for example, of injurers as drivers of automobiles and of victims as bicyclists, or of injurers as parties conducting blasting operations and of victims as passersby. Injurers and victims will each have two kinds of decisions to make (or at least potentially so): a decision whether, or how much, to engage in a particular activity; and a decision over the degree of care to exercise when engaging in an activity. The number of miles an individual drives, for instance, might be interpreted as his level of activity, and the precautions he takes when on the road (slowing for curves, paying attention to the presence of bicyclists) as his level of care. Similarly, how often a bicyclist rides where there is automobile traffic might be regarded as his level of activity, and his precautions when riding (staying close to the side of the road, use of a brightly colored vest) as his level of care.

Injurers may face liability for accidents they cause. They will be assumed, however, not to make contractual or other agreements with victims to pay for accident losses or to reduce accident risk, for injurers and victims will be supposed to be strangers to one another.
In addition, injurers and victims will be assumed to make their decisions on the basis of evaluations of their expected utility, as described generally in the Introduction. The level of utility of an injurer or a victim will be taken to equal the amount he holds of a single, abstract good.\(^2\) Hence, if a party faces, say, a 10% chance of losing 100 units of the good, his expected utility will be lowered by 10% \(\times 100 = 10\). Notice that a party's expected utility would also be lowered by 10 if he faced instead a 1% chance of losing 1,000 units or a .1% chance of losing 10,000. This illustrates that under the present assumption, parties' decisions will not be influenced by the potential magnitude of their losses \textit{per se}. Their decisions will be affected only by their \textit{expected losses}, that is, by the potential magnitude of their losses \textit{multiplied} by the probability of suffering the losses.\(^3\) Parties will therefore be said to be \textit{risk neutral}.\(^4\) Making the assumption of risk neutrality will greatly simplify the analysis of liability and deterrence. Moreover, having studied the situation under the assumption will prove helpful later, in Chapters 8 through 10, when the often more realistic assumption is made that parties are risk averse and are concerned not only about their expected losses, but also about the possible size of their losses.

Given the assumption of risk neutrality and the other assumptions, the effect of liability rules on the behavior of parties and on specified measures of social welfare will now be considered in several increasingly general versions of the
model. The analysis of each version of the model will proceed in the same way. First the socially ideal situation will be discussed, and then the situation in the absence and in the presence of different liability rules will be examined.

A. Levels of care the only determinant of risk

unilateral accidents

$\S A.1$ In this version of the model, it will be supposed that accidents are unilateral in nature: while injurers' behavior will be assumed to affect accident risks, victims' behavior will not, and they will have no role in the analysis. Where an airplane crashes into a building, for example, or where a break in a water main causes a flood in a basement, the victims presumably could not have done much to prevent harm, so that the accidents might be seen as unilateral -- as might automobile-bicycle accidents where it is believed that bicyclists' actions are of minor importance in reducing risks.

It will also be supposed that the only way injurers affect accident risks is through their exercise of care; their level of activity, in other words, will be assumed fixed.

In addition, the social goal will be taken to be minimization of the sum of the costs of care and of expected accident losses. This sum will be called total accident costs.

$\S A.2$ social welfare optimum. Before determining how injurers are led to act in different situations, it is of
interest to identify the level of care that minimizes total accident costs. This socially optimal level of care will clearly reflect both the costs of exercising care and the reduction in accident risks that it would accomplish. Consider\textsuperscript{5}

Example 1. The relationship between injurers' care and the probability of accidents that would cause losses of 100 is as follows.

<table>
<thead>
<tr>
<th>level of care</th>
<th>cost of care</th>
<th>accident probability</th>
<th>expected accident losses</th>
<th>total accident costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>0</td>
<td>15%</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>moderate</td>
<td>3</td>
<td>10%</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>high</td>
<td>6</td>
<td>8%</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

To understand why exercising moderate care minimizes total accident costs, observe on the one hand that raising the level of care from none to moderate reduces expected accident losses by 5, but involves costs of only 3; it thus lowers total accident costs. On the other hand, observe that raising care beyond the moderate level would reduce expected accident losses by only 2, yet involve additional costs of 3; hence it would not be worthwhile.\

Note that the example illustrates the obvious point that the optimal level of care may well not result in the lowest
possible level of expected accident losses (for that would require the highest level of care).

Let us now examine how much care injurers will be led to exercise in various situations.

§A.3 no liability. In the absence of liability, injurers will not exercise any care. Total accident costs will therefore generally exceed their optimal level; in Example 1, for instance, they would be 15 rather than 13.

§A.4 strict liability. Under this form of liability, injurers must pay for all accident losses that they cause. Hence, injurers' total costs will equal total accident costs; and because they will seek to minimize their total costs, injurers' goal will be the social goal. Consequently, injurers will be induced to choose the socially optimal level of care. In Example 1, the column of expected accident losses will become injurers' expected liability and the column of total accident costs will become injurers' total costs. Accordingly, injurers will decide to exercise the optimal, moderate level of care.

§A.5 negligence rule. Under this rule, an injurer will be held liable for accident losses he causes only if he was negligent, that is, only if his level of care was less than a level specified by courts called due care. If the injurer exercised a level of care that equalled or exceeded due care, he will not be held liable.

If due care is chosen by courts to equal the socially optimal level of care, then injurers will be led to exercise
due care. Thus the outcome will be socially optimal. To see why, first reconsider Example 1. If courts define due care to be the socially optimal, moderate level, then the following table describes injurers' situation.

Table 2.2

<table>
<thead>
<tr>
<th>level of care</th>
<th>cost of care</th>
<th>injurer liability</th>
<th>expected liability</th>
<th>injurer's total costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>0</td>
<td>liable</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>moderate</td>
<td>3</td>
<td>not liable</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>high</td>
<td>6</td>
<td>not liable</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

Hence, injurers will indeed be best off exercising moderate care.

More generally, there are two reasons why injurers will necessarily be led to take due care if it is chosen to equal the optimal level. First, injurers plainly would not take more than due care: for they will escape liability by taking merely due care; taking greater care would therefore be to no advantage yet would involve additional costs. Second, injurers would not wish to take less than due care, provided that due care is the socially optimal level. The logic behind this assertion is that if injurers took less than due care, they would expose themselves to liability, so that their expected costs would equal total accident costs. Thus, injurers would want to choose their level of care so as to
minimize total accident costs. But this in turn means that they would wish to raise their level of care to the socially optimal point -- which by hypothesis equals due care and therefore allows them to avoid liability entirely.

§A.6 liability rules compared. Both forms of liability result in the same, socially optimal behavior, but they differ in what courts need to know to apply them. Under strict liability, a court needs only to determine the size of the loss that occurred, whereas under the negligence rule, a court needs in addition to determine the level of care actually taken (a driver's speed) and to calculate the socially optimal level of due care (the appropriately safe speed). To do the latter, in turn, a court needs to know the cost of taking different levels of care and their differential effectiveness in reducing accident risks.

§A.7 remark on the comparison where care has several dimensions. Suppose, as would be usual, that there is more than one dimension of an injurer's behavior that affects accident risks (not only a driver's speed, but also the frequency with which he looks at the rear view mirror). In this situation, under strict liability, an injurer would be led to choose optimal levels of all dimensions of care in order to minimize his expected total costs. But under the negligence rule he would not have a motive to choose optimal levels of those dimensions of care that are not incorporated into the due care standard. And in fact there will usually be excluded dimensions of care. This is because of difficulties that
courts would face in ascertaining certain elements of care (How would a court obtain information about the number of times per minute a driver usually looks at his rear view mirror?) or in determining proper standards in respect to these dimensions of care.

**bilateral accidents**

§A.8 It will now be assumed that victims as well as injurers can take care and thereby lower accident risks.

In this bilateral version of the model of accidents, the way in which injurers choose to behave may depend on the way victims behave, and conversely. For example, how watchful drivers are for bicyclists may depend on how cautious bicyclists generally are (drivers might be very watchful if bicyclists are not very cautious); and how cautious bicyclists generally are may depend on the usual attentiveness of drivers.

The possible interdependence of parties' behavior means that in showing that injurers and victims will act in a particular way, two things will have to be demonstrated: that injurers will choose to act in the asserted way, given that victims act in the asserted way; and that victims will choose to act in the asserted way, given that injurers act in the asserted way. A situation with these two characteristics will be called an equilibrium, since neither victims nor injurers will have a motive to alter their behavior.10

Injurers' and victims' behavior in such equilibria will be determined in the analysis below under several different forms
of strict liability and negligence rules. The social goal considered will continue to be minimization of total accident costs, which here will be the sum of injurers' and victims' costs of care, plus expected accident losses.

§A.9 social welfare optimum. This should reflect the parties' joint possibilities for reducing accident risks and their costs of care. Consider

Example 2. The probability of an accident that would cause losses of 100 is related to the different possible combinations of injurers' and of victims' levels of care as shown below, where it has been assumed for simplicity that there is only one positive level of care -- labeled "care" -- for parties of each type.

<table>
<thead>
<tr>
<th>Table 2.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>levels of care</td>
</tr>
<tr>
<td>injurers</td>
</tr>
<tr>
<td>none</td>
</tr>
<tr>
<td>none</td>
</tr>
<tr>
<td>care</td>
</tr>
<tr>
<td>care</td>
</tr>
</tbody>
</table>

Thus it is socially optimal for both injurers and victims to take care. To see why, observe, for instance, that if injurers alone take care, expected losses are 10, whereas if victims also do so, at a cost of 2, expected losses fall
by 4; hence total accident costs are reduced by victims also taking care. Similar reasoning shows that the situation where victims alone take care can be improved by injurers also taking care.//

While in this example it is socially optimal for both injurers and victims to take care, other examples can, of course, be constructed in which it is optimal only for injurers or only for victims to take care (or for neither to do so). These possibilities are not the focus here (but see §A.20), because in most real situations one supposes that it would be best for both injurers and victims to take a positive degree of care, however small.

§A.10 no liability. As before, injurers will not take care in the absence of liability, and the outcome will therefore generally depart from the optimal. However, because victims will bear their accident losses, they will have a reason to take care. In Example 2, though injurers will not take care, victims will, as this will cost them 2 yet lower expected accident losses from 15 to 12.11

§A.11 strict liability. Since injurers will be liable for the accident losses they cause under strict liability, they will have a proper motive to take care, but because victims will be fully compensated by injurers for accident losses, victims will be indifferent as to the occurrence of accidents. Therefore, victims will not take care,12 and the outcome will not be optimal. In Example 2, injurers will take care because
this will reduce their expected liability from 15 to 10 and
cost only 3; but victims will not take care.

§A.12 strict division of accident losses. By strict
division of accident losses is meant that injurers and victims
each bear a positive fraction of any accident losses that
occur. The fraction is assumed to be independent of their
levels of care and, in particular, independent of whether
someone was negligent (thus the division is called "strict").

Under this form of liability, injurers and victims may be
led to exercise too little care, and the outcome may not be
socially optimal. The reason is that as parties of each type
will bear only a portion of accident losses, what they will
save by taking care is only a portion of the true reduction in
expected accident losses that taking care accomplishes; hence
their incentive to take care may be socially inadequate.

In Example 2, for instance, suppose that injurers and
victims each bear half of any accident losses, that is, 50
rather than 100. Then neither will take care. To verify that
this is so, observe first that injurers will not wish to take
care given that victims do not; for injurers would reduce their
expected liability from 15%×50 = 7.5 to 10%×50 = 5 by taking
care if victims do not; but since taking care would cost 3,
injurers will decide not to do so. Observe likewise that
victims will not wish to take care given that injurers do not;
for victims would reduce their expected losses from 15%×50 =
7.5 to 12%×50 = 6 by taking care; but since that would cost 2,
they too will not find taking care worthwhile. 13
Altering the fraction of liability borne by injurers will not necessarily solve the problem. In the example, while injurers would be induced to take care if their liability was raised substantially above 50, that would even further dilute victims' incentives to take care. On the other hand, while victims would be led to take care if the portion of losses they had to bear was raised substantially above 50, that would further dilute injurers' incentives to take care. Thus, in the example, and in general, there is no "magic" allocation of accident losses that would induce both injurers and victims to take appropriate care.

$\text{SA.13 \ strict liability with the defense of contributory negligence.}$ Under this rule, an injurer is liable for the accident losses he causes unless the victim's level of care was less than a due care level; and when that is so, the victim is said to be \textit{contributorily negligent} and must bear his losses.\(^{14}\)

If the level of due care for victims is chosen by courts to equal the socially optimal level of care, then victims will be induced to exercise due care and injurers also will be induced to take their socially optimal level of care. Thus, the socially optimal outcome will be achieved. To establish this, note, on the one hand, that injurers will exercise optimal care, given that victims take due care. This is so because if victims take due care and therefore will not have to bear their accident losses, injurers will be liable for accident losses. Hence, injurers will have a socially appropriate motive to take
care. (If bicyclists take due care, then drivers will be liable for accident losses and will decide to take optimal care.) Note, on the other hand, that victims will take due care if injurers exercise optimal care. This is true essentially because victims will wish to avoid being found contributorily negligent and thus having to bear their own losses. The specific reasoning is analogous to that in §A.5 explaining why injurers will take due care under the negligence rule.\textsuperscript{15}

To verify the claim in Example 2, assume that due care for victims equals "care," since their taking care is socially optimal. Then presuming that victims take care, injurers will be liable for accident losses they cause. Therefore, their expected liability will be reduced from 12 to 6 if they spend 3 to take care, so that they will do this. Conversely, assuming that injurers take care, victims will be induced to take care; for if victims do not take care, they will bear their expected accident losses of 10; whereas if they take care at a cost of 2, they will not bear their losses.\textsuperscript{16}

§A.14 \textbf{strict liability with the defense of relative negligence}. Under this rule as under the last, an injurer is liable for the accident losses he causes if the victim took due care. If, however, the victim failed to take due care, the victim now does not bear all his losses; rather, he bears only a fraction of them, the fraction depending on his actual level of care relative to due care.
It can be demonstrated that if this fraction is sufficiently large and if due care is chosen by courts to equal the socially optimal level of care, then victims will be induced to take due care and injurers will also be led to take the socially optimal level of care. This should be evident from the previous section (and details are therefore omitted), since if the fraction of losses that contributorily negligent victims would have to bear is high, then the rule here resembles the last one. In Example 2, for instance, if a victim who does not take care will have to bear more than half of his losses, he will be induced to take care if injurers do so, as the victim will then save more than 2 in expected losses by taking care, and so forth.

§A.15 negligence rule. The description of this rule is virtually the same as in the unilateral case: If an injurer takes at least due care, he will not be liable for accident losses he causes; otherwise he will be liable, regardless of how the victim acted.

It is easy to see that if due care is chosen by courts to equal the socially optimal level, then again injurers will be led to take due care and victims will also be induced to take the optimal level of care. Injurers will be motivated to take due care to avoid liability, by the argument of §A.5. And because victims will bear their losses if injurers take due care, victims will have a proper incentive to take care. (Drivers will be led to take due care; and bicyclists, knowing
that they will bear their losses, will decide to take appropriate care.)

To illustrate, assume in Example 2 that due care for injurers equals "care." Then if injurers do not take care, their expected liability will be 12, presuming that victims take care; thus injurers will choose to avoid liability by spending 3 on care. Also, as victims will bear their losses when injurers take due care, victims will reduce their expected losses from 10 to 6 by taking care; since this will cost victims 2, they too will decide to take care.

§A.16 negligence rule with the defense of contributory negligence. According to this rule, an injurer will not be liable for accident losses he causes if he takes at least due care; and even if he does not, he will still escape liability if the victim too failed to take due care.

By an argument very close to that of the previous section, it can readily be seen that if injurers' and victims' levels of due care are chosen by courts to equal the socially optimal levels, then both injurers and victims will be led to take due care and the socially optimal result will be achieved. Injurers will wish to take due care to avoid liability, under the assumption that victims take due care and thus will not bear their losses on account of contributory negligence. Also, victims will wish to take due care, presuming that injurers take due care; for as victims will then bear their losses, they will be led to take the socially optimal level of care, which
by assumption is due care. (This may be verified in Example 2 exactly as it was in the preceding section.)

Notice that the defense of contributory negligence is a superfluous addition to the negligence rule with respect to the objective of inducing victims to act optimally, for it was seen that victims take optimal care when the negligence rule is unaccompanied by the defense. The explanation should be clear on reflection. Under the negligence rule without the defense of contributory negligence, injurers take due care to avoid liability. Consequently, victims bear their losses, and this by itself supplies them an incentive to take appropriate care. Accordingly, there is no need to provide victims another incentive to take care. 17

§A.17 comparative negligence rule. Under this rule, like under the last, an injurer will not be liable for accident losses he causes if he takes due care. But this rule differs from the previous rule in the situation where both an injurer and a victim fail to take due care. In that situation, each party bears a fraction of the accident losses, where the fraction is determined by a comparison of the amounts by which the two parties' levels of care depart from the levels of due care 18 -- thus the term "comparative" negligence.

Under the comparative negligence rule, if courts choose optimal levels of due care, then both injurers and victims will be led to take due care. The rationale for the result is precisely that of the last section. (Injurers will take due care to avoid liability if victims take due care, and so on.)
The reason there is no difference between the outcomes under the comparative negligence rule and under the negligence rule with the defense of contributory negligence is in essence this: Under both rules, if parties of one type take due care, then parties of the other type will reason that they alone will be found negligent if they fail to take due care. The allocation of accident losses when both injurers and victims are negligent -- the distinguishing feature of the comparative negligence rule -- therefore turns out to be irrelevant to parties' calculations in equilibrium. 19

§A.18 liability rules compared. In the present bilateral version of the model, it was seen that strict liability does not lead to the socially optimal outcome for the obvious reason that it fails to supply victims a motive to take care. And while strict division of accident losses may provide victims some incentive to take care, it may leave injurers with an inadequate incentive to do so. A comparison of these two forms of strict liability with each other and with not having liability at all therefore depends on the importance of modifying injurers' as opposed to victims' behavior. The more important it is for injurers to take care, the greater the relative appeal of strict liability -- or a division of losses in which injurers pay a high fraction -- over no liability.

It was also seen that strict liability with the defense of contributory (or relative) negligence and all forms of the negligence rule result in the socially optimal outcome. Under these rules, parties have one of two sufficient reasons to take
optimal care: Either taking optimal care allows them to avoid entirely the bearing of accident losses (victims' situation under strict liability with the defense of contributory negligence, injurers' situation under the negligence rules); or else taking care reduces the expected losses that parties in fact bear (injurers' situation under strict liability with the defense of contributory negligence, victims' situation under the negligence rules).

To apply each of the rules leading to optimality, courts need to determine the magnitude of accident losses, and the actual level of care and the optimal level of due care for injurers or victims. Moreover, to ascertain the optimal level of due care for one party, a court must generally determine (if only implicitly) the optimal level of care for the other, since the optimal level of care for one party will in principle depend on the other's costs of and possibilities for reducing risk. This latter point makes the comparison of liability rules with respect to their ease of application different from what it might at first seem to be.

Consider, for instance, the rule of strict liability with the defense of contributory negligence and the negligence rule with the defense. It may seem initially that strict liability with the defense of contributory negligence is the easier to apply, because courts are not directly concerned with injurers' behavior under the rule. But to apply the defense of contributory negligence, courts must determine optimal due care for victims, and, as just remarked, this ordinarily effectively
requires courts to determine the optimal level of care for injurers. Therefore, the main difference affecting the ease of application of the two rules is only that under the strict liability rule, courts do not need to determine the actual (as opposed to the optimal) level of care of injurers.

$A.19$ remark on the comparison where care has several dimensions. It was noted in $A.7$ that there may be dimensions of injurers' care (such as the frequency with which drivers look at their rear view mirrors) that would not be taken into account in the negligence determination because of difficulties courts would encounter in assessing them. Injurers would therefore not choose these dimensions of care in an optimal way under the negligence rule, but they would be led to do so under strict liability. It is clear that a similar point applies where there are dimensions of victims' care (such as the frequency with which bicyclists look behind themselves) that could not be included in their standard of due care. Specifically, victims would not choose these dimensions of care optimally under strict liability with the defense of contributory negligence, but they would do so under the negligence rule (because they would bear their losses under that rule). In consequence, to know how the present consideration affects the comparison of liability rules, one must make a judgment about the relative importance of the dimensions of injurers' and of victims' behavior that would be excluded from their respective standards of due care.
§A.20 remarks on the notion of the least cost avoider. This notion applies in situations where if either injurers or victims take care, the risk of accidents will be eliminated. In such situations, it is clearly wasteful for both injurers and victims to take care. What is optimal is for the parties who can prevent accidents at least cost, the "least cost avoiders," alone to take care. Suppose, for example, that injurers can prevent accident losses of 100 by taking a precaution that costs 10, and that victims also can prevent the losses by taking a precaution that costs 20. Then injurers alone ought to take precautions, because they are the least cost avoiders.

Consideration of the model of the least cost avoider may be misleading for thinking about the class of bilateral accident situations examined in this book. In the situations examined here, there simply are no least cost avoiders who alone ought to take care, since both injurers and victims generally ought to do something to avoid risk, and the functioning and comparison of liability rules is therefore different from in the least cost avoider model. In that model, for instance, if injurers are the least cost avoiders, an optimal outcome will be achieved under strict liability unaccompanied by the defense of contributory negligence. But in the bilateral model studied here, of course, the defense of contributory negligence must accompany strict liability in order to induce victims to take appropriate care.
Similarly, consideration of the least cost avoider model may lead to the belief -- though mistaken -- that the liability rules asserted here to result in optimal behavior can result in suboptimal behavior. To illustrate, the following might be said about the example above where the injurers who can prevent losses at a cost of 10 are the least cost avoiders: "Since victims can also prevent the losses of 100 at a cost of 20 and 20 is less than 100, they could be found contributorily negligent for failure to take precautions. Thus, use of the defense of contributory negligence can lead to the undesirable result that victims rather than injurers take precautions." The error in this argument lies in the assumption that victims might be found contributorily negligent for failure to take precautions. If due care for victims is assumed to be optimally determined -- which, recall, has always been the assumption in showing that rules of liability lead to optimal results -- then the due care requirement for victims in the example should be vacuous: There should not be any duty for victims to take precautions. 22

These remarks about the least cost avoider model are not meant to suggest that the model cannot be helpful, but rather that the model is special and must be interpreted with caution. 23

In the United States, the major rules of liability for accidents between strangers are the comparative negligence rule, the negligence rule with the defense of contributory negligence, and strict liability with that defense. 24 In England,
France, Germany, and the Soviet Union, the usual forms of liability are the comparative negligence rule and strict liability with what was called here the relative negligence defense.  

§A.22 remark on the actual determination of due care and its "as if" interpretation. Negligence in American law as defined in the Restatement of Torts, Second, is "conduct which falls below the standard [of due care]... for the protection of others against unreasonable risk of harm," and the concept is similar in other legal systems.  

To decide on the standard of due care often requires some sort of weighing of the magnitude of risk against the disutility or cost of more careful conduct.  

Now as the reader has seen in the analysis here, the level of due care that minimizes total accident costs implicitly involves just such a weighing of risk against the cost of care. This suggests that due care is in fact found by a process that operates as if it were designed to identify total accident cost minimizing behavior.  

I hasten to say that the words as if are stressed because the interpretation is hardly that individuals or courts think in terms of the mathematical goal of minimizing a sum. They obviously do not do anything so unnatural. Rather, they appear to gauge the appropriateness of behavior by a rough consideration of risk and the costs of reducing it, ordinarily on the basis of felt notions of fairness. Likewise, the as if interpretation carries with it no specific implications
about the degree to which individuals or courts concern themselves about goals of deterrence.\textsuperscript{31}

With these caveats in mind, observe that the \textit{as if} interpretation is borne out not only by the mere fact that there is a weighing involved in the negligence determination, but also by a consideration of its more particular character. First, the elements that are taken into account in finding due care and their effect on its level are what would be expected were the aim to minimize total accident costs: the level of due care is generally higher the greater the likelihood of harm, the larger the probable size of harm, the more individuals who are at risk, and the easier it is for parties to alleviate risk.\textsuperscript{32} Second, the choice of due care levels probably reflects the \textit{joint} nature of the possibilities for injurers and victims to reduce accident risks, as is consistent with the bilateral model of accidents. Consider, for instance, accidents in which bicyclists run into car doors as the doors are opened. My surmise is that most of us would say that bicyclists should not have to proceed so slowly that were a car door to open suddenly, they could virtually always stop in time, and that before persons open car doors to leave, they should look around to see if anyone is approaching. I suggest too that in coming to this view, most of us would have at the back of our minds -- if not in our conscious thoughts -- such ideas as that it would be a burden for bicyclists to have to go so slowly that they could stop immediately before running into car doors, that it is relatively easy for persons leaving cars
to look for danger, and that it is not necessary for bicyclists to go very slowly if persons are properly cautious when leaving their cars. In other words, when deciding on the care that parties of one type ought to exercise, we quite naturally factor into our thinking the ability to take care of parties of the other type and what their care would accomplish.

§A.23  note on the literature. The first writer to study in an analytical way the theory of the effect of liability rules on parties' behavior was Calabresi (1961, 1965, 1970). He examined the desirability of different rules, emphasizing versions of strict liability and assuming for the most part the goal of minimization of total accident costs. Posner (1972, 1973a,b) later made significant contributions, especially in his analysis of the various principles and doctrines governing use of the negligence rule. While both these writers used suggestive numerical examples, neither recognized that liability rules would, as a general matter, lead calculating parties to take total accident cost minimizing levels of care. Brown (1973) contains the first clear statement and formal proofs of this result. He showed that the rules of strict liability with the defense of contributory negligence and the negligence rule (with or without the defense) induce injurers and victims to take total accident cost minimizing levels of care in equilibrium.
B. Levels of care and levels of activity: the determinants of risk

unilateral accidents

§B.1 Now that the effect of liability rules on parties' exercise of care has been studied, the influence of the rules on parties' levels of activity will also be considered. The analysis will begin with the unilateral case, where only injurers' actions affect risk, and will rely on two assumptions about injurers' level of activity: First, an increase in their level of activity will result in a proportionate increase in expected accident losses, given their level of care. Thus, a doubling in the number of miles that individuals drive will result in a doubling in the number of accidents they cause, given the care with which they drive; or a doubling in the number of times individuals walk their dogs will result in a doubling in the risk that their dogs will bite strangers, given the care taken (leashing) to prevent that. Second, an increase in injurers' level of activity will result in an increase in their utility (at least up to some point); the more individuals drive or the more they walk their dogs, the greater will be their utility (until their need to drive is met or until walking their dogs turns into a chore).

The social goal will be taken to be maximization of the utility injurers derive from engaging in their activity less total accident costs, that is, less their costs of care and expected accident losses. It makes sense, of course, to
introduce the utility injurers derive from their activity into the measure of social welfare, since the level of their activity is to be studied. 37

§B.2 social welfare optimum. This will reflect two elements: As before, when an injurer engages in his activity, he ought to take care commensurate with its effect in reducing accident losses and with its costs. But now, also, he should engage in his activity at the level that appropriately balances the utility he obtains against the additional risks he creates. Consider

Example 3. Assume that the situation each time injurers engage in their activity is as was described in Example 1. Namely, if injurers behave optimally, they will take moderate care at a cost of 3 and will reduce expected accident losses to 10. Consequently, if an injurer engages in his activity twice, taking optimal care each time, his total costs of care will be 6, and the expected accident losses he causes will be 20; if he engages in his activity three times, the figures will be 9 and 30, respectively; and so forth. This explains the third and fourth columns in the table below. Supposing that the second column shows the total utility injurers derive from engaging in the activity, the last column can be calculated.
Table 2.4

<table>
<thead>
<tr>
<th>activity level</th>
<th>total utility from activity$^{38}$</th>
<th>total costs of care</th>
<th>total expected accident losses</th>
<th>social welfare = total utility - total costs of care expected accident losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>3</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>6</td>
<td>20</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>69</td>
<td>9</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>71</td>
<td>12</td>
<td>40</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>70</td>
<td>15</td>
<td>50</td>
<td>5</td>
</tr>
</tbody>
</table>

The optimal activity level is 2 because social welfare is highest at that level. To see better why, observe that each time an injurer engages in the activity, he will increase total accident costs by $3 + 10 = 13$. Therefore, social welfare will be enhanced by his engaging in the activity another time if and only if the marginal utility he would gain exceeds 13. Since the utility he obtains from engaging the first time is 40, the marginal utility he obtains from the second time is 20, and that from the third time is only 9, it is best that he stop at the second time.//

The general point illustrated by this example is that the socially optimal behavior of injurers can be determined in two steps: first, by finding (as in $\$A.2$) the level of care that minimizes total accident costs incurred each time injurers engage in their activity; and then by raising the level of
activity as long as the marginal utility injurers derive exceeds the increment to total accident costs.

§B.3 no liability. In the absence of liability, not only will injurers fail to take care, they also will engage in their activity to too great an extent. Indeed, they will continue to engage in it as long as they obtain any additional utility (individuals will go for a drive or walk their dogs on a mere whim) -- rather than, as would be socially desirable, only as long as they obtain additional utility exceeding the costs of optimal care plus the expected accident losses they cause. In Example 3, injurers will not take care, and thus will choose activity level 4, the level at which they cease to gain utility from their activity, rather than the optimal activity level of 2.

§B.4 strict liability. Under this rule, an injurer's utility net of his expected costs will be equal to the measure of social welfare, since he will pay for the accident losses he causes and he will naturally enjoy the benefits of engaging in his activity and will bear the costs of care. Accordingly, injurers will behave so as to maximize social welfare; they will thus choose both the optimal level of care and the optimal level of activity.

More directly, injurers will choose the optimal level of care because this will minimize the expected costs they bear each time they engage in their activity. And they will choose the optimal level of activity because they will wish to engage in the activity only when the extra utility they derive exceeds
their costs of care plus their added expected liability payments for accident losses caused. (People will walk their dogs only when their utility gain out weighs the disutility of having to leash the dogs and the added liability risk due to dog bites.) In Example 3, for instance, we know (from §A.4) that strictly liable injurers will take the moderate level of care. Hence, the last column from the table will become injurers' utility net of their expected costs, and they will therefore choose the optimal activity level of 2.

§B.5 negligence rule. As the reader recalls from previous analysis, injurers will be led to take optimal care under the negligence rule, assuming that the level of due care is chosen by courts to equal the optimal level of care. Because they will take due care, however, injurers will escape liability for any accident losses they cause. They will therefore not have a reason to consider the effect that engaging in their activity has on accident losses. Consequently, injurers will be led to choose a socially excessive activity level. Specifically, they will engage in their activity whenever the utility they derive net of the cost of care is positive (whenever the pleasure from walking their dogs net of the disutility of leashing them is positive), rather than only when their net utility exceeds the additional expected accident losses they create.

This can be seen in Example 3, where we know (from §A.5) that if due care is the optimal, moderate level, injurers will take due care. Hence, injurers will not be liable for accident losses, and their situation will be as follows.
<table>
<thead>
<tr>
<th>activity level</th>
<th>total utility from activity</th>
<th>total costs of care</th>
<th>total utility net of total costs of care</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>6</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>69</td>
<td>9</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>71</td>
<td>12</td>
<td>59</td>
</tr>
<tr>
<td>5</td>
<td>70</td>
<td>15</td>
<td>55</td>
</tr>
</tbody>
</table>

From the last column, it is evident that injurers will choose the activity level 3 rather than the optimal activity level 2. And they will do this for the reasons just explained: They will increase their activity level from 2 to 3 because this will raise their utility by 9 and their costs of care by only 3; they will not consider that increasing their activity level will also raise expected accident losses by 10, for they will not be liable for these losses.

§8.6 liability rules compared. While under both strict liability and the negligence rule injurers are led to take socially optimal levels of care, it has been seen that under the negligence rule they engage in their activity to too great an extent because, unlike under strict liability, they do not pay for the accident losses they cause.39
The importance of this defect of the negligence rule will clearly depend on the expected magnitude of the losses caused by an activity. If an activity is by its nature very dangerous even when carried out with appropriate precautions, then the fact that under the negligence rule the level of the activity would be excessive might be significant. If the walking of dogs of a vicious breed or if blasting creates high risks of harm despite the use of all reasonable care, then the fact that under the negligence rule people would walk the dogs excessively (rather than exercising them in a yard or rather than owning dogs of another breed) or that firms would blast excessively (rather than employing other methods of demolition) might be of real consequence. If, however, an activity creates only a low risk of accidents when due care is taken, then the importance of any excess in the level of activity under the negligence rule will be small. This is true, one suspects, of many and perhaps most of our everyday activities (mowing a lawn, playing catch, walking the friendly, domesticated dog).

§B.7 remarks on the source of the defect of the negligence rule. The failing of the negligence rule that is under discussion can be regarded as due to an implicit assumption that the standard of behavior used to determine negligence is defined only in terms of care. Were the standard defined also in terms of the activity level, injurers would make sure not to engage in their activity to an excessive extent.
This consideration, however, immediately raises questions as to the reason it is assumed that courts do not include the activity level in the negligence determination. A possible reason concerns the information that they would need to do so. To formulate a standard for the level of activity, courts would need to determine the character of the benefits parties derive from their activities. (Courts would have to inquire into the pleasure obtained from walking a dog or the need for and importance of driving somewhere.) Because these benefits often seem practically unknowable, attempts by courts to determine appropriate levels of activity would probably quickly land them in the most speculative of realms. Deciding on appropriate levels of care, although by no means an easy task, usually appears to be less problematic. (We can say with fair confidence that a dog that snaps at others should be leashed, or that a person should not drive at 60 miles per hour along a residential street.)

Aside from the difficulties that courts would have to face in formulating appropriate standards for parties' levels of activity, courts would have to ascertain what parties' levels of activity actually were. This additional burden might be a substantial one in some situations, especially because determining a party's level of activity would require knowledge of what the party did in the past. (How many times did a person walk his dog before the last time when it bit someone?) By contrast, assessing a party's level of care often requires knowledge of his behavior only at the time of an accident.
Nevertheless, there may be situations where a court would have sufficient information to incorporate the level of activity into the negligence determination. Notably, if having engaged in an activity even once was very dangerous despite the exercise of care and if the utility obtained from the activity was obviously small, then a party could be called negligent merely for having engaged in the activity.\footnote{42}

**bilateral accidents**

§B.8 In this most general case, victims as well as injurers will be assumed to choose levels of activity and levels of care. As with injurers' levels of activity, increases in victims' levels of activity will be assumed to raise their utility, at least up to some point, and will result in proportionate increases in expected accident losses. Thus, when a bicyclist rides an extra mile, he will enjoy extra utility and will increase his chances of being involved in an accident. The measure of social welfare will be taken to be the utility that victims and injurers derive from their activities, less their costs of care and expected accident losses.

The analysis that follows will be brief because most conclusions can be explained by appeal to previous cases.

§B.9 **social welfare optimum.** This will reflect not only the costs of care and its effect on accidents risks, but also the utility that injurers and victims obtain from their activities. Consider
Example 4. Suppose for simplicity that victims either engage in their activity or they do not, and the same for injurers; in other words, for parties of each type, there is only one possible positive level of activity. Suppose also that if parties of one type engage in their activity and the others do not, no accidents can occur -- it takes the presence of both injurers and victims for there to be accidents. Hence, if parties of only type engage in their activity, it would be point-less and socially wasteful for them to take care. Last, suppose that if both injurers and victims engage in their activities, there will be a risk of accidents, as described in Example 2. Thus, in this case, injurers ought to take care, which, recall, costs 3; victims also ought to take care, which costs 2; and expected accident losses will be 6. Therefore, total accident costs will be $3 + 2 + 6 = 11$ if both injurers and victims engage in their activities.

Given these assumptions, it is easy to determine when it is optimal for injurers and for victims to engage in their activities, as a function of the utilities they would each derive from so doing. Were parties of only one type to engage in their activity, none of the accident costs of 11 would be borne (since no accidents could occur and no care would be taken). Therefore, it will be social welfare maximizing for both injurers and victims to engage in their activities only when each would obtain utility exceeding 11 from their activity. Otherwise, it will be best for the
parties that would enjoy the greater utility to engage in their activity, unaccompanied by the other parties.

To verify this, suppose for instance that injurers would obtain utility of 35 and victims 25 from engaging in their activities. Then, if both injurers and victims engage in their activities, social welfare will be 35 + 25 - 11 = 49; if only injurers engage in their activity, social welfare will be 35; if only victims do so, social welfare will be 25; thus it will indeed be optimal for both injurers and victims to engage in their activities. On the other hand, suppose that while injurers would obtain 35 from engaging in their activity, victims would obtain only 8. Then if both injurers and victims engage in their activities, social welfare will be 35 + 8 - 11 = 32; if injurers alone do so, social welfare will be 35; if victims alone do so, social welfare will be 8; and it will be best for injurers alone to engage in their activity. Similar calculations show that if injurers would obtain 8 and victims 25 from engaging in their activities, then it will be optimal for victims alone to engage in their activity. The simplifying feature of this example, that parties either do not engage in their activity or engage in it at only one positive level, should not disturb the reader. The points to be illustrated below will carry over in obvious ways to the more realistic case where there are many different positive levels of activity.
§B.10 **strict liability with the defense of contributory negligence.** As the reader knows from previous analysis, if courts select the optimal level of due care, then under strict liability with the defense of contributory negligence both injurers and victims will be led to take optimal care when they engage in their activities. Furthermore, since victims will take due care, injurers will pay for the accident losses they cause, and thus, as in §B.4, will choose the correct level of their activity given victims' behavior.

Yet because victims will be compensated for their losses, victims may engage in their activity too often. A victim's only cost of engaging in his activity will be his cost of taking due care. Therefore, he will engage in his activity whenever his utility from so doing would exceed the cost of taking due care. But what would be desirable is that he engage in his activity only when his utility would exceed the cost of taking due care plus the expected accident losses that would result from his engaging in his activity. (A bicyclist will go for a ride whenever the pleasure he would gain exceeds the disutility from having to exercise appropriate care, rather than only when the pleasure exceeds the disutility of exercising care plus the increment to expected accident losses.)

To illustrate, consider in Example 4 the case where injurers would obtain utility of 35 and victims utility of only 8 from their activities, and thus where it is not optimal for victims to engage in their activity. Under strict liability with the defense of contributory negligence, victims would need
only to take due care, at a cost of 2, to be assured of compensation for accident losses suffered. Hence, when they compare the utility of 8 that they would obtain from engaging in their activity to the cost of care of 2, victims will, undesirably, decide to engage in their activity (along with the injurers, for they will compare their utility of 35 to their cost of care of 3 plus their expected liability of 9).

§8.11 negligence rule with or without the defense of contributory negligence. Again, the reader knows from before that under this rule if courts select optimal due care levels, both injurers and victims will be induced to take optimal care when they engage in their activities. Also since injurers will escape liability by taking due care, it is evident from the argument of §8.5 that injurers may engage in too high a level of their activity.

Victims, however, will choose the correct level of their activity given injurers' behavior. As victims will bear their own losses, they will engage in their activity another time only if the utility they would obtain net of the costs of taking care exceeds the addition to expected losses. In Example 4, consider the situation where injurers would obtain utility of 8 and victims utility of 25 from engaging in their activities. In this situation, it is optimal for victims alone to engage in their activity; and while under the negligence rule they will do so (for they will compare 25 to 2 + 6), so will injurers engage in their activity (for they will compare 8 to 3).
§B.12 other liability rules. It follows from the last section and from the earlier analysis in §A.17 that under the comparative negligence rule, the outcome will be the same as under the negligence rule. It should be clear too that under strict liability without the defense of contributory negligence, victims will not only fail to take care, but they will also engage in their activity too often, whereas injurers will choose socially appropriate levels of care and of their activity given victims' behavior.

§B.13 liability rules compared. It should be evident from §B.10 and §B.11 that strict liability with the defense of contributory negligence will result in higher social welfare if its disadvantage -- that victims engage too often in their activity -- is not as important as the disadvantage of the negligence rules -- that injurers engage too often in their activity. That is, strict liability will result in higher social welfare if it is more important to control injurers' level of activity than victims'.

Whether injurers' level of activity will be more important to control than victims' will depend on context. As discussed before, where an activity of injurers (walking dogs of a vicious breed) creates substantial risks despite the use of due care, the activity will be desirable to control. This point is not fundamentally altered if account is now taken of the activities of victims that expose them to risk. Especially if these activities are just the activities of ordinary life (walking about, going to work), we would not want the
activities constrained in favor of injurers'. Conversely, where an activity of injurers (playing baseball) is not very dangerous if appropriate care is taken, the importance of controlling the activity will not be great; and once victims' behavior is taken into account, we may see some advantage in their moderating engagement in certain activities that subject them to particular risks (such as pushing a baby carriage across a baseball field while a game is in progress).

§8.14 nonexistence of a liability rule leading to optimal levels of activity. As no rule that has been examined induces both injurers and victims to choose optimal levels of their activities, one is led to ask whether there exists any conceivable liability rule which always results in optimal levels of activities. The answer is "No." The reason, in essence, is that for injurers to choose the correct level of their activity, they must bear accident losses, while for victims to choose the correct level of their activity, they too must bear accident losses. Yet injurers and victims cannot each bear accident losses.

Three comments should be made about this conclusion. First, the explanation just given for it directly suggests methods that in principle would lead to optimal behavior. For example, let injurers pay fines to the state equal to harm done -- or taxes equal to expected harm -- and let victims bear their losses. Then the expected payments of injurers and of victims would each equal expected accident losses and they would each choose optimal levels of their activity.45
Second, the conclusion depends on the assumption that courts cannot incorporate parties' levels of activity into the negligence or contributory negligence determination (an assumption that may be justified by what was said in §B.7). If negligence and contributory negligence could be defined in terms of levels of activity as well as levels of care, then the usual liability rules would lead injurers and victims to choose optimal levels of care and of their activity. Third, the conclusion should not be interpreted as an unduly negative one. As more factors are incorporated into a model, it naturally becomes less likely that a hypothetically ideal outcome can be achieved.

§B.15 remark on the reciprocal nature of harm. It is a truism that harm has a reciprocal aspect in that a victim must be present to suffer harm just as much as an injurer must be present to do harm. This observation has sometimes been said to imply that injurers should not necessarily pay for harm done, that harm should not necessarily be "internalized" to injurers. That, of course, was a conclusion of the analysis here. As emphasized in §B.13, either strict liability or negligence rules could turn out to be best.

But the fact that harm has a reciprocal aspect has also occasionally been suggested to mean that it is conceptually impossible to decide whether strict liability or the negligence rule should be applied, and even that the very notion of harm is rendered ambiguous. There is, however, no difficulty in principle in deciding in models whether strict liability or the
negligence rule will be better in a given situation (there was no difficulty in deciding the question in Example 4, for instance), and there is nothing problematic about the notion of harm.

§B.16 **Actual use of strict liability and negligence rules.** The choice between the two main forms of liability has been made in approximately the same way for accidents between strangers in different legal systems. Namely, negligence is the usual basis of liability; strict liability is employed only in certain areas of accident. In Anglo-American law, liability for accident losses is "for most significant purposes governed by the concept of negligence"; use of strict liability is restricted to harms caused by wild animals, to certain types of harms due to fire, and to harms arising from "abnormally dangerous" or "ultrahazardous" activities (such as blasting, storage of flammable liquids, transport of nuclear materials). Most of the provisions of the German Civil Code impose liability only if the injuring party was at fault; strict liability is adopted in connection with harms due to animals other than domestic animals, and, according to special legislation, in connection with harms arising from rail, road, and air traffic and from use of electricity, gas, and atomic energy. The situation in France is similar. Two important articles of the Civil Code specify fault or negligence as the general principle of liability; strict liability applies to harms due to animals or to certain dangerous inanimate objects (including automobiles and
aircraft). Likewise, in the Soviet Union, fault is the general basis of liability, with strict liability being reserved for harms due to "sources of increased danger."

§8.16 remark on the actual use of strict liability and negligence rules in light of the theory concerning levels of activity. As stressed in the analysis, use of strict liability rather than the negligence rule in areas where activities create high risks despite the exercise of reasonable care has the advantage that it will tend to reduce in a desirable way participation in these activities.

This theoretical advantage seems consistent with fact in the sense that the impression given by the description of the foregoing section is that the areas of strict liability are generally more dangerous than those of negligence (certainly the reverse is not true). The areas of strict liability do not appear uniformly more dangerous than those of negligence, however; the choices made between strict liability and negligence are not always easy to explain on the basis of differences in riskiness. (Is the chance of a wild animal escaping from the zoo and doing harm, for which strict liability would probably apply in the United States, clearly greater than that of an automobile running down a pedestrian, for which the negligence rule would govern?) Moreover, differences among countries in the areas of strict liability and negligence are sometimes difficult to explain in terms of differences in dangerousness. (Why should the negligence rule govern liability for automobile-pedestrian accidents in the
United States, but strict liability apply in Germany, the Soviet Union, and France? It seems that the conformity of the observed pattern of use of strict liability and negligence rules to what would be suggested by the theoretical considerations of this chapter is somewhat rough.

Putting aside questions concerning the actual dangerousness of the areas of strict liability versus those of negligence, it should be emphasized that one of the aims of the law is to impose strict liability on activities that are dangerous, or, more precisely, that are dangerous even if conducted with reasonable care. A particularly direct expression of this aim is provided by the Restatement of Torts, Second, which says that in deciding whether an activity should be subject to strict liability, notice ought to be taken of possible "inability to eliminate risk by the exercise of reasonable care." Further, the Restatement draws a contrast with most "ordinary activities" which can be made "safe by the taking of all reasonable precautions" and for which liability should be based on negligence.

But it should be added that the deterrent effect of strict liability on the level of participation in activities is not mentioned in the Restatement and is only infrequently noted in other places. Evidently, the mere creation of an unusual risk is seen as a justification for imposition of strict liability.

§8.16 note on the literature. In Shavell (1980a), I introduced the issue of the choice of the level of activity as distinct from the level of care and developed the points of
this part. The issue is elaborated upon in Landes and Posner (1981a).
Appendix to Chapter Two

Liability and Deterrence: Basic Theory

The main elements of the theory concerning how liability affects behavior that influences accident risks will be presented formally here.

The assumptions will be much as were described in the text: There is a single good, in terms of which all variables are defined. Parties are risk neutral; that is, a party's (von Neumann-Morgenstern) utility is taken to equal the amount of the good he possesses, so that his expected utility equals the expected amount of the good he possesses.\(^1\) The measure of social welfare\(^2\) is the sum of parties' expected utilities; equivalently, it is the sum of the expected amounts of the good that they possess.

Parties are of two types, injurers and victims, and are strangers to one another. Injurers are all identical, as are victims, and only the victims suffer losses if accidents occur. The probability and severity of accidents may be influenced by parties' behavior.

If an accident occurs, a rule of liability will apply. The rule will determine whether and how much the injurer involved shall pay the victim.

The behavior of parties in equilibrium will be studied under various liability rules and will be compared with social welfare maximizing behavior. An equilibrium is a
situation where injurers have no motive to alter their behavior, taking as given victims' behavior and the liability rule; and likewise, where victims have no motive to alter their behavior, taking as given injurers' behavior and the liability rule.3

The analysis of equilibrium behavior under liability rules will be carried out first assuming that parties affect accident risks only through their choice of their levels of care; and then assuming that they affect accident risks through their choice of their levels of activity as well. In each case, unilateral accidents -- in which injurers' behavior alone affects risks -- and bilateral accidents -- in which victims' behavior too affects risks -- will be separately considered.

The notation, definitions of liability rules, and certain other assumptions will be introduced in the course of the analysis.

A. Levels of care the only determinant of risk

§A.1 unilateral accidents. Let

\[ x = \text{level of care of an injurer; } x \geq 0; \text{ and } \]

\[ \ell(x) = \text{expected accident losses}^4 \text{ caused by an injurer given } \]

\[ x; \ell(x) \geq 0; \ell'(x) < 0 \text{ and } \ell''(x) > 0 \text{ where } \ell \text{ is positive.} \]

Thus, the exercise of care reduces expected accident losses, but at a decreasing rate.
The social goal will be to minimize total accident costs,⁵

\[(2.1) \ x + \ell(x).\]

Let \(x^*\) denote the \(x\) that minimizes \(2.1\). This socially optimal value, \(x^*\), is unique, by our assumptions, and is illustrated in Figure 1. Assuming as we shall that \(x^*\) is positive, it is determined by the first-order condition

\[1 = -\ell'(x),\]

that is, the marginal cost of care must equal the marginal benefit in terms of the reduction in expected accident losses.⁶

Consider now injurers' behavior under liability rules.⁷ As injurers will seek to maximize the expected amount of the good they hold, they will act to minimize their expected expenses, namely, \(x\) plus their expected liability. Hence under the "rule" of no liability, injurers will choose \(x = 0\); total accident costs will thus be \(\ell(0)\) and will not be minimized. Under strict liability, injurers will be liable for any accident losses they cause. Therefore, an injurer's expected liability will be \(\ell(x)\) and he will minimize \(2.1\) and choose \(x^*\); this is shown in Figure 1. Under the negligence rule, injurers will be liable for accident losses they cause if and only if their level of care was less than a due care level, \(\bar{x}\), specified by courts. If \(\bar{x} = x^*\), then, as should be apparent from Figure 1, injurers will choose \(x^*\): Certainly, an injurer will not choose \(x > x^*\), for if \(x\) merely equals \(x^*\), he will avoid liability and his
expenses will be \( x^* \); choosing a higher \( x \) would cost him more but give him no advantage. To show that an injurer will not choose \( x < x^* \), observe that if he did, he would be liable for accident losses caused, so his expected expenses would be \( x + \ell(x) \). Moreover, since \( x \) would be unequal to \( x^* \) and since \( x^* \) is unique,

\[ x + \ell(x) > x^* + \ell(x^*) \]

would be true. And since \( x^* + \ell(x^*) \geq x^* \),

\[ x + \ell(x) > x^* \]

would hold. This, however, says that the injurer's expected expenses would be higher than if he chose \( x^* \), and establishes the claim.

We may summarize as follows.

**Proposition 1.** In the absence of liability, injurers will take no care. Under strict liability they will choose the socially optimal level of care; and they will do so under the negligence rule as well, assuming that due care equals the socially optimal level.

**Remarks.** (i) To employ strict liability, courts need only observe \( \ell \). To employ the negligence rule, courts need to observe \( \ell \) and \( x \) and to know the function \( \ell(\cdot) \) (in order to calculate \( x^* \)).

(ii) If care \( x \) is a multidimensional variable, the proofs that strict liability and the negligence rule will lead to the socially optimal outcome \( x^* \) still apply. However, if there are dimensions of \( x \) that courts cannot observe and cannot include in the due care standard under
the negligence rule, then injurers will not choose optimal levels of these dimensions of x under that rule.

§A.2 bilateral accidents. Let
\[ y = \text{level of care of a victim; } y \geq 0; \]
and redefine \( l \) as
\[ l(x, y) = \text{expected accident losses given } x \text{ and } y; \ l(x, y) \geq 0; \]
\[ l_x(x, y) < 0 \text{ and } l_y(x, y) < 0 \text{ where } l \text{ is positive;} \]
\[ l \text{ is a strictly convex function of } x \text{ and } y \text{ where } l \text{ is positive.} \]

Thus, the care of both injurers and victims now affects expected accident losses, and the care of each lowers expected losses, at a decreasing rate, given the care of the others.

The social goal will be to minimize total accident costs, which are here
\[(2.2) \ x + y + l(x, y). \]
Let \( x^* \) and \( y^* \) denote the (unique) socially optimal values of \( x \) and \( y \), and assume that these are positive. Hence \( x^* \) and \( y^* \) satisfy
\[ l = -l_x(x^*, y^*), \]
\[ l = -l_y(x^*, y^*). \]
Also, let \( x^*(y) \) be the value of \( x \) that minimizes \( (2.2) \) given \( y \); equivalently, therefore, \( x^*(y) \) minimizes \( x + l(x, y) \); and let \( y^*(x) \) be defined similarly. Observe from these definitions that \( x^* = x^*(y^*) \) and \( y^* = y^*(x^*). \)
Now consider the actual behavior of parties, assuming that injurers will minimize their expenses taking as given victims' level of care, and that victims will minimize their expenses taking as given injurers' level of care. If there is no liability, injurers will choose $x = 0$ (whatever is $y$); thus, in particular, the outcome will not be socially optimal. Also, as $x = 0$, victims will select $y$ to minimize $y + \ell(0,y)$. Hence, they will choose $y^*(0)$.

Under strict liability, victims will choose $y = 0$ (whatever is $x$) since they will be compensated for any losses they sustain; thus, again, the outcome will not be socially optimal. As $y = 0$, injurers will choose $x$ to minimize $x + \ell(x,0)$, so they will choose $x^*(0)$.

Under strict division of accident losses, injurers must pay a fraction $f$ of losses caused, where $0 < f < 1$. Hence, injurers will choose $x$ to minimize $x + f\ell(x,y)$. Thus, if $x$ is positive, it will satisfy the first-order condition $1 + f\ell_x(x,y) = 0$; but as $x^*(y)$ satisfies $1 + \ell_x(x,y) = 0$ and as $f < 1$, $\ell_x < 0$, and $\ell_{xx} > 0$, it follows that $x$ will be less than $x^*(y)$. Similarly, victims will choose $y$ to minimize $y + (1 - f)\ell(x,y)$, so that $y$ will be less than $y^*(x)$. In particular, then, the outcome will not be socially optimal.

Under strict liability with the defense of contributory negligence, an injurer will be liable for accident losses he causes unless the victim's level of care $y$ was less than a due care level $\bar{y}$, in which case the victim must bear his losses. Let us establish that if $\bar{y} = y^*$, then both injurers
and victims acting socially optimally is an equilibrium and it is the only equilibrium. To show this, let us suppose first that \( y = y^* \), and verify that injurers will choose \( x = x^* \). Now as \( y = y^* = \bar{y} \), victims will not bear their losses; an injurer's expected liability will therefore be \( l(x,y^*) \), and he will choose \( x \) to minimize \( x + l(x,y^*) \); hence he will choose \( x^*(y^*) = x^* \). Next, suppose that \( x = x^* \). Then a victim will not choose \( y > y^* \), for he will avoid having to bear his losses merely by choosing \( y = y^* \); and he will not choose \( y < y^* \) by the logic of the argument in the unilateral case demonstrating that under the negligence rule injurers would not take less than due care. Hence, the victim will choose \( y^* \).

It remains to prove that there does not exist any other equilibrium (given that \( \bar{y} = y^* \)). To prove this, let us assume there is another equilibrium and demonstrate that that leads to a contradiction. In another equilibrium, it must be that \( y \neq y^* \) (for if \( y = y^* \), the argument in the preceding paragraph shows that \( x = x^* \), which would contradict the assumption that the equilibrium is different). Now if \( y > y^* \), victims would obviously be better off reducing \( y \), for they can avoid liability so long as \( y \geq y^* \). Hence, it must be that \( y < y^* \) in the equilibrium. But if this is true, victims will bear their losses, so injurers will choose \( x = 0 \). Thus, a victim's expected expenses must be \( y + l(0,y) = 0 + y + l(0,y) > x^* + y^* + l(x^*,y^*) > y^* \), which means that victims would have been better off choosing \( y^* \) and avoiding bearing their losses, a contradiction.
The rule of strict liability with the defense of relative negligence differs from the rule just considered only in that if the victim's care $y$ was less than due care $\bar{y}$, then the victim will bear a fraction $f$ of his losses, where $0 < f < 1$ and $f = f(y)$ with $f'(y) < 0$. If the limit of $f$ as $y$ approaches $\bar{y}$ is sufficiently close to 1, then it can be verified using the arguments of the last two paragraphs that if $\bar{y} = y^*$, both injurers and victims acting socially optimally is a unique equilibrium.

Next, let us consider the negligence rule and show that if $\bar{x} = x^*$, both injurers and victims acting socially optimally is a unique equilibrium. Suppose first that $y = y^*$. Then the argument concerning injurers' behavior under the negligence rule in the unilateral case can be applied to show that injurers will choose $x^*$. Now suppose that $x = x^*$. Then, as victims will bear their losses, they will choose $y$ to minimize $y + \ell(x^*,y)$, which means that they will choose $y^*(x^*) = y^*$.

To demonstrate that the equilibrium is unique (given that $\bar{x} = x^*$), essentially the same argument that was used under strict liability with the defense of contributory negligence can be employed. The argument is thus only sketched: If there is another equilibrium, it must be such that $x \neq x^*$ (since if $x = x^*$, then $y = y^*$); but $x > x^*$ cannot be, for an injurer could always do better by choosing a lower $x$ that exceeds $x^*$. Hence, $x < x^*$. If, however, this is true, victims will not bear their losses, so they
will choose $y = 0$. This leads to a contradiction as before (reversing the roles of victims and of injurers).

Under the negligence rule with the defense of contributory negligence, an injurer will be liable for accident losses he causes if and only if two conditions are met: his care was less than his due care level $\bar{x}$; and the victim's care was not less than his due care level $\bar{y}$. Otherwise the victim will bear his losses. We wish to show that if $\bar{x} = x^*$ and $\bar{y} = y^*$, then injurers and victims acting socially optimally is a unique equilibrium. Thus, assume that $y = y^*$. Then, since injurers will be liable if $x < x^*$, the argument used in the analysis of the negligence rule in the unilateral case may be applied to show that injurers will choose $x^*$. Now assume that $x = x^*$. Then, since victims will bear their losses, they will, as under the negligence rule, decide to choose $y^*$.

To show that the equilibrium is unique, assume otherwise, so that $x \neq x^*$ (for $x = x^*$ implies $y = y^*$) and $y \neq y^*$ (for $y = y^*$ implies $x = x^*$). Now $x > x^*$ cannot be, for injurers could then do better by lowering $x$ slightly. Hence, $x < x^*$. But then $y > y^*$ cannot be, for since $x < x^*$, victims can still avoid bearing losses if they lower $y$ slightly; thus $y < y^*$. This, however, means that victims will necessarily bear their losses, so that injurers will choose $x = 0$. Hence, a victim's expected expenses must be $y + \ell(0,y) > y^*$ (as shown in the argument regarding strict liability with the defense of contributory negligence),
implying that the victim would have been better off choosing \( y^* \) and thereby shifting his losses to injurers.

Finally, under the **comparative negligence rule**, an injurer will be liable for a fraction \( f \) of accident losses he causes if his care \( x \) was less than \( \bar{x} \) and the victim's care was less than \( \bar{y} \), where \( 0 < f < 1 \) and \( f = f(x,y) \) (usually, with \( f_x < 0 \) and \( f_y > 0 \)). On the other hand, if \( x \) was less than \( \bar{x} \) and \( y \) was at least \( \bar{y} \), the injurer will be liable for the entire losses; and if \( x \) was at least \( \bar{x} \), the victim will bear his own losses. Under this rule too, if \( \bar{x} = x^* \) and \( \bar{y} = y^* \), then both injurers and victims acting socially optimally is a unique equilibrium. The proof that this is an equilibrium is identical to that under the negligence rule with the defense of contributory negligence, and verification is left to the reader. To show that the equilibrium is unique, suppose otherwise. Then reasoning as in the case of the negligence rule with the defense of contributory negligence, we deduce that \( x < x^* \) and \( y < y^* \). But this means that \( x + f\ell(x,y) < x^* \) (else \( x < x^* \) would not have been chosen) and similarly that \( y + (1 - f)\ell(x,y) < y^* \). Adding these two inequalities gives \( x + y + \ell(x,y) < x^* + y^* \), which contradicts the social optimality of \( x^* \) and \( y^* \).

**Proposition 2.** In the absence of liability or under strict liability, the outcome will generally not be socially optimal: injurers will take no care in the absence of liability and victims will take no care
under strict liability; injurers, however, will choose the optimal level of care given the behavior of victims under strict liability, and conversely for victims in the absence of liability. Under strict division of accident losses as well, the outcome will generally not be optimal; victims will take less care than is optimal given injurers' behavior, and vice versa.

But under each of the other rules -- strict liability with the defense of contributory or relative negligence, the negligence rule with or without the defense of contributory negligence, and the comparative negligence rule -- the outcome will be socially optimal. Specifically, if due care levels are socially optimal, then both injurers and victims taking socially optimal levels of care will be an equilibrium, and no other equilibrium exists.

Remarks. (i) To employ any of the rules leading to the socially optimal outcome, courts need to observe $l$ and to know the function $l(\cdot, \cdot)$ (to be able to calculate either $x^*$ or $y^*$). The differences among the rules are thus that courts need to observe only $y$ under strict liability with the defense of contributory (or relative) negligence; only $x$ under the negligence rule; but both $x$ and $y$ under the negligence rule with the defense of contributory negligence. See, however, the discussion in §A.18 of the text.

(ii) If $x$ and $y$ are multidimensional variables, then the liability rules stated as leading to socially optimal
behavior will continue to do so. If, however, there are
dimensions of x or y not included in injurers' or victims'
standards of due care, the situation changes. See the
discussion in §A.19 of the text.

§A.3 note on the literature. The results presented
here were, except for several minor differences, first
proved by Brown (1973); but the result that use of the
negligence rule with the defense of contributory negligence
leads to socially optimal behavior was also shown shortly
after by Diamond (1974a,b).

B. Levels of care and levels of activity the determinants of
risk

§B.1 unilateral accidents. Let us now reconsider the
unilateral case assuming that injurers may vary their level
of activity as well as their level of care. Define

\[ s = \text{level of activity of an injurer}; \ s \geq 0; \text{ and} \]
\[ u(s) = \text{gross utility to an injurer of engaging in his} \]
\[ \text{activity at level } s; \ u(s) > 0, \ u'(s) > 0, \ u''(s) < 0 \text{ for } s < \hat{s}; \ u'(') = 0. \]

Thus the gross utility increases with the level of activity,
but at a decreasing rate, up to some point of satiation \( \hat{s} \).
The net utility to an injurer in the absence of liability
will be assumed to equal the gross utility less his costs of
care,

\[ (2.3) \ u(s) - sx; \]
the interpretation is that if an injurer engages in his activity $s$ times taking care of $x$ each of these times, his total costs of care will be $sx$. The expected accident losses the injurer causes will similarly be assumed to equal $s\ell(x)$.

The social goal will be to maximize the net utility injurers obtain from their activities less expected accident losses,\textsuperscript{11}

\begin{equation}
(2.4) \quad u(s) - sx - s\ell(x) = u(s) - s[x + \ell(x)].
\end{equation}

Denote the optimal values of $s$ and $x$ by $s^*$ and $x^*$ and assume that they and $\ell(x^*)$ are positive. From the right-hand side of (2.4), it is clear that $x^*$ is the $x$ that minimizes $x + \ell(x)$; hence $x^*$ is the optimal $x$ described in §A.1. It also follows from the right-hand side of (2.4) that $s^*$ is determined by maximizing $u(s) - s[x^* + \ell(x^*)]$. Thus, $s^*$ is determined by

\begin{equation}
(2.5) \quad u'(s) = x^* + \ell(x^*),
\end{equation}

the interpretation of which is that the marginal utility from an increase in the level of activity must equal the costs of taking optimal care plus the increase in expected accident losses.

Consider now injurers' behavior. If there is no liability, an injurer's utility will be $u(s) - sx$, so he will choose $x = 0$ and $s$ such that $u'(s) = 0$, that is, $s = \hat{s}$, the activity level of satiation. Thus, the level of activity will be excessive (this follows from the facts that $u'(\hat{s}) = 0$; that (from (2.5)), $u'(s^*) > 0$, and that $u'' < 0$).
Under strict liability, since injurers will be liable for all accident losses that they cause, they will choose \( s \) and \( x \) to maximize \( u(s) - s[x + l(x)] \), namely, (2.4). Hence, they will choose \( x^* \) and \( s^* \).

Under the negligence rule, assume first that the due care level \( \bar{x} \) is \( x^* \). Then injurers will wish to take due care to avoid liability whatever \( s \) they choose; the proof of this is essentially the same as the proof of the result of §A.1 that injurers will take due care and is omitted.

Consequently, injurers will choose \( s \) to maximize \( u(s) - sx^* \), so that \( s \) will be determined by

\[
(2.6) \quad u'(s) = x^*.
\]

Comparing (2.6) to (2.5), we see that \( s \) will exceed \( s^* \). (As emphasized in the text, the problem is that because injurers avoid liability by taking due care, the marginal cost to them of increasing their level of activity is only the cost of due care \( x^* \); it does not include the increase in expected accident losses \( l(x^*) \).)

Let us demonstrate that social welfare under the negligence rule would be raised if a due care level \( \bar{x} \) somewhat higher than \( x^* \) were employed. In proving this, we restrict attention to \( \bar{x} \) to which injurers would be induced to adhere.\(^{12}\) (If \( \bar{x} \) is so high that injurers would not adhere to it, the negligence rule devolves into strict liability -- which we know would result in a socially optimal outcome -- and it does not seem natural to analyze this under the heading of the negligence rule.) Thus, injurers will choose \( s \) to
maximize \( u(s) - s\bar{x} \), so that \( s \) will obey the first-order condition \( u'(s) = \bar{x} \). Writing \( s = s(\bar{x}) \), differentiating (2.4) with respect to \( \bar{x} \), and making use of (2.6), we obtain

\[
(2.7) - s'(\bar{x})l(\bar{x}) - s(\bar{x})[1 + l'(\bar{x})].
\]

This is positive for \( \bar{x} \leq x^* \): As \( s' \) is negative (differentiate \( u'(s(\bar{x})) = \bar{x} \) to obtain \( s'(\bar{x}) = 1/u''(s(\bar{x})) < 0 \), the first term is positive, and the second term is non-negative since \( x + l(x) \) is convex in \( x \) and its derivative \( 1 + l'(x) \) is zero at \( x^* \). As (2.7) is positive for \( \bar{x} \leq x^* \), it must be socially beneficial to choose an \( \bar{x} \) exceeding \( x^* \). (The explanation is that raising \( \bar{x} \) makes engaging in his activity more costly to an injurer and thus offers an indirect means of reducing his excessive activity level; on the other hand, the first-order direct effect on social welfare of raising \( \bar{x} \) above \( x^* \) is zero, for \( x^* \) is socially optimal.)

The results are summarized as follows.

**Proposition 3.** In the absence of liability, the outcome will not be socially optimal: injurers will not take care and will choose too high a level of activity.

Under strict liability, the outcome will be socially optimal: injurers will take optimal care and will choose the optimal level of activity.

Under the negligence rule, the outcome will not be socially optimal: if due care equals the socially optimal level of care, injurers will take optimal care but their level of activity will be excessive. Also, if due care is raised above socially optimal care,
social welfare will be higher than if due care equals socially optimal care.

Remark. If under the negligence rule, an injurer would be liable if either his care \( x \) was less than \( x^* \) or if his activity level \( s \) exceeded \( s^* \), then he would choose \( x^* \) and \( s^* \). As suggested in the text, however, courts probably do not have sufficient information about \( u(\cdot) \) to compute \( s^* \) reasonably well.

§B.2 Bilateral accidents. Finally, let us consider the general case where both injurers and victims choose levels of care and levels of activity. Thus, define

\[
\begin{align*}
t &= \text{level of activity of a victim; } t \geq 0; \text{ and} \\
v(t) &= \text{gross utility to a victim of engaging in his activity at level } t; \; v(t) > 0, \; v'(t) > 0, \text{ and } v''(t) < 0 \text{ for } t < t; \; v'(t) = 0.
\end{align*}
\]

The interpretation of \( t, \dot{t}, \text{ and } v(t) \) is analogous to that of \( s, \dot{s}, \text{ and } u(s) \). Expected accident losses will be assumed to equal \( \text{SEL}(x,y) \), and the net utility to a victim of engaging in his activity to equal

\[
(2.8) \quad v(t) - ty - \text{SEL}(x,y)
\]

exclusive of any liability payments received.

The social goal will be to maximize the sum of injurers' and victims' net utilities,

\[
(2.9) \quad [u(s) - sx] + [v(t) - ty - \text{SEL}(x,y)] = u(s) + v(t) - [sx + ty + \text{SEL}(x,y)],
\]
that is, their gross utility from engaging in their activities less their costs of care and expected accident losses. It will be assumed that there is a unique optimum and that all variables are positive at the optimum; the optimal variables will be denoted by an asterisk. The first-order conditions determining \( x^*, y^*, s^*, \) and \( t^* \) are

\[
1 + t_\ell x(x, y) = 0, \\
1 + s_\ell y(x, y) = 0, \\
u'(s) = x + t_\ell(x, y), \\
v'(t) = y + s_\ell(x, y).
\]

The interpretation of the first two conditions is analogous to that made in §A.2, and the interpretation of the latter two conditions is analogous to that made with respect to \( s \) in (2.5).

Rather than to reconsider each of the different liability rules of the previous section, which would be tedious, let us discuss only strict liability with the defense of contributory negligence and the negligence rule, and suggest why neither result in the socially optimal outcome. Then it shall be proved that in fact there does not exist any liability rule which would result in social optimality.

Under strict liability with the defense of contributory negligence, suppose that the due care level \( \bar{y} \) is \( y^* \) and that victims would be led to adhere to it (otherwise the outcome would clearly be suboptimal). Then victims will choose \( t \) to maximize \( v(t) - t\bar{y} \), so \( t \) will satisfy \( v'(t) = \bar{y} \); thus victims will generally choose a socially excessive level of their
activity given injurers' behavior (for from (2.10), social optimality requires \( v'(t) = \bar{y} + s\ell(x, \bar{y}) \)). Injurers, however, will choose the socially optimal \( x \) and \( s \) given victims' behavior. This is because injurers, being liable for losses, will maximize \( u(s) - [sx + st\ell(x, \bar{y})] \), which differs from (2.9) by a term that does not depend on \( x \) or \( s \).

Similar reasoning (essentially that of the last section) shows that under the negligence rule (with or without the defense of contributory negligence), while injurers may be induced to take due care, they will engage in too high a level of their activity. Victims, however, will take socially correct levels of care and of their activity given injurers' behavior.

Comparing the two rules, it is clear that either could result in a higher level of social welfare; very roughly, strict liability with the defense of contributory negligence will be the better rule if the problem of controlling injurers' level of activity is more important than that of controlling victims' level of activity; otherwise the negligence rule (with or without the defense) will result in the higher level of welfare. (This statement could be made precise by introducing parameters describing the influence of activity levels on utility and on expected accident losses, but that does not seem worthwhile.)

Now let us prove that there does not exist any liability rule that will induce injurers and victims to act socially optimally. The liability rules considered are assumed to be
functions depending on the amount of losses and possibly on injurers' and victims' levels of care, but not on their levels of activity. Moreover, the expected payments of an injurer under a liability rule are assumed to be of the form \( stq(x,y) \), where \( q(x,y) \) is his expected liability given \( x \) and \( y \) if his and the victim's level of activity are each 1.

(For instance, under strict liability with a defense of contributory negligence, \( q(x,y) \) would equal \( l(x,y) \) if \( y \geq \overline{y} \) and 0 otherwise.) Supposing, then, that a liability rule that leads to optimal behavior exists, injurers in particular must behave optimally. Thus, injurers must choose \( x^* \), and they must also choose \( s^* \). The latter implies that \( s^* \) is the solution to the injurers' problem,

\[
\max_{s} u(s) - sx^* - st^*q(x^*,y^*) = u(s) - s[x^* + t^*q(x^*,y^*)],
\]

so that \( u'(s^*) = x^* + t^*q(x^*,y^*) \). On the other hand, since \( s^* \) is socially optimal, we know from (2.10) that \( u'(s^*) = x^* + t^*l(x^*,y^*) \). From this and the fact that \( u'' < 0 \), we obtain \( x^* + t^*q(x^*,y^*) = x^* + t^*l(x^*,y^*) \); and from this and the fact that \( t^* \) is positive, we conclude that \( q(x^*,y^*) = l(x^*,y^*) \). (The expected liability of injurers must have been equal to expected losses for them to have chosen the correct level of activity.) Similarly, \( t^* \) must be the solution to the victims' problem,

\[
\max_{t} v(t) - ty^* - s^*t^*l(x^*,y^*) + s^*t^*q(x^*,y^*) = v(t) - t[y^* + s^*l(x^*,y^*) - s^*q(x^*,y^*)],
\]

so that \( v'(t^*) = y^* + s^*l(x^*,y^*) \). But from (2.10), \( v'(t^*) = y^* + s^*l(x^*,y^*) \), which leads to the conclusion
that \( l(x^*, y^*) - q(x^*, y^*) = l(x^*, y^*) \), or that \( q(x^*, y^*) = 0 \). But this contradicts what we earlier showed, that \( q(x^*, y^*) = l(x^*, y^*) \).

We thus have

**Proposition 4.** None of the usual liability rules will lead injurers and victims to act socially optimally; strict liability with the defense of contributory negligence will result in victims engaging in too high a level of their activity, and negligence rules will result in injurers engaging in too high a level of their activity. In fact, there does not exist any liability rule that will lead both injurers and victims to engage in optimal levels of their activities.

**Remarks.** As should be clear from the remark following Proposition 3, the result that there does exist any liability rule that leads to optimal levels of activities depends on the assumption that liability rules are not functions of levels of activity. Otherwise, it would be easy to design rules leading to optimality.\(^{14}\) The result also depends on the characteristic of a liability rule that what the injurer pays, the victim receives. Were this not true, it would again be easy to induce parties to act optimally; for instance, if injurers were to pay fines to the state equal harm done and were victims to bear their losses, both would be led to act optimally.

\(^{14}\)\textit{§B.3 note on the literature.} The analysis of the last two sections is taken from Shavell (1980a), where I intro-
duced the issue of the choice of levels of activity as distinct from levels of care.
Chapter Four

FACTORS BEARING ON THE NEGLIGENCE DETERMINATION

Having analyzed the basic theory of liability and deterrence where accidents involve strangers and where they involve firms and strangers or customers, I begin in this chapter a more detailed consideration of the liability system, and take up three topics concerning the functioning of the negligence rule. The first topic has to do with whether differences among parties should result in differences in standards of due care. The second topic deals with certain "prior" decisions of parties (such as whether to obtain information about risk) that affect their ability to exercise care later, at the time when accidents may occur. And the third topic involves uncertainty and error over the negligence determination.

A. Differences among parties

Parties may differ with respect to the costs they incur in exercising care and with respect to the effect that their exercise of care will have in reducing accident risks. For example, individuals presumably differ in their ability to clear ice from their sidewalks, and if they do so, will reduce accident risks by varying degrees (depending on the amount of foot traffic, and the like). In what follows, reference will be made, for simplicity, only to differences in parties' cost of taking care, although what will be said will plainly apply
also to differences in the effectiveness of their exercise of care. The relevance of individual differences in the cost of taking care to the determination of due care levels, to levels of care actually taken, and to levels of activity will now be analyzed.

§A.1 levels of care. The first point to make is that the socially optimal level of care of a party depends on his cost of taking care. The socially optimal level of care of a party for whom the cost of taking care is low will generally be higher than the optimal level of care of a party for whom the cost of taking care is high. Thus it may be desirable for a young, able-bodied person to clear his sidewalk of ice, but undesirable for an elderly individual to do that.

Because the socially optimal level of care will vary among parties, levels of due care must vary among them if they are to be led to act optimally under the negligence rule. Accordingly, where it is simple for courts to determine differences in the cost of taking care among parties, this should be done, and levels of due care should be individualized. If courts can distinguish the young and able-bodied person who can readily clear his sidewalk of ice from the elderly person who cannot, the first should be found negligent for failing to clear ice but not the second.

On the other hand, where differences in the cost of taking care among parties in some class are hard (or impossible) for courts to assess, then it may be best for courts to employ (or there may be no choice but for them to employ) a uniform level
of due care in determining negligence for the parties. For instance, courts would presumably experience problems were they to attempt to ascertain how difficult it was to clear ice on a given day for a particular young and able-bodied individual. (To ascertain this, courts would have to evaluate that particular individual's stamina, how well he happened to have been feeling on the given day, and so forth.) Hence, courts may well find it best to determine negligence for the whole class of young and able-bodied individuals by reference to a single amount of ice that must have been cleared to avoid liability. For courts to choose this amount of ice optimally, they can be imagined to do the following: to begin by considering a low amount of ice as a standard, and then to raise the contemplated standard as long as the advantage of doing so -- that a greater number of individuals who ought to clear at least the contemplated standard amount of ice will be led to do that -- outweighs the disadvantage -- that a greater number of individuals who ought not to clear the contemplated amount of ice will nevertheless be led to do that. In other words, the optimal uniform level of due care involves an implicit balancing of advantages and disadvantages due to the varying cost of taking care among individuals.

It should be noticed that the optimal uniform level of due care will actually correspond to the individually optimal level of care for some "representative" individual within the class of individuals under consideration (though the optimal uniform level of due care will be either too high or too low for all
other individuals). Thus one can, if one wants, phrase the problem of determining the optimal uniform level of due care as the problem of determining the optimal individualized level of due care for an appropriately chosen representative individual. ¹

§A.2 engagement in activities. If the choice whether to engage in an activity is now considered, a reason can be identified why it would be socially beneficial for courts to hold parties to some "moderate" level of due care -- even where courts know that a particular party's ability to have exercised this level of care is slight or that he simply could not have done so, and thus that the optimal level of care for him was lower than the moderate level.

The reason is that insisting on a moderate level of care will discourage from engaging in an activity those parties who would create especially high risks of accidents because of their inability to meet that level of care. As was discussed in §B.2 of Chapter 2, it is socially undesirable for a party to engage in an activity when the expected accident losses he would cause exceed the utility he would derive less his costs of care. Hence, when a party would not take moderate care and thus would cause high expected losses, it would be desirable for him to refrain from engaging in an activity (unless he would derive unusually great utility from it). This is what will happen if courts hold a party liable when he does not take moderate care; for then a party who cannot take moderate care will tend to decide against engaging in an activity. For
example, a peculiarly inept person might be dissuaded from practicing archery (unless he were extremely enthusiastic about the idea) if he thought that he would probably be found negligent for failing to meet the standard of due care required for archery to be reasonably safe. To put the point a little differently, with respect to the inept person, archery may be regarded as an ultrahazardous activity; thus it makes sense, in effect, to impose strict liability on the inept person if he engages in archery.

§ A.3 actual determination of due care in view of differences among parties. 5 Many differences among parties that affect their ability to take care, and that are relatively easy for courts to observe, influence the determination of due care. Blindness, lameness, or infirmity, for instance, may lower the standard of care to which an individual would otherwise be held, while an individual's strength or size, or his special knowledge or professional skill may raise it.

But the consideration of individual characteristics in determining negligence is limited in a variety of ways. Small, difficult to observe, differences in physical attributes are not taken into account by courts and, normally, neither are differences in intelligence or in temperament; instead, due care is found by asking what would be appropriate for a "reasonable man" possessing "average" intellect and physical powers. Moreover, in some circumstances, even an individual's easily ascertainable characteristics do not influence due care. A nearly blind person, a child, or a mental incompetent
would probably be held responsible for causing an automobile accident, even if he drove with all the care of which he was capable; and it is doubtful whether a person's clumsiness, if established, would help to relieve him of liability. In addition, the fact that a person might weigh the costs of exercising care or the significance of the probable harm in a manner that departs from community norms would not affect the determination of due care.

§A.4 Several justifications are offered by courts and legal scholars for why certain characteristics of a party are not considered in determining negligence or, as it is sometimes expressed, for why negligence is often found by reference to "external" or "objective" standards rather than to individual or "subjective" standards. One justification is "the obvious difficulties of proof," for instance, of "drawing any satisfactory line[s] between... variations of temperament [and] intellect." This was discussed in §A.1 above. (Recall that finding the optimal level of care for the class of individuals among whom the law cannot distinguish may be interpreted as finding the optimal level of due care for the fictitious reasonable man.) Another justification is the notion that if an individual acts in a highly dangerous way, he simply deserves to be held responsible for harm done. Thus the driver with very poor vision should be found liable for accidents caused by his failure to notice what a person with normal vision would have noticed; as should "a man... born hasty and awkward [who] is always having accidents and hurting himself or
his neighbors...[for] his slips are no less troublesome to his neighbors than if they sprang from guilty neglect." This notion is very close to that used generally to justify strict liability, and it comports with the discussion here in §A.2. A third justification for use of objective standards of due care is unwillingness to credit idiosyncratic sensitivities in the comparison of the costs of care against the risks of harm.  

§A.5  note on the literature. Diamond (1974a) and Posner (1977) make the argument of §A.1 that it is best for levels of due care to be individualized to the extent that the courts can distinguish differences among parties.  

B. Prior precautions

In this part, what will be called prior precautions -- actions taken by parties that influence their ability to take care at subsequent times -- will be examined. One important example of a prior precaution is acquisition of information about risk (safety features of a power tool, likelihood of adverse reactions to a drug). The possession of such information will enhance a party's ability to prevent accidents. Another example concerns consumption of alcohol, since if a person is careful about indulging, he will be better able to avoid doing harm, notably when driving.  

The determination of socially optimal prior precautions will now be considered, and then the incentive to take prior precautions under the negligence rule will be discussed.
§3.1 optimal prior precautions. Consider for simplicity the unilateral model of accidents, and suppose the social goal to be minimization of the costs of prior precautions, the costs of care, and expected accident losses. Then the socially optimal behavior of a party is easily determined, as is illustrated by

Example 1. Taking a prior precaution will enable a party to exercise care at a cost of 3; and taking care will lower the risk of a loss of 100 from 10% to 2%. On the other hand, not taking the prior precaution will mean the party will not be able to exercise care to reduce risk. Therefore, taking the prior precaution will allow the party to lower total accident costs by 5 (since the exercise of care will decrease expected accident losses by 8 at a cost of only 3). Hence, if the cost of the prior precaution is less than 5, it will be socially worthwhile.\13

More generally, a prior precaution will be socially worthwhile not only if its cost is sufficiently low, but also if its effect on the ability to take care results in a sufficiently high expected reduction in total accident costs.

§3.2 negligence rule. Assume that the level of due care used to determine negligence is that which is socially optimal provided that parties have taken optimal prior precautions. Then parties will be led both to take optimal prior precautions and to exercise optimal care. The claim, in other words, is that if the level of due care is not relaxed for drunk drivers but is held to the level appropriate for sober drivers, then
drivers will be induced both to be sober and to exercise the level of due care appropriate for sober drivers. Likewise, if individuals are held to the level of care that presume they have made reasonable efforts to learn about risk, they will be led to make such efforts as well as to take the then called-for level of care.

The reason for these conclusions is as follows. If parties do not take optimal prior precautions, they will find it excessively costly or impossible to take due care. Hence they will be led to take optimal prior precautions. And just because they will do so, they will be induced, by familiar logic, to exercise due care. For instance, in Example 1, suppose that the cost of the prior precaution is 1 (so that taking the precaution is socially optimal) and, accordingly, care must be taken to avoid being found negligent. If the party takes the prior precaution, since he will then find it in his interest to take care to avoid liability, his expenses will be only 1 + 3 or 4. But if he does not take the prior precaution, his expected liability will be 10. Therefore, he will take the prior precaution and exercise care.14

§8.3 remark. It should be stressed that under the version of the negligence rule just considered, courts do not determine whether a party actually took optimal prior precautions. They determine only whether a party took due care. This is an advantage to the extent that it would be difficult for courts to ascertain a party's prior precautions (what knowledge he really possessed, how much alcohol he in fact consumed).
$B.4$ actual determination of negligence in view of prior precautions. Courts normally determine negligence in the manner described in $B.2$: they hold parties to the level of due care that would be required if their prior behavior had been appropriate.\textsuperscript{15} Thus, courts generally hold individuals to the level of care that would be required of a sober individual.\textsuperscript{16} And courts usually assume that parties have made reasonable efforts to apprise themselves of risk. They state that parties "know or should have known" easily obtained facts about risk, that parties ought to possess "common knowledge" about risk; that if parties are able to acquire knowledge about risk (as are physicians regarding the chance of adverse reactions to a drug), they are deemed to have that knowledge.\textsuperscript{17} Similarly, in other contexts, courts usually presume that parties have taken appropriate prior precautions; in determining how quickly a driver should have been able to come to a halt, for example, courts will ordinarily assume that he had taken the prior precaution of keeping his brakes in good repair.

C. Uncertainty, error, and misperception

Factors leading to uncertainty over the finding of negligence, and the consequences of such uncertainty, will be considered in the initial sections of this part. Then the effect of systematic, anticipated error in the choice of due care levels will be analyzed. Last, the effect of misperception of due care levels will be discussed.
$\text{C.1 uncertainty over the finding of negligence.}$

One factor leading to uncertainty over the finding of negligence is that courts may err in assessing a party's true level of care. For example, a court might not accept a physician's claim that he had performed a diagnostic test (listened carefully to a person's heartbeat after a series of exercises) when in fact he had. The possibility that a court would make an error of this type might lead a physician to administer redundant or uncalled for tests (an electrocardiogram) because that would reduce his chance of being found negligent by mistake. Of course, the possibility that a court would make an opposite type of error may also exist. A court might conclude that a physician had taken proper care when in truth he had not. (Suppose medical records omitted information showing that a patient should have been referred to a specialist when this was not done.)

However, the significance of the two types of error is not likely to be the same. The disadvantage to a party of being found negligent by mistake is that he will have to pay the victim's losses. This disadvantage will often dominate in importance the savings in the cost of care that the party could obtain by reducing his level of care somewhat and hoping that he would erroneously escape liability if an accident occurred.

Hence, the reader should not be surprised to learn that a general consequence of uncertainty over the assessment of true levels of care is that parties will tend to be led to take more than due care -- and thus to take socially excessive levels of
care (presuming that due care is set at socially optimal levels). To illustrate, consider

Example 2. The probability of an accident that would cause a loss of 100 is related to the level of care as shown.

Table 4.1

<table>
<thead>
<tr>
<th>level of care</th>
<th>cost of care</th>
<th>accident probability</th>
<th>expected accident losses</th>
<th>social costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>0</td>
<td>15%</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>moderate</td>
<td>3</td>
<td>10%</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>high</td>
<td>5</td>
<td>9%</td>
<td>9</td>
<td>14</td>
</tr>
</tbody>
</table>

The socially optimal level of care, which is assumed to be due care, is moderate care. If there were no chance of mistake in courts' assessment of care, parties could avoid liability for sure by taking moderate care, at a cost of 3; they would not take high care, since that would involve a greater cost of 5.

Suppose, however, that there is a 33% chance of courts misperceiving care by one level, and a 5% chance of courts' misperceiving care by two levels. That is, there is a 33% chance that no care would be seen by courts as moderate care and a 5% chance that no care would be seen as high care. There is a 33% chance that moderate care would be seen by courts as none and a 33% chance that moderate care would be seen as high care. And there is a 33% chance that high care would be seen by courts as moderate care and a 5% chance that high care would be seen as none.

Then parties will take high care. If they take no care, their expected expenses will be 62% x 15% x 100 = 9.3 (since they
will mistakenly escape liability $33\% + 5\% = 38\%$ of the time). If they take moderate care, their expected expenses will be $3 + 33\% \times 10\% \times 100 = 6.33$ (since they will mistakenly be found liable $33\%$ of the time). Yet if they take high care, their expected expenses will be only $5 + 5\% \times 9\% \times 100 = 5.45$ (since they will mistakenly be found liable only $5\%$ of the time). //

As this example shows, if raising the level of care reduces the chance of being found negligent by mistake, parties may decide to take more than due care, even where the chances of courts overestimating care are as large as the chances of their underestimating care. 19 The example illustrates also the point that despite parties' increasing their level of care, they may still face a positive risk ($5\%$ in the example) of being found negligent if they cause accidents.

Much the same conclusions hold with respect to two other factors leading to uncertainty over the finding of negligence. One of these factors is that a party may be unable to control completely his momentary level of care. A driver may be unable to control completely his level of care at each instant (due to a lapse of attention, a sudden glare, a sneeze); or a physician may be unable to act with all the care he intends with each of his patients on each of their visits. But since it is the driver's care at the time of an accident and the physician's treatment of the particular patient that courts will ordinarily consider in their negligence determination, the driver and the physician will generally bear some uncertainty over whether they will be found negligent. A little reflection should convince the reader that such uncertainty will lead parties to
take more than due care in an average sense in order to reduce the likelihood of their momentary level of care falling short of due care and thus of their being found negligent. (The logic behind this assertion is essentially that of the last paragraphs, that the disadvantage of being found negligent will outweigh the advantage of conserving on the cost of taking care.)

The other factor leading to uncertainty over the negligence determination is uncertainty over the level of due care the courts will apply. This uncertainty will arise where the way in which courts will assess the cost of care or its effectiveness in reducing risk is uncertain. There may be uncertainty, for instance, over how courts will evaluate the cost to a physician in time and effort of performing a diagnostic test or over how courts will assess the value of the test in providing information about a disease, and therefore uncertainty over whether courts will see failure to perform the test as negligence. It should be clear to the reader that such uncertainty will tend to induce parties to take higher than desirable levels of care because this will help them to guard against being found liable by mistake.

§C.2 remarks on uncertainty. (a) The relative importance of the three sources of uncertainty -- courts' errors in assessing true levels of care, parties' inability to control their momentary level of care, and courts' errors in calculating levels of due care -- will depend on context. Where there are few witnesses to, or there is little evidence
concerning, a party's behavior, errors in assessing true levels of care may be important; where courts will not be able to obtain or to evaluate reliably information about the costs and benefits of care, errors in the calculation of the level of due care may be important (a problem that may be of general significance for physicians and other professionals, for firms using new technology); and so forth.

(b) With respect to parties' inability to control their momentary levels of care, three comments seem worth making. First, an individual's momentary level of care can be regarded as an imperfect indicator of his true, and inherently unobservable, level of care, namely, the degree to which he adopts a prudential mental attitude. Therefore, uncertainty over the finding of negligence due to an individual's inability to control his momentary level of care may in strict logic be viewed as due to courts' inability to assess an individual's true prudential mental attitude. Second, one wonders whether courts might sometimes lower the level of due care in implicit recognition of parties' problems in controlling their momentary level of care. (Might not courts allow for some irregularity in driving behavior, knowing that individuals cannot maintain full concentration at all times?) Third, there are two types of situation that appear to involve uncertainties similar to those over the momentary level of care: situations where parties are responsible for the negligence of subordinates whose behavior they cannot control completely; and situations
where parties operate machines which occasionally function erratically.

(c) The more general interpretation of the fact that uncertainty over the level of due care may induce parties to take socially excessive care is that uncertainty over the law may lead parties to take socially undesirable steps to avoid liability.

§C.3 anticipated errors in the choice of due care. Now suppose that parties know in advance that the level of due care will be different from the socially optimal level, and how so. (But, for simplicity, suppose that courts can correctly measure parties' true levels of care and that parties can control completely their levels of care.)

It might be that parties know the due care level will be less than the optimal level. This would be true, for instance, where parties know they will not be found negligent for failure to use a particular safety device despite its low cost and substantial effectiveness in reducing risk. In such a situation, parties will obviously not purchase the safety device; they will not take more than due care.

The other possibility is that parties know due care will exceed the optimal level (that a safety device will be required despite its high cost and low effectiveness in reducing risk). In this situation, parties will take due care unless its level is so high that they are better off acting negligently. In the latter case, parties will take optimal care since they will, in effect, be strictly liable.
§C.4 misperception of the level of due care. Suppose finally that parties misperceive the level of due care that courts will apply. Then parties will take the level of care that they believe constitutes due care, unless it exceeds optimal care by so much that they are better off acting in a way they think is negligent, in which case they will take optimal care. Hence, if parties overestimate due care, they will either take more then due care, or they will take optimal care; while if they underestimate due care, they will take less than due care.

§C.5 note on the literature. Diamond (1974a) first studied how parties will respond to uncertainty over the finding of negligence, making many of the points discussed here; and Calfee and Craswell (1984) further developed the subject.

D. Miscellaneous comments

§D.1 why parties are found negligent. The reader will recall that in the basic theory of Chapter 2, parties were never found negligent because it was in parties' interests to act with due care and because the negligence determination was perfect. It is evident from the analysis in this chapter, however, that there are a variety of reasons why parties may be found negligent. To recapitulate, it is clear from §A.1 that (i) limitations in courts' capacity to distinguish differences among parties may lead to findings of negligence against parties for whom taking due care is relatively difficult. It is apparent from §A.2 that (ii) the undesirability of the
courts' reducing the level of due care below a certain minimum, even if courts know some parties simply cannot meet it, will lead to findings of negligence against these parties. It is clear from §C.1 that (iii) uncertainty in the negligence determination, and parties' lapses of attention and inability to control their momentary level of care, mean that they may be found negligent; from §C.4 that (iv) if courts set due care sufficiently above the optimal level, then parties will decide against taking due care; and from §C.5 that (v) if parties underestimate the actual level of due care and thus take only what they believe to be due care, they may be found negligent.22

§D.2 significance of occurrence of negligence. There are two points that should be made. First, the occurrence of findings of negligence implies that there is an element of strict liability -- of having to pay for harm done -- associated with use of the negligence rule. Hence, much of what was said in Chapters 2 and 3 about strict liability carries over to a degree to the negligence setting. For example, the fact that under strict liability injurers will take into account the losses their activity creates has relevance under the negligence rule; injurers will take some account of the losses their activity creates because they will face some risk of being found negligent.

Second, the occurrence of findings of negligence means that injurers bear a risk of liability under the negligence rule. This observation will be called upon later, in Chapter 9, to
explain why injurers should wish to purchase liability insurance against being found negligent.

§D.3 strict liability and its comparison with negligence. It is interesting to note that because an injurer's goal will be the social goal under strict liability, injurers will act optimally under that rule in the situations studied in this chapter. Under strict liability, each injurer will exercise his individually optimal level of care, and will engage in an activity only when that is desirable; injurers will take appropriate prior precautions; and, of course, injurers will act without any concern over whether there would have been uncertainty surrounding a negligence determination.

All this constitutes an argument in favor of a strict liability approach over the negligence rule as far as injurers' behavior is concerned. But the problems with injurers' behavior under the negligence rule become problems with victims' behavior under strict liability with the defense of contributory negligence. Uncertainty over the contributory negligence determination, for instance, may lead victims to take excessive care. This complicates the comparison of strict liability and negligence rules.

§D.4 the different versions of the negligence rule reconsidered. In Chapter 2 parties acted optimally under the negligence rule whether or not the rule was accompanied by the defense of contributory negligence, and parties also acted optimally under the comparative negligence rule. When, however, the factors studied in the present chapter are
introduced into the bilateral model -- so that findings of negligence and contributory negligence will occur -- parties will not necessarily act optimally under the different versions of the negligence rule, and the different versions of the rule will have different effects.

As a general matter, to the extent that injurers will be found negligent, victims will have the greatest motive to take due care under the negligence rule with the defense of contributory negligence, a lesser motive to take due care under the comparative negligence rule, and the least motive under the negligence rule without the defense of contributory negligence. This is because, of course, under the negligence rule with the defense of contributory negligence, a victim who does not take due care will receive no compensation from a negligent injurer, whereas under the comparative negligence rule he will receive partial compensation, and under the negligence rule without the defense of contributory negligence, he will receive complete compensation.

On the other hand, to the extent that victims will be found contributarily negligent, the situation is reversed. Injurers will have the greatest motive to take due care under the negligence rule without the defense of contributory negligence, they will have a lesser motive to take due care under the comparative negligence rule, and the least motive under the negligence rule with the defense of contributory negligence.
This suggests that the comparison among the different versions of the negligence rule will depend importantly on the likelihood of victim negligence versus injurer negligence. 23
Chapter Nine

LIABILITY, RISK-BEARING, AND INSURANCE: BASIC THEORY

The discussion of the last chapter will allow account to be taken in the analysis here of the dual nature of the accident problem: that it involves not only the goal of reducing appropriately the risks of accident losses, but also the objective of properly allocating those accident losses that do occur. The chapter will begin by describing the socially ideal solution to the accident problem, and will then consider the problem in the absence and in the presence of liability and of insurance.¹ For simplicity, the analysis will be of the unilateral model of accidents in which all losses are pecuniary; the main points will carry over in obvious ways to the bilateral model, and the case of nonpecuniary losses will be discussed in the next chapter.

A. The socially ideal solution to the accident problem

§A.1 Under the socially ideal² solution to the problem, parties will make decisions about engaging in activities and about their exercise of care in the way that was previously described (in Chapter 2) as optimal. In addition, risk averse parties -- be they victims or injurers -- will not bear risks, which is to say, their risks will be perfectly spread through insurance arrangements or will be shifted to risk neutral parties. Consider

- 311 -
Example 1. If injurers engage in an activity and spend 60 on care, they will reduce the risk of causing an accident loss of 10,000 from 8\% to 6\%; and if they spend an additional 90 on care, they will further reduce the risk to 5\%. Hence, if injurers engage in the activity, it will be optimal for them to spend \(60 + 90 = 150\) on care; and they should therefore engage in the activity if the benefits they would obtain from it exceed 650 (the cost of care plus expected accident losses).

With regard to the allocation of the 5\% risk of losses that will exist whenever it is optimal for injurers to engage in the activity, suppose first that both victims and injurers are risk averse. In this case it will be optimal that neither bear risk. Therefore, victims will be compensated through an insurance arrangement where they would otherwise bear losses, and injurers will be compensated through an insurance arrangement where they would otherwise pay for victims' losses. If victims are risk averse and injurers are risk neutral, then it will be optimal that the victims be compensated through an insurance arrangement where they would otherwise bear losses, but it will also be optimal for injurers to bear the risk of victims' losses; and conversely if injurers are risk averse and victims are risk neutral.//

$A.2$ remarks. (a) It should be stressed that protection of risk averse injurers against risk is just as important a determinant of social welfare as is protection of risk averse victims, other things being equal.
(b) In thinking about the attitudes towards risk of injurers and of victims, the reader may wish to keep in mind the following types of situation: (i) Where large firms may do harm to individuals; here the injurers might often be regarded as risk neutral and the victims as risk averse. (ii) Where individuals may cause losses for large firms (as where an individual driver damages a firm's vehicle); here the injurers might well be regarded as risk averse and the victims as risk neutral. (iii) Where individuals may do harm to other individuals; here both the victims and the injurers might be regarded as risk averse. (iv) Where large firms may cause losses for other large firms; here both the victims and the injurers might frequently be regarded as risk neutral.

(c) In making judgments about the importance of risk aversion, one must consider the size of losses in relation to parties' assets (see §A.2 of Chapter 8), and one must bear in mind that while firms may usually be fairly seen as less risk averse than individuals, there are two important reasons why they should not always be: First, not all firms are large (stores and restaurants are often owned by small numbers of persons; professional services are often provided by single or only a few individuals), and many individuals are well to do. Second, even large firms may cause losses that are high relative to their assets, yet that are the aggregation of only relatively modest losses for each victim. This possibility may arise, for example, where design errors affect a high percentage of the units of a firm's product. In such cases, a
firm's risk aversion could be a more important consideration than victims'.

B. The accident problem in the absence of liability

§B.1 If injurers will not be liable for accident losses, they generally will not reduce risk appropriately. They may engage in risky activities to an excessive extent, and will have no motive to take care. In Example 1, for instance, injurers will engage in their activity as long as it provides them any benefits, and when doing so, they will take no care and thus will cause expected losses of 800. On the other hand, injurers will bear no risk; this aspect of the outcome is socially desirable if injurers are risk averse.

Since victims will not be able to obtain judgments from injurers, they will be left bearing risk if accident insurance is not available. This is socially undesirable if victims are risk averse. But if accident insurance is available, risk averse victims will purchase full coverage, so that the only difference in the outcome from the ideal will be that injurers do not act to reduce risk appropriately.

§B.2 remark. A restatement of the preceding is that where victims can secure accident insurance coverage (or can be supplied with social insurance), the main advantage of use of the liability system is that it provides injurers incentives to reduce risk.
C. The accident problem given liability alone

Assume here that injurers are subject to liability, but that neither liability insurance nor accident insurance is sold.

§C.1 strict liability. Under this rule, victims will, by definition, be compensated for any losses they sustain; it is injurers who will bear risk, and it is their attitude towards risk that is of interest. If injurers are risk neutral, then their bearing of risk will not matter to the optimal allocation of risk. Also, as is known from before, injurers will take optimal care if they engage in an activity and will decide to do that when appropriate. Thus, the outcome will be socially optimal.

But if injurers are risk averse, the outcome will not be socially optimal. Social welfare will be lowered relative to the ideal if injurers engage in an activity not only because injurers will bear risk, but also because they may be led to exercise excessive care to avoid liability (consider how cautiously risk averse and uninsured individuals would drive if subject to strict liability). In addition, for these reasons, injurers may be discouraged from engaging in an activity in the first place even though doing so would be best (individuals might decide against owning and driving automobiles despite that being socially desirable). This is illustrated by Example 2. In the situation of the previous example, suppose that injurers are risk averse, have initial wealth of 30,000, and a graph of utility of wealth as in the examples of
Chapter 8. Then if injurers engage in the activity, their bearing of a risk of losing 10,000 means that the outcome will not be optimal. In addition, they may exercise too much care. For instance, if they can reduce the risk from 5% to 4% by spending an additional 110, they will do so even though that will not be socially worthwhile. Also, they may decide against engaging in the activity when they would engage in it in the socially optimal situation. That will be the case, for example, if the benefits injurers would obtain from the activity are 700. One way of alleviating such problems is to reduce the magnitude of liability, and indeed, it can be shown always to be beneficial to do that to some degree. In other words, if injurers are risk averse, it is not socially desirable to "internalize" fully the harm they do.

§C.2 negligence rule. The situation is quite different under the negligence rule, since, at least in the absence of uncertainty over the negligence determination, injurers will not bear risk provided that they take due care, which they will decide to do. Hence, there will be no particular problems where injurers are risk averse, and they will not be led to take excessive care nor be undesirably discouraged from engaging in an activity (although, of course, there will be the usual problem of their undesirably deciding to engage in an activity).

Victims, on the other hand, will bear their losses, presuming that injurers are not mistakenly found negligent. As
a consequence, social welfare will be lower than optimal if victims are risk averse and are not insured.

§C.3 comparison of rules. It has been seen that under strict liability, risk will be borne by injurers, and under the negligence rule, largely by victims. Therefore, in the absence of insurance, the relative appeal of strict liability will be enhanced where injurers are risk neutral or, more generally, where they are less risk averse than victims; and the relative appeal of the negligence rule will be enhanced where the reverse is true.

It should be observed also that the appeal of liability rules that have the effect of dividing losses between injurers and victims (the comparative negligence rule and strict division of liability) will be increased where both types of party are risk averse.\(^8\)

§C.4 remark. In assessing the relevance of the assumption of this part that risk averse parties are not insured, one must keep in mind that there are a variety of reasons why in fact they may not be insured (or very little insured),\(^9\) and one should also note that the sale of liability insurance may be proscribed.\(^10\)

D. The accident problem given liability and insurance

Now reconsider the situation under liability rules assuming that both accident insurance for victims and liability insurance for injurers are available.\(^11\)
$D.1\;\textbf{strict liability.}$ Suppose initially that liability insurers can determine injurers' levels of care. The outcome will then be socially optimal. The only problems that could arise concern the situation of risk averse injurers, since victims are protected against risk by definition of strict liability. However, risk averse injurers will purchase liability insurance policies supplying them with full coverage. Moreover, the policies will have provisions inducing injurers to take optimal care (see $B.3$ of Chapter 8). Also, since injurers will pay premiums equal to the expected losses they cause, they will decide to engage in an activity only when they should. Thus, the outcome will indeed be optimal, as the following example illustrates.

Example 3. In the situation of Example 1, if injurers are risk averse and decide to engage in their activity, they will obtain full liability insurance coverage and will spend 150 on care. For spending only 60 on care will reduce their premium from 800 to 600, and spending another 90 will reduce it to 500.\textsuperscript{12} As their cost of engaging in the activity will also include the 150 cost of care, injurers will choose to engage in the activity if and only if the benefit they will obtain exceeds 650. Thus, they will behave optimally. Also, as injurers will be insured if they engage in the activity (and victims will be paid for any losses), no risk averse parties will bear risk.\textsuperscript{/}

It should be noticed here that the ability to purchase liability insurance is socially beneficial because it leads to the socially ideal outcome. In $B.1$, when liability insurance
was unavailable, injurers were exposed to risk and there were accompanying problems of their exercising excessive care and possibly of being undesirably discouraged from engaging in their activity.

Next suppose that liability insurers cannot determine injurers' levels of care and thus cannot link the terms of policies to it. In this case, the outcome will be more complicated, and will depart from the socially optimal one. Specifically, while risk averse injurers will purchase insurance policies, the policies will usually involve less than complete coverage. Policies with less than complete coverage will tend to be favored by injurers because these policies will leave injurers with some motive to reduce the risk of liability. This will mean that the premium rate will be lower than if they had full coverage (recall again the argument of §8.3 of Chapter 8). Of course, the fact that injurers are risk averse and will bear part of the risk of losses means immediately that the outcome will not be ideal. In addition, the fact that injurers will be protected from part of the risk of losses suggests that they will generally not take optimal care. In the situation of Example 3, for instance, if injurers have coverage of, say, 7,000 against their possible 10,000 liability, they will be led to spend only 60, rather than the optimal 150 amount, on care.

Yet despite the possibility that the purchase of liability insurance will result in a level of care that differs from the optimal, it should be emphasized that the availability of
liability insurance will still be socially desirable. The reasoning is as follows. The availability of liability insurance cannot affect the welfare of victims under present assumptions. Victims will be compensated for accident losses whether or not injurers have liability insurance. Because of this, it will not matter to victims what the risk of accidents is, or what influence injurers' ownership of liability insurance may have on the risk. Hence, the only way the availability of liability insurance can affect social welfare is by changing the welfare of injurers. But since injurers choose to buy liability insurance, it must be that the insurance makes them better off.

§D.2 negligence rule. Under this rule, assume first that there is no uncertainty over the negligence determination. Then injurers' levels of care will be optimal, as will the bearing of risk by them and by victims. Specifically, this will be the case if injurers take due care, since injurers will then bear no risk and victims will purchase full accident insurance coverage against losses. Thus, what needs to be shown is that injurers will take due care. The usual argument establishing that it is rational for injurers to take due care will apply if injurers do not possess liability insurance protecting them against their negligence. And, as it turns out, we can see on reflection why injurers will not purchase such insurance: the premium for the insurance would be too high to make it worth purchasing. For all injurers who owned the insurance would act negligently, meaning that the insurers'
costs and the premium charged would reflect the level of expected accident losses produced by negligent behavior. Injurers would therefore be better off not buying the insurance and taking due care. 16

Now assume that there is uncertainty over the negligence determination. In this case, the main difference to note is that injurers might be found negligent even if they try to take due care. Thus, injurers will, as is realistic, decide to purchase liability insurance. On the other hand, it is important to say that the type of policy injurers will purchase will protect them primarily against being found negligent through some type of error or lapse. The policy will not protect injurers so broadly as to induce them definitely to act negligently. 17 The reason is essentially that given at the end of the last paragraph. A policy inducing injurers definitely to act negligently would not be purchased because the premium for the policy would be too high. Also, I believe that it can be argued again, along the lines of what was said about the situation under strict liability, that the availability of liability insurance will be socially desirable.

§D.3 summary. Three points about liability and insurance will summarize the analysis of this part.

(a) First, since liability insurers pay for some or all of losses for which injurers are found liable, the manner in which liability rules will alter injurers' behavior is to a significant degree indirect, being associated with the terms of their liability insurance policies (with the link between
premiums or the payment of claims and injurer behavior, and with whether the level of coverage is complete).  

(b) Second, the availability of liability insurance is socially desirable. The particular arguments demonstrating this result depended on the form of liability and insurers' information about insureds' behavior. The arguments were, roughly, based on the following considerations. The availability of liability insurance increases the welfare of risk averse injurers because it protects them from risk and mitigates the problems that they would otherwise take excessive care or be discouraged from engaging in desirable activities. Moreover, the availability of liability insurance does not necessarily dilute injurers' incentives to reduce risk; and where it does do that, the dilution of incentives will be moderate; for policies that would substantially increase risks would be so expensive that they would not be attractive for purchase.

This is hardly to deny, however, the existence of circumstances where liability insurance might be socially undesirable, and some such circumstances will be discussed later. But it is to say that thinking about the issue of the social desirability of liability insurance should proceed from the understanding that in the basic model of liability studied here, the insurance is socially desirable.

(c) Third, the availability of accident and liability insurance limits the importance of the allocation of risk as a factor to be considered in evaluating liability rules. For
example, the fact that in some area of accident the typical injurers might be large, essentially risk neutral firms and the victims, risk averse individuals, will not constitute an argument in favor of imposing liability to the extent that the individuals are insured against their losses.

§D.4 remark on the cheaper insurer. Even though the availability of insurance limits the importance of the allocation of risk in the evaluation of liability rules, it is relevant to consider whether victims or injurers can more cheaply insure because of administrative cost advantages. In this regard, two cautionary points should be borne in mind. First, it is sometimes argued that since a firm need purchase only one liability insurance policy, a firm can more cheaply insure than the many individual victims the firm may harm. But victims ordinarily purchase insurance policies protecting them against generally described risks (such as loss of life due to any cause). Hence, it is not necessary for victims to purchase additional policies protecting them against each new risk they bear; for victims to be insured against additional risks does not mean that additional administrative must be incurred. Second, even if injurers can more cheaply insure than victims, it may still be administratively cheaper for victims to bear their losses. This is because administrative costs must be incurred by victims in obtaining awards or settlements from liable injurers (see §A.2 of Chapter 11).
§D.5 actual use of accident and liability insurance. 21

The ownership or public provision of accident insurance and of liability insurance is today widespread, with the major fraction of victims' losses being compensated by their insurers or by injurers' liability insurers. The only real exception to this is that in the Soviet Union liability insurance is not allowed, 22 because of the fear that otherwise the deterrent of liability would be dulled. 23 Historically, similar worries were expressed by the legal community about liability insurance in Western countries, and in some cases there was considerable resistance to its sale. 24

§D.6 note on the literature. In Shavell (1982a), I demonstrated the results presented here in an analysis of liability and insurance. Related discussions of the subject are contained in James (1948), in ch. 4 and (interspersed) in other chapters of Calabresi (1970), in ch. 10 and secs. 3 and 4 of ch. 23 of Atiyah (1980), in McNeely (1941), and in Williamson, Olson, and Ralston (1967).
Chapter 10
LIABILITY, RISK-BEARING, AND INSURANCE:
EXTENSIONS TO THE BASIC THEORY

This chapter continues the previous analysis by considering insurance against nonpecuniary losses; the possible divergence between awards that are optimal for purposes of compensation and awards that are optimal for purposes of deterrence; the rule determining whether victims' receipts of collateral insurance benefits are to be set off against liability; injurers' inability to pay for harm done; and the structure of a system of pure accident insurance.

A. Nonpecuniary losses and insurance
§A.1 nonpecuniary losses and an individual's need for money. In many cases, suffering a nonpecuniary loss will not alter an individual's need for money, or, more exactly, the utility to him of having additional money. If, for example, an irreplaceable family portrait with great sentimental value but no market value is destroyed, there is no obvious reason to believe that the owner's need for money will increase, however much he regrets the loss. Similarly, if a person loses a small toe in an accident, then, aside from the cost of any required medical treatment, his need for money may well not change.

In some cases, though, events with adverse nonpecuniary consequences will result in a person placing a higher value on money. An individual who is crippled may value money more, even after being compensated for medical expenses and foregone
income, because of a desire to obtain household help, special transportation services, and the like.

It is also possible that suffering a nonpecuniary loss will lower the value of money to an individual. The individual who is crippled could turn out to value money less because venturing out to spend is difficult for him. Perhaps the most important example of valuing money less due to a nonpecuniary loss involves death. The value of money in that contingency is, in effect, its value as a bequest, and this will often be less than the value of money to a person if he is alive.

§A.2 optimal insurance coverage for nonpecuniary losses.

The amount of insurance coverage against nonpecuniary losses that an individual will wish to purchase will clearly depend on the need for money that such losses would create.

If nonpecuniary losses will not result in an added need for money, then under the expected utility maximimizing insurance policy, a person will not arrange for coverage against the nonpecuniary losses; coverage will be restricted to pecuniary losses, if any. Thus, a person might not insure against the loss of his family portrait, and might limit coverage against loss of a toe to medical expenses. The following numerical example demonstrates the point.

Example 1. Suppose that a person’s utility is the sum of utility he derives from money, as specified in the graph in Figure 7 of Chapter 8, and of utility derived from nonpecuniary elements. Suppose also that his initial wealth is 30,000 and that he faces a 10% risk of an accident that would involve
a pecuniary loss of 10,000 and a nonpecuniary loss of 25 units of utility.

Then calculations will show that the person will be best off purchasing coverage against only the pecuniary loss of 10,000 -- even though he can purchase higher coverage sufficient to compensate him in utility terms for the nonpecuniary component of loss. If the person arranges for coverage of only 10,000, his premium will be 1,000, and his wealth will be 29,000 regardless of the occurrence of an accident. Hence, his utility will be 994.49 if an accident does not occur and 994.49 - 25 if an accident does occur. Therefore, his expected utility will be 90% \times 994.49 + 10\% \times 969.49 = 991.99.

On the other hand, suppose the person purchases 15,270 unit coverage. This is the amount necessary to compensate him for the 10,000 pecuniary loss plus the 25 unit nonpecuniary loss. In particular, if he buys the 15,270 coverage, his premium will be 1,527, his wealth if there is no accident will therefore be 28,473, and his utility from wealth will thus be 991.36. His wealth if there is an accident will be 33,743 (that is, 28,473 - 10,000 + 15,270), and his utility from wealth will be 1016.36. Hence, the utility from having an extra 5,270 in insurance coverage will be 1016.36 - 991.36 = 25, which will compensate him exactly for his nonpecuniary loss of 25. But if the person purchases 15,270 coverage, his expected utility will be 991.36, so he will indeed be worse off than if he purchases just the 10,000 coverage against his pecuniary loss.
Notice that since it will be optimal to insure only against pecuniary losses when a nonpecuniary loss will not alter a person's need for money, optimal insurance coverage will not make a person "whole", that is, just as well off as if he does not suffer any losses.

If the value a person will place on money will increase as a result of a nonpecuniary loss, optimal insurance coverage will exceed his pecuniary loss. Thus, a person may purchase greater coverage against the possibility of being crippled than an amount equal only to the costs of medical treatment and foregone earnings. It is unlikely, however, that he will purchase coverage sufficient to make him whole (if this is even possible).

If the value a person will place on money will decrease as a result of a nonpecuniary loss, expected utility maximizing insurance coverage will be less than his pecuniary losses. A person who has little desire to leave a bequest will rationally purchase little or no life insurance, despite the possibility that the earnings foregone by his death will be large.

In the discussion so far, it has implicitly been assumed that the risk of losses is fixed. If the risk is influenced by the behavior of insureds, then, as was discussed in Chapter 8, ownership of insurance may dull insureds' incentives to take care. But this problem is reduced in importance where the risk includes nonpecuniary elements. Since it is generally not in an individual's interest to purchase enough coverage against a nonpecuniary loss to make himself whole, he will be made worse
off by the occurrence of a nonpecuniary loss even after having received his insurance coverage. Hence, he will have a reason to lower risk. A person who has purchased optimal coverage against loss of a toe or of his life will still have a very strong motive to avoid injury.

§A.3 actual coverage against nonpecuniary losses. Insurance coverage against loss of property does not ordinarily seem to reflect its sentimental value, only its market value or replacement cost. Coverage against personal injury usually approximates only direct medical expenses and foregone earnings. Insurance against death is ordinarily bounded by lost earnings; if a person (such as an unmarried or elderly individual) has no dependents, he normally possesses little or no coverage; parents do not often carry significant coverage on the life of their children. Evidently, insurance coverage is intended mainly to remedy pecuniary needs created by losses, not to compensate for the disutility due to losses.

§A.4 note on the literature. Arrow (1974), Zeckhauser (1973), and Cook and Graham (1977) first developed the theory of insurance for nonpecuniary losses.

B. Divergence between awards that are optimal for purposes of compensation and awards that are optimal for purposes of deterrence

§B.1 The general point to be made here is that the size of the award that is best for a victim to receive, in view of his need for money and the theory of insurance, may be less than
the amount that is best for an injurer to pay, in order that his incentives to reduce risk be appropriate. The discussion will begin with, and stress, the case where the possibility of nonpecuniary losses causes such a divergence between awards optimal for compensation and awards optimal for deterrence, but other reasons for a divergence shall also be mentioned.

§B.2 nonpecuniary losses and the socially ideal solution to the accident problem. As emphasized in the last part, the amount of insurance an individual will purchase against nonpecuniary losses, and therefore the amount of money he will receive under the socially ideal solution to the accident problem, will be based on the value he will place on money if he suffers nonpecuniary losses; the amount he will receive will not generally make him whole.² By contrast, the socially optimal level of care taken by injurers (and their level of activity) will reflect the nonpecuniary elements of accident losses as well as the pecuniary (recall the discussion in Part D of Chapter 6). Thus, for instance, it will be best that injurers take substantial care to reduce the risk of accidentally killing children even though the death of children may not mean that their parents will need much more money and consequently, in the ideal, may not call for the parents to receive significant amounts.

§B.3 socially ideal solution cannot be achieved under the liability system when there is a divergence between awards optimal for compensation and awards optimal for deterrence. Since under the ideal solution to the accident problem,
injuries will take a degree of care reflecting both nonpecuniary and pecuniary components of victims' losses, the magnitude of payments injurers make under liability rules must reflect both these components of losses for them to be led to take optimal care. But if injurers' payments are this high, then the amount victims receive will exceed optimal compensation, since that will usually approximate their pecuniary losses. On the other hand, if injurers' payments equal only optimal compensation, injurers' incentives to take care will be inadequate. Thus, the socially ideal outcome cannot be achieved under the liability system. The magnitude of liability will inevitably result in a compromise between awarding victims correctly and creating appropriate incentives for injurers to reduce risk. 3

§8.4 reasons other than nonpecuniary losses for a divergence between awards optimal for compensation and awards optimal for deterrence. It was noted before that where liable injurers will not always be identified as responsible for harm done, the amount they pay if they are identified and sued must be raised for their incentives to reduce risk to be maintained at the correct level. 4 Where, for instance, the likelihood is 50% that a liable injurer will be successfully sued, the amount he pays if so sued must be on the order of twice the victim's losses. Optimal payments by injurers may therefore exceed optimal compensatory awards by a substantial factor.

It was also explained before that where a party obtains socially illicit benefits from an act, or incurs socially
illicit utility costs when exercising care, it will be desirable for him to pay an amount greater than the harm sustained by the victim. Thus, again, optimal payments by injurers may exceed optimal compensatory awards.

Another reason why optimal payments by injurers may be greater than optimal awards concerns taxes that would have been paid on income foregone by accident victims. It was argued previously that for injurers' incentives to reduce risk to be proper, they must pay an amount based on before-tax income foregone by victims. Yet the amount of money that victims will in fact lose, and thus the amount that will constitute the optimal compensatory award, is after-tax foregone income.

An additional reason why optimal payments by injurers may exceed optimal compensatory awards concerns victims' receipts of insurance benefits or of gifts. In this case, optimal compensatory awards will equal only the shortfall between victims' receipts and their losses, but injurers must pay victims' entire losses to be adequately deterred.

§8.5 the case for fines as a supplement to liability. An improvement over the situation under the liability system can be achieved under a regime in which liability is supplemented by fines collected by the state. With the use of fines, the total amount that injurers are made to pay can be raised to the point that their incentives to reduce risk are appropriate, while at the same time liability can be held to the lower level equal to optimal compensatory awards. Thus, under the contemplated regime, injurers would pay fines reflecting
nonpecuniary losses, the probability of escaping suit, taxes on victims' foregone earnings, and so forth, while victims would usually receive in liability awards payments reflecting only their otherwise uncompensated pecuniary losses.  

§B.6 remarks on supplemental fines. (a) It would ordinarily be best for the fines used as a supplement to liability to be insurable. The general argument made in Chapter 9 that it is desirable to allow risk averse injurers to purchase coverage against liability can be employed to demonstrate this result.  

(b) The use of supplemental fines would tend to resolve certain tensions arising from application of standard principles of tort law to calculate liability. It is frequently observed, for instance, that if damages for wrongful death are based on the present value of foregone earnings, payments for the death of children or the elderly may be negligible, an uncomfortable result. But although liability awards for the death of children or the elderly may properly be low, optimal supplemental fines could well be significant.  

The receipt by tort victims of punitive damages or of large awards for pain and suffering presents different issues. One sometimes encounters uneasiness over the "windfall" character of punitive damage awards and over victims' incentives to exaggerate pain and suffering. Such incentives would be diminished, and there would be no windfalls -- no inappropriate allocation of risk in the terminology of economics -- if pain and suffering and punitive damage awards were incorporated into fines paid to the state.
(c) To calculate optimal fines for nonpecuniary losses, it can be shown that one can use as a guide extrapolations from the following: the amount optimally insured individuals would be willing to pay for a small reduction in the probability of suffering nonpecuniary losses. Suppose, for instance, that an individual who has the insurance he desires against the medical expenses and foregone income that would result from losing his left arm would be willing to pay $1,000 for a 1% reduction in the likelihood of that event. Then the optimal fine would be approximately $100x$1,000 or $100,000. Information about willingness to pay for reductions in risk could in theory be obtained by survey, or in some cases, perhaps, by attributing wage differences to differences in risks of accidents.  

§8.7 Remark on liability of firms to customers. To the extent that firms will be liable for (or would have to pay a supplemental fine for) nonpecuniary losses caused by their products or services, and that customers do not wish to insure against these losses, customers will be undesirably discouraged from making purchases, other things equal. This is because prices will reflect firms' expected liability. Consider, for instance, the price of sending a child to camp, supposing that camps will be liable for large losses if children die in accidents. Then the price of camp will include a component to cover camps' large liability insurance premiums. Thus, in paying for camp, parents will in effect be purchasing life insurance on children, a form of insurance that they may well not wish to carry. It is possible that this forced purchase of
life insurance will discourage parents from sending children to
camp, when it would be desirable that the children go to camp.
This, then, is an argument disfavoring imposition of
(especially strict) liability on firms when damages include
nonpecuniary elements against which customers do not wish to
insure.\textsuperscript{13}

\textsection{B.8 note on the literature.} Spence (1977) originally
demonstrated the desirability of employing fines in addition to
liability in a study of firms' liability for accidents causing
nonpecuniary losses.\textsuperscript{14}

\textbf{C. Victims' collateral insurance benefits and injurers' liability}

\textsection{C.1 When a victim is insured against losses (here assumed
to be only pecuniary) for which an injurer is liable, two legal
rules may apply: A rule under which a victim is allowed to
collect both his collateral insurance benefits\textsuperscript{15} as well as a
judgment from a liable injurer -- so that the victim may
receive in total an amount exceeding his losses; and a rule
under which a victim can collect in a judgment only the
difference, if any, between his losses and his insurance
benefits. The first rule will be called the "no subtraction
rule" and the second, the "subtraction rule." These rules were
discussed briefly before (in \textsection{F.7 of Chapter 6). They will be
reconsidered here, along with related issues, in light of the
theory of insurance.
§C.2 subrogation. Before examining the two rules, the insurance arrangement known as "subrogation" must be defined. Under this arrangement, an insurer will pay a victim for his losses, but the insurer rather than the victim will then have the right to sue and collect from a liable injurer. 16

§C.3 socially optimal outcome regarding victims' receipts and injurers' liability. Subrogation and the rules concerning subtraction of collateral benefits will be evaluated below with reference to the socially optimal outcome regarding victims' receipts and injurers' liability. This optimal outcome will have two characteristics. First, a liable injurer will pay an amount equal to a victim's losses, for that is necessary to maintain injurers' incentives to reduce risk. Second, a risk averse victim will receive from his insurer and a liable injurer together an amount equal to -- but not more than -- his losses, for that is an aspect of optimal risk bearing. 17

Notice that the outcome if victims' insurance policies involve subrogation arrangements will have both these characteristics, and therefore will be optimal.

§C.4 outcome under the no subtraction rule. Under this rule, victims will prefer to purchase insurance policies involving subrogation arrangements. Hence, the outcome under the rule will be optimal.

There are two reasons why victims will in principle prefer to purchase policies involving subrogation arrangements. The first is in essence the point from §B.2 of Chapter 8 that risk averse parties will generally prefer purchasing policies giving
them full coverage to purchasing, at a greater cost, policies giving them more than full coverage. Here the interpretation of the point is that victims will generally prefer purchasing policies where insurers have subrogation rights to purchasing for higher premiums policies where insurers do not have these rights and where, under the no subtraction rule, victims will collect both from their insurers and from liable injurers. (Of course, premiums will be higher for policies without subrogation rights because insurers will not be able to collect judgments from injurers.) This is illustrated by

Example 3. A victim with initial wealth of 30,000 and utility of wealth as in Chapter 8 faces a 10% risk of suffering losses of 10,000 for which no injurer will be liable, and a 10% risk of suffering the same losses for which an injurer will be liable. Suppose that the victim purchases an insurance policy giving him full coverage and allowing the insurer subrogation rights. The victim will then receive exactly 10,000 if he suffers losses, whether or not an injurer is liable, and the insurer will bear an expense of 10,000 only when there is no liable injurer. Hence, the premium for the policy will be only 10%×10,000 = 1,000, and since the victim's wealth will be 29,000 for sure, his utility will be 994.49.

If the victim purchases a policy under which the insurer does not have subrogation rights, the victim will receive 10,000 if he suffers losses and there is no liable injurer; the victim will receive 10,000 plus a judgment of 10,000 if an injurer is liable; and as the insurer will bear an expense of
10,000 whenever the victim suffers a loss, the premium for the policy will be 2,000. Thus, the victim’s wealth will be 28,000 with probability 90% and 38,000 with probability 10%. His expected utility will therefore be 90% \times 988.40 + 10\% \times 1,028.85 = 992.44, which means that he will be worse off under the policy without subrogation and will not purchase it.\(^{18}\)

The second reason why, in theory, insurance policies will be expected to involve subrogation rights is that because victims can, under the no subtraction rule, otherwise collect both insurance benefits and judgments, they will otherwise profit from suffering losses. This could set up undesirable incentives for victims actually to cause losses or to fabricate them. This, in turn, would result in high premiums and consequently in policies that would be unattractive for purchase.

§C.5 outcome under the subtraction rule. Under the subtraction rule, the outcome will be identical to that under the no subtraction rule. Victims will again purchase policies involving subrogation arrangements, and the outcome will thus be optimal. The reason that victims will prefer to purchase policies giving insurers subrogation rights is plain. If a policy does not include subrogation rights, then a victim will be compensated by his insurer and will not also be able to collect from a liable injurer because of the subtraction rule. If a policy does include subrogation rights, the victim will as before collect only from his insurer, but his insurer will be able to collect from a liable injurer;\(^{18}\) this will reduce the
insurer's costs and will lead to a lower premium. Hence, victims will prefer policies with subrogation arrangements. Consider

**Example 4.** If the victim of the last example purchases a policy with full coverage under which the insurer has subrogation rights, then as noted, the victim's premium will be 1,000 and his wealth will be 29,000 for sure. If, however, the victim purchases a policy without subrogation rights, his premium will be 2,000, and his wealth will be only 28,000 for sure. Therefore, he will decide to purchase the policy with subrogation rights.//

$\S$C.6 **remark on departures from theoretically predicted outcomes.** It has just been argued that in theory victims will decide to purchase policies giving insurers subrogation rights, and thus that under the no subtraction rule, victims will not collect both insurance benefits and judgments from injurers. Yet this may not be true in fact, for there may be legal constraints against subrogation; or victims may be covered by social insurance plans allowing them both plan benefits and recovery from liable injurers; or victims may possess coverage from multiple insurers, and owing to difficulties of coordination among the insurers, victims may simply find that they are able to collect both insurance benefits and judgments.19

$\S$C.7 **comparison of the no subtraction and subtraction of benefit rules.** To the extent that, for the reasons just mentioned, victims may collect more than their losses under the
no subtraction rule and that injurers' liability may be reduced under the subtraction rule, it will matter which rule is chosen. The no subtraction rule will be the better rule if it is more important to maintain injurers' incentives to reduce risk than to avoid overcompensation of victims.\textsuperscript{20}

\textit{\textsection C.8 remark on constraints against subrogation.} These seem to be against the interests of insureds, since as has been emphasized, insureds should prefer to give insurers subrogation rights and to pay lower premiums. A failure to take into account this beneficial effect of subrogation on premiums (or disbelief in the effect) is apparently what lies behind the commonly heard argument that subrogation represents a detriment to insureds.\textsuperscript{21}

\textit{\textsection C.9 remark on repayment arrangements.} A type of insurance arrangement resembling subrogation but under which victims retain their right to sue injurers should be mentioned. According to this arrangement, a victim must repay his insurer for insurance coverage received if he sues and collects from an injurer.\textsuperscript{22} Thus, if the victim sues and collects, the eventual result will be the same as under subrogation; the victim will obtain only his losses, and the insurer will break even.

Under a repayment arrangement, a victim will, however, have little motive to bring suit, since he will have to return to his insurer most or all of his award.\textsuperscript{23} Therefore, suits will usually not be brought, meaning that insurers' costs, and premiums charged, will be higher than under policies with
subrogation arrangements. As a consequence, it is not surprising that policies with subrogation arrangements greatly predominate over policies with repayment arrangements.²⁴

§C.10 actual relationship between victims' insurance benefits and injurers' liability, and the nature of victims' insurance policies. The United States is one of few countries in which the general rule is that a victim's insurance benefits are not subtracted from a liable injurer's judgment.²⁵ Nevertheless, the situation turns out to be similar in the United States and in other countries in certain important areas of insurance. This is because there are exceptions made to the general rules that result in a like pattern of use, or of denial of use, of subrogation. For example, in regard to property insurance, subrogation rights are nearly universal.²⁶ In the areas of life and personal injury insurance, subrogation is prohibited, or at least not encouraged, and victims' insurance benefits are not subtracted from judgments.²⁷ In respect to workers' compensation, the usual outcome, though reached in various ways, is that victims collect only once and that their insurers may seek reimbursement from liable parties.²⁸

In other areas of insurance, however, the situation differs substantially among countries. This is the case, for example, with regard to private medical insurance and social security.²⁹
D. Injurers' inability to pay for losses and liability insurance

§D.1 This part first discusses the effect of injurers' inability to pay fully for harm done on their decisions to purchase liability insurance and on their incentives to reduce risk. Then the part examines regulation of liability insurance.

§D.2 diminished motive to purchase liability insurance. If injurers' assets are lower than the harm they may cause, a portion of the liability insurance premium they will pay for full coverage will in fact pay for losses that they would not otherwise bear. It follows that risk averse injurers may rationally decide against purchasing full coverage, and may decide to buy none. This is illustrated by

Example 2. Suppose the risk averse injurer described in Chapter 8 has assets of 30,000 and faces a 20% risk of being found liable for 100,000. If he does not purchase coverage, he will enjoy his 30,000 with probability 80% and lose his 30,000 (for that is all he can possibly pay) with probability 20%. Hence, his expected utility will be \(0.8 \times \$1,000 + 0.2 \times \$0 = \$800\).

If, on the other hand, the injurer purchases full coverage of 100,000, his premium will be \(0.2 \times \$100,000 = \$20,000\) (of which 14,000 will pay for the 70,000 of losses that he would not bear if he does not insure). His wealth will therefore be 10,000 with certainty. Therefore, his utility will be 665.24. In other words, he will be worse off with full coverage than with none at all. Indeed, it can be shown he will be worse off with any positive amount of coverage than with none at all.\(^{30}\)//
While in this example, the injurer is best off not buying any coverage, injurers may sometimes decide to purchase positive, though less than complete coverage, and in particular, to purchase policies with ceilings on coverage.31

§D.3 remark. The likelihood that parties will not purchase much or any coverage will depend on the magnitude of their assets in relation to the harm they might do and on their degree of risk aversion. For example, an individual with assets of $20,000 might wish to buy liability coverage on the order of $200,000 if he were fairly risk averse, but an almost risk neutral electric utility company with assets of $500 million might well decide not to purchase $1 billion coverage against the possibility of a catastrophic accident at its nuclear power plant.

§D.4 functioning of the liability and insurance system. It was emphasized before, in the risk neutral setting without liability insurance, that injurers' inability to pay for harm done dilutes their incentives to take care and leads to too high a level of risky activity.32 This point remains largely unchanged where risk averse injurers choose not to purchase liability insurance.

If, however, injurers do decide to purchase insurance coverage, the point must be qualified. Injurers' purchase of coverage will result in their bearing more of the expected losses with which their activity is associated, and thus will tend to reduce their levels of activity desirably. If, for instance, injurers buy full coverage and are strictly liable,
they will make correct decisions about engagement in risky activity. If the electric utility were to purchase full coverage, it would not build or operate the nuclear facility unless the benefits outweighed the full expected losses.

On the other hand, whether injurers' incentives to take care will be altered for the better by their purchase of liability coverage depends on the ability of insurers to determine levels of care and to link the premium (or other policy terms) to it. When an insurer can do this (when it can tell if the utility has installed needed safety devices in the nuclear plant) and injurers purchase complete coverage, then they will be induced to take adequate care. But when insurers cannot determine levels of care (how the utility selects and trains its staff, whether pipes in the plant have been properly welded), then injurers' care will be further diluted by their having purchased coverage. In other words, the initial problem of injurers' exercising too little care, because of limited assets, will be exacerbated.

Finally, injurers' incentives aside, it should be observed that victims will rationally purchase insurance coverage against the risk that liable injurers will be unable to pay for the losses they cause.33

§D.5 regulation of insurance. It has been seen that injurers' inability to pay fully for losses creates two types of problems: their incentives to take care and to engage in risky activity may be incorrect; and because they may choose not to purchase adequate liability coverage, they may bear risk.
Consider first the effect on these problems of a requirement to purchase complete liability insurance coverage. This will eliminate the problem of injurers' bearing risk, and it will also alleviate the problem of their excessive engagement in risky activity. However, if liability insurers are unable to determine injurers' levels of care and injurers possess full coverage, they will exercise no care at all. Hence, it is not clear that a requirement to purchase coverage will be socially desirable.

Now consider a prohibition against purchase of liability insurance. This will increase the problem of the bearing of risk by injurers, for their entire assets will be at stake. Yet just because of that, barring coverage may reduce injurers' levels of activity from an initially excessive level. If such a reduction is moderate, it could be beneficial; but it could also be disadvantageously large. Likewise, prohibiting liability insurance may turn out to have either a desirable or an undesirable effect on injurers' exercise of care. Prohibiting liability insurance will tend to increase levels of care in the case where the purchase of coverage would dull incentives, that is, where insurers would not be able to observe levels of care. But prohibiting insurance will reduce levels of care where insurers would be able to observe care.

The conclusion is that whether requiring purchase of liability insurance, prohibiting it, or not regulating it will be best will depend on the situation. Where insurers can observe levels of care, requiring purchase of coverage will be
the superior policy; where insurers cannot do this, requiring coverage may be dominated by not regulating coverage or, conceivably, by prohibiting coverage.

§D.6 remark. The problems considered here did not include what is often viewed as the main problem with injurers' inability to pay. Namely, victims may not be compensated for their losses. This did not arise as a problem in the analysis because, as noted in §D.4, risk averse victims were assumed to decide to purchase insurance coverage against not being compensated by liable injurers. In reality, of course, victims may not purchase such coverage for a variety of reasons. Hence, in fact, requiring purchase of liability insurance would help to solve victims' problem; but so would requiring that victims purchase their own coverage or establishing a public insurance fund for victims.

§D.7 note on the literature. Calabresi (1970), p. 58, observed that there will be a diminished motive to purchase insurance where parties' assets are less than the losses for which they may be held responsible; Keeton and Kwerel (1984) and Huberman, Mayers, and Smith (1983) first investigated the point formally; and in Shavell (1986), I studied the implications of the point for the functioning of the liability system and the regulation of liability insurance.

E. Structure of a system of pure accident insurance

§E.1 This part briefly considers the nature of a system in which risk averse victims insure against accident losses but do
not enjoy the right to sue injurers. (As will be explained in subsequent chapters, such a pure accident insurance system may be justified even though the liability system, having been eliminated, will no longer provide injurers' incentives to avoid doing harm.)\textsuperscript{34} The main question to be addressed concerns the connection between the cause of a victim's loss and his coverage under an optimal insurance policy. This question will be discussed first assuming that victims do not influence the risk of suffering losses, and then that they do.

\textit{\$E.2 optimal coverage where victims cannot influence risks.} In this case, under an expected utility maximizing insurance policy, a victim's coverage will depend only on the losses he suffers, not on how they come about. A victim will thus receive the same amount of coverage if his house burns due to lightning striking it or if it burns due to a fire that his neighbor negligently allows to spread.

That only the magnitude of a loss, and not its cause, should influence insurance coverage may seem obvious to the reader. (The reader probably says to himself that the only thing that matters to a victim is his loss, so the victim should have no reason to arrange for different coverage on account of the loss arising from one or another cause.) But the point does seem worth justifying. The reader will recall that in \textit{\$B.2} of Chapter 8, it was demonstrated that where victims cannot influence risk, expected utility maximizing insurance coverage will equal the loss suffered. It follows that if a loss of some amount, say, 1,000, can come about either through
cause A or cause B, optimal coverage will be the same, namely, 1,000, whether the cause of the loss is A or B. In other words, and as claimed, optimal insurance coverage will not depend on the cause of losses.35

§E.3 remark. There are affirmative advantages to policies under which coverage does not depend on the cause of losses. Under such policies, insurers need not expend resources in investigating the cause of losses; hence, the cost of supplying insurance will be lower. Furthermore, victims themselves need not make efforts to establish the cause of losses.

§E.4 optimal coverage where victims can influence risks. Here it may be the case that under an expected utility maximizing insurance policy, coverage will depend on the cause of losses. Suppose, for example, that a victim's house can burn not only due to a fire caused by lightning or to his neighbor's negligence, but also due to the spontaneous combustion of oily rags -- a risk that the victim can lower by removal of the rags. Then it may be optimal for a victim to receive only partial coverage if a fire is caused by oily rags, in order to give victims an incentive to remove them. (See §B.3 of Chapter 8.) By contrast, it will not be optimal for a victim to receive partial coverage if a fire is caused by lightning, assuming that making victims bear some of their losses in that event will not create any beneficial incentives.

The general point illustrated by this example is that it may be optimal for coverage to depend on the cause of losses if
victims have greater influence on accident risks that come about through certain causes than through other causes.

§8.5 remarks. (a) The causes of losses that may in principle be relevant to optimal insurance coverage range widely. For example, it may be relevant whether the cause of a person's lung cancer is his smoking habit or a carcinogen the exposure to which was beyond his control, since making coverage lower in the former instance may induce individuals not to smoke. It may even be relevant whether the cause of an accident was an injurer's fault (as determined by an accident insurer), for where injurers fail to take care, it may be that victims are unable to avoid harm.\textsuperscript{36}

The question is sometimes asked whether it is desirable for distinctions in (especially social) insurance coverage to be made between losses caused by disease versus those caused by accidents.\textsuperscript{37} The answer is that it may be optimal for a distinction to be made if the distinction pertains to a victim's ability to reduce risk. To illustrate, suppose that a person can lose an arm either due to a disease that he cannot prevent or in a type of accident that he can do something to avoid. Then optimal coverage may well be less for losing an arm due to an accident than due to the disease.\textsuperscript{38}

(b) The degree to which it will be desirable for insurance coverage to depend on the cause of losses will depend not only on the incentives thereby created, but also on the feasibility and cost of determining causation. Thus, under optimal insurance policies (private or public), coverage will have to depend on broad categories of fairly easily ascertained causes.
Chapter Eleven
LIABILITY AND ADMINISTRATIVE COSTS

By "administrative costs" are meant the various expenses borne by parties in resolving the disputes, or the potential disputes, that arise when harm occurs. Administrative costs thus include the time and effort spent by injurers, victims, and their legal counsel and insurers in coming to settlements and in litigation, as well as the publicly borne operating expenses of the courts. This chapter will first discuss several factors determining the magnitude of administrative costs. Then it will contrast victims' incentives to make use of the liability system with the social interest in their doing so, given that that now involves administrative costs.

A. **Factors determining administrative costs.**

§A.1 total administrative costs. When a victim is harmed by an injurer, the victim will decide for or against making a claim. If a victim does make a claim, he will either settle with the injurer or he will go to trial. Total administrative costs will equal the number of claims settled multiplied by the administrative cost per settled claim, plus the number of litigated claims multiplied by the cost per litigated claim. Equivalently, total administrative costs will equal the total number of claims multiplied by the average cost per claim, the average being calculated over both settled and litigated claims.
§A.2 remarks. (a) A more detailed description of the
determination of administrative costs would, of course,
distinguish among claims settled at different stages of the
litigation process (after this or that amount of negotiation,
after the filing of certain motions, after discovery), and
would allow for the possibility that a victim will abandon his
claim. Moreover, it would take into account the various costs
incurred by insurers.

(b) Because the costs of coming to settlement may be
important and because the frequency of settlement is high
(apparently over 90%), the proportion of total
administrative costs associated with settlement may be
significant in many areas of litigation. It would be a mistake
to attribute administrative costs mainly to the costs of trials.

(c) The data that exist suggest that in the United States
the administrative costs of the liability system are
substantial. Most studies find that administrative costs,
averaged over settled and litigated claims, exceed the amounts
received by victims. It is not clear, however, to what
extent these administrative costs should be viewed as intrinsic
to the liability system or as a feature of the particular
system that has developed.

(d) By contrast, the administrative costs of accident
insurance are often on the order of from 1% to 15% of what
victims receive. The reason that the administrative costs
associated with provision of accident insurance are relatively
low is that accident insurers have much less need than courts
to inquire into the cause of losses or about injurers' behavior, that accident insurers have adopted comparatively simple procedures for verifying the magnitude of insureds' claims, and that accident insurers are not in an adversarial relationship with insureds. 5

§A.3 administrative costs under strict liability versus under the negligence rule. Administrative costs under strict liability may be usefully compared to those under the negligence rule by considering, on the one hand, the total number of claims, and, on the other, the average administrative costs per claim.

The total number of claims is likely to be larger under strict liability than under the negligence rule. Under strict liability, a victim will have an incentive to make a claim whenever his losses exceed the costs of making a claim. 6 Under the negligence rule, a victim will not have an incentive to make a claim so often, because he will also be concerned about establishing the injurer's negligence. If a victim and an injurer both believe that a court will find the injurer free of fault, the victim will be unlikely to make a claim under the negligence rule.

Although the volume of claims should therefore be greater under strict liability, the average administrative cost per claim should tend to be higher under the negligence rule. Under the negligence rule, it is more probable that a claim will be litigated than under strict liability, for under the negligence rule there is an additional element of dispute --
that of the injurer's negligence -- and hence more room for disagreement leading to trial. Since the probability of trial should be greater under the negligence rule and since trials will usually be more costly than settlements, we have one reason for saying that average administrative costs per claim are likely to be larger under the negligence rule. A second reason is that the costs of trial are likely to be higher under the negligence rule than under strict liability because the issue of negligence must be adjudicated under the former rule.

In sum, then, the comparison of the size of administrative costs under the two forms of liability is ambiguous as a theoretical matter. While one would predict that a greater number of claims will be made under strict liability, one would expect the average cost of resolving claims to be higher under the negligence rule because of both a higher propensity to go to trial and a higher cost per trial.

§A.4 remark. Consideration of the role of insurers appears to diminish the importance of the factors determining differences between administrative costs under the two liability rules. First, the significance of the point that under strict liability there will be more claims made against injurers than under the negligence rule is lessened when one takes into account claims made by victims against their accident insurers. That is, when victims do not make claims against injurers under the negligence rule, victims will generally seek to collect from their accident insurers, with accompanying administrative costs. Second, the significance of
the point that under strict liability, there will be no need for a legal determination of injurers' negligence is lessened when one takes into account the behavior of liability insurers. Liability insurers may well wish to ascertain injurers' negligence, or something like it, so as to provide injurers with incentives to reduce risks (see §B.3 of Chapter 8).

B. Social versus private incentives to make use of the liability system given its costs

§B.1 in general. Suppose that administrative costs are included in the measure of social welfare. Then it will be socially desirable for victims to make use of the liability system, that is, to make claims against injurers, if and only if the associated social benefits will outweigh the administrative costs. Yet whether victims will decide to make claims depends on the quite different "private" comparison they make between their own benefits and costs. Specifically, a victim's private benefit from making a claim will be the expected settlement or judgment he will receive. As shall be seen, this private benefit may be higher or lower than the social benefit. A victim's own cost of making a claim will be lower than the administrative costs due to his claim, if it is assumed that the victim will not bear the injurer's costs or the costs to courts themselves. While this difference suggests that victims may have too great an incentive to make claims, it can well turn out that victims have too little incentive to
make claims, since their private benefits from making claims may be smaller than the social benefits.

§B.2 a simple model. To gain an appreciation of the circumstances under which the number of claims will be socially excessive or will be undesirably low, and to see what policies might remedy these problems, a simple version of the unilateral model of accidents will be examined. In this model, parties will be assumed risk neutral, victims will bear a cost if they make a claim, injurers will bear a cost if claims are made against them, and if injurers are found liable, they will pay victims an amount equal to their losses.\footnote{7} The social goal will be to minimize total costs, defined as expected accident losses, plus costs of care, plus costs due to the making of claims. The situations under strict liability and under the negligence rule will be examined separately, since there are differences between them relevant to the issues of present interest.

§B.3 private versus social incentives to make claims under strict liability. A victim's private incentive to make a claim under strict liability is clear in the above model: a victim will make a claim if and only if his cost of doing so will be less than the losses he suffered. It will be socially desirable that victims make claims, however, if and only if their costs of making claims plus injurers' costs exceed the social benefits inhering in the incentives to reduce risk that will result from claims. The next example illustrates the possibility that the difference between private and social
incentives will lead victims to make claims when it would be best that they did not.

Example 1. Suppose that the losses victims will suffer in accidents are 1,000; that the probability of accidents will be 4% if injurers take care, at a cost of 8, and 5% if they do not take care; and that a victim's cost of making a claim will be 300 and an injurer's cost of defending, 200.

Victims will therefore make claims whenever accidents occur, since making a claim will cost a victim only 300, whereas he will obtain 1,000 in damages. It follows that injurers will be induced to exercise care: this will cost an injurer 8 yet will reduce his expected liability and defense costs by a greater amount, namely, by \(1\% \times (1,000 + 200) = 12\). As, the probability of accidents will be 4%, total costs will be \(8 + 4\% \times (1,000 + 300 + 200) = 68\).

From the social perspective, this outcome is undesirable. If victims did not make claims, then, although it is true that injurers would have no incentive to take care and the accident risk would consequently be 5%, total costs would be only \(5\% \times 1,000 = 50.//\)

The reason that the making of claims was not socially desirable in the example was that the incentives towards safety that were created by the claims did not reduce accident losses plus the costs of care by as much as the administrative costs due to the claims. This fact was of no moment to victims, however. They had no reason to take into account that the incentives created
by claims were small; they were concerned only with the judgments they would obtain. 8

The following example illustrates the opposite possibility, that victims may not make claims even though it would be desirable that they do.

Example 2. The losses victims will suffer in accidents are 100, and an expenditure of 1 by injurers will reduce the probability of accidents from 10\% to 1\%. The costs of making and defending against claims are as in the previous example.

In this case, victims will not make claims, since doing so will cost a victim 300, but yield him a judgment of only 100. Hence, injurers will have no reason to take care to reduce risk, and total costs will be 10\%x100 = 10.

On the other hand, it would be desirable for victims to make claims. For if they do, injurers will be led to lower risk to 1\%. Therefore, total costs would be only 1 + 1\%x(100 + 300 + 200) = 7.//

Here, the making of claims would be socially worthwhile because of the significant reduction in accident losses that would result. (And observe that this is true despite the fact that the social resources that would be expended in making a claim, namely, 500, exceed a victim's losses of 100.) But no victim took the social benefits of making claims into account. Each victim looked only to his own gain from making a claim, which was small.
§B.4 corrective policy. If a social authority has information sufficient to know whether and how the social and private incentives to make claims differ, the authority can remedy the situation. The authority can simply ban suit if victims would make claims when that would be undesirable. Alternatively, the authority can raise the cost to victims of making claims sufficiently to discourage victims from doing this; or, in some cases, the authority could insist that victims pay the total administrative costs due to making claims. 

Conversely, the authority can subsidize the making of claims if victims would otherwise fail to make claims when it would be desirable that they did, or the authority can shift victims' costs to injurers (use the "British" system) to encourage claims.

While a social authority can therefore act to improve social welfare, it is obvious that the authority's appropriate choice of policy depends on whether the problem faced is one of too few claims or too many. There is no policy that will correct automatically for both types of problem. Shifting costs to liable injurers, for instance, will hardly be helpful if the problem is that victims make too many claims.

§B.5 private versus social incentives to make claims under the negligence rule. Assume initially that if injurers take due care, they will not be found liable by mistake and victims know this. Then it will be socially desirable for victims to be willing to make claims against negligent injurers however great the administrative costs would be. The reason is that if
victims stand ready always to make claims against negligent injurers, injurers will be induced never to act negligently. Thus, there will never turn out to be any claims made, and no administrative costs will be incurred. 12 Although it will be socially desirable for victims always to be willing to make claims against negligent injurers, victims will not do so if the cost of making claims will exceed their losses. Therefore, a policy to encourage the making of claims, such as a subsidy or the shifting of victims' costs to negligent injurers, may be desirable.

Now assume, more realistically, that because of various errors and misperceptions, victims may make claims even if injurers take due care. Victims may make claims because, for example, they erroneously believe that non-negligent injurers were negligent, or because they correctly believe that the courts would erroneously find non-negligent injurers negligent, and so forth. For such reasons, administrative costs will in fact be incurred under the negligence rule. Hence, the situation is qualitatively similar to that under strict liability; there may be too many claims as well as too few, calling for corrective policy either to discourage or too encourage the making of claims.

§B.6 Remarks. (a) Not only does the presence of administrative costs mean that social versus private incentives to use the liability system becomes an issue, it also introduces a new consideration into the determination of optimal levels of care: Accidents are socially more expensive
if they involve administrative costs in addition to victims' direct losses. Thus, optimal levels of care should be higher on account of administrative costs.\textsuperscript{13}

(b) It should be clear from the initial discussion and from the logic of the arguments made in the analysis of the model, that the main conclusion -- that private incentives to make use of the legal system may lead to either too many or too few claims -- will apply in a general setting, incorporating the complexities of the litigation process, and also social benefits of litigation in addition to its incentive effects, such as the creation of precedent and compensation of the uninsured.

(c) Differences between private and social incentives to make use of the liability system may well be significant, and may constitute a serious reason for social intervention.

On one hand, because the administrative costs of the liability system seem generally to be large (as noted in §A.2), the incentives towards safety created by use of the liability system must be substantial to warrant its use. Yet these incentives may not always be very strong, especially when account is taken of factors other than liability that already operate to reduce risk. In the area of automobile accidents, for instance, criminal liability, enforcement of safety regulations, and the fear of being injured in accidents already provide incentives towards accident avoidance. Where the added incentives created by the liability system are not sufficient to justify its use, a point of this part is that it may still
be very much in the private interest of victims who have sustained large losses to bring suit. Hence, some sort of social (presumably legislative) intervention will be required to reduce use of the liability system. Indeed, various attempts to limit litigation (passage of no fault statutes, shifting of legal fees to losing plaintiffs) can perhaps be seen as reflecting perceived problems of excessive private motives to bring suit.

On the other hand, there undoubtedly exist areas in which victims of harm have too little incentive to bring suit. This may be true where the cost of bringing suit will typically exceed the magnitude of losses, and where injurers, realizing that they will not be sued, fail to behave appropriately even though it would be cheap for them to do that. What, for instance, would be the incentive of a taxi cab company to have its drivers avoid doing small harms (scratching a parked car when pulling up to discharge a passenger), or the incentive of a restaurant not worried about its reputation (say it is located on a superhighway and its customers will usually visit it only one time) to avoid causing minor cases of food poisoning? Where the private incentives to sue are low and injurers take advantage of this, social intervention to encourage suit may be desirable. And again, certain social efforts to affect litigation, here to promote it (through establishment of small claims courts, availability of the class action), might be interpreted in light of the present theory.
§B.7 note on the literature. In Shavell (1982b), I examined the contrast discussed here between the social and private incentives to use the legal system in view of its costs; and Menell (1983) and Kaplow (1986) further analyzed the issue. See also the related work of Ordover (1978) and Polinsky and Rubinfeld (1986).
Chapter Thirteen

CRITICAL COMMENTS ON THE ANALYSIS

The present chapter, which is written in question and answer form, will make a variety of comments about the analysis of the book and will attempt to respond to certain commonly heard criticisms of economic analysis of law.¹

A. Predictive and normative analysis

§A.1 Of what value is the analysis of the book for predicting and understanding behavior? It seems self-evident that the analysis should be of value for these purposes, as it seeks carefully to determine the decisions that calculating actors will make given the rules of liability and opportunities to insure. How much value the analysis will have will depend on whether the assumptions studied capture important elements of reality, on the degree to which the analysis helps to organize thought about the effects of liability and the insurance system, and on the extent to which the analysis identifies effects that the reader does not consider obvious. With regard to the latter, the reader should ask himself, for example, whether he had recognized at the outset that under error-free application of the negligence rule, the calculating actor will necessarily, and not just sometimes, be led to take due care; whether he had appreciated that uncertainty surrounding the negligence determination may lead systematically to the exercise of excessive, defensive, precautionary measures; whether he was well aware of the manner
in which the terms of insurance policies are influenced by the ability of insurers to obtain information about insureds; whether he had focused on the point that individuals are unlikely to insure against many nonpecuniary losses; or whether he had realized that the collateral source rule is unlikely to have an effect much different from the opposite rule. On reflection, I believe many readers will agree that their understanding of the working of the liability and insurance systems has been enhanced by the analysis presented here.

§A.2 Does the prospect of liability really deter? Was undue importance attached to this whole issue in the analysis? It is, of course, an empirical question how much the threat of liability affects behavior, and given the relative lack of statistical study of the question, one must rely mainly on intuition.² One may surmise, first of all, that the possibility of liability may significantly affect the actions of business enterprises in view of their habits of calculation and the magnitude of the harms they may cause. One must not discount, moreover, the influence of potential liability on individuals. Even if, as I suspect, some readers believe that the number of times an individual's potential liability will impinge on his behavior is small, this does not mean that the effect of the liability system on the accident rate will be negligible. (While the decision of a homeowner to fence in a backyard swimming pool for fear of liability may be made only once in his lifetime, the fence may significantly reduce the risk of a drowning.) In addition, although it is often said
that the ownership of liability insurance undoes the deterrent effect of liability, this view is oversimple. As I have emphasized, insurance policies contain many features that create incentives towards safety (the homeowner may fence his pool to obtain a reduction in his liability insurance premium). It thus seems that there is ample reason for theoretical study of the effect of liability on behavior -- though I doubt many would seriously have argued against this in the first place.3

§A3 What is the normative value of the analysis? The analysis should be of aid in assessing the desirability of legal rules, presuming as I do that the criterion of social welfare in which the reader is interested reflects the value to parties of engaging in their activities, the costs of taking precautions, the losses due to accidents, compensation of the risk averse parties, and administrative costs. How much help the analysis will offer in answering normative questions will depend, as was true in respect to predictive questions, on the degree to which the analysis leads to conclusions that were not fairly clear in the first place. In considering this, the reader should again pause to review some of the conclusions that were reached. For example, he should recall the point concerning the advantage of strict liability over the negligence rule in controlling injurers' levels of activity; the argument showing that the sale of liability insurance is socially beneficial; the discussion of the possible gains to be had from accompanying liability by fines where achieving
optimal deterrence requires that injurers pay more than it is
optimal for victims to receive; and the examination of the
social versus the private incentives to make use of the
liability system given its costs. I trust that most readers
will not have found all of these and the other conclusions of
the book familiar, and that the analysis will have helped to
clarify thinking.

§A.4 What consistency is there between observed and
theoretically optimal liability law? One can point to many
instances of approximate agreement between liability law as
observed and theoretically optimal liability law, that is, the
law that appears best given the measures of social welfare that
were examined and certain guesses about the actual effects of
liability on behavior. I suggested, to mention an important
example, that the observed use of strict liability in areas
where risks are apparently high is consistent with the
theoretically optimal use of strict liability. Indeed, I
indicated throughout the first seven chapters of the book how
desirable incentives are created by this or that feature of
tort law. (Recall, for instance, the discussions of the
reasonable man and of prior precautions, the analysis of
causation, and the treatment of the subject of damages.)

On the other hand, one can readily adduce examples of
apparent disagreement between observed and theoretically
optimal law. Supplemental fines, for instance, are generally
not used despite the strong advantages they offer. One can
also find instances where the theoretically optimal law is not
clear. To illustrate, it was shown in the basic analysis of Chapter 2 that use of the defenses of contributory negligence and of comparative negligence will result in the same, and optimal, behavior; no persuasive theoretical arguments were offered indicating the desirability of one defense over the other. One can find as well instances where the law differs among countries in ways that cannot plausibly be explained on the basis of the factors studied here. In this regard the defenses of contributory and comparative negligence again afford an example: It is not apparent why, on the theory in this book, the first defense should sometimes be employed in the United States, but not in England, France, Germany, or the Soviet Union. Likewise, it is not easy to see reasons why, on the theory in the book, the liability of drivers towards pedestrians should be governed in this country by the negligence rule whereas usually by a strict theory elsewhere.

Thus, not only does there seem to be considerable consistency, but there also seems to be substantial ambiguity and inconsistency between the liability system that we observe and the regime that is best given the criteria of optimality and the models examined here.  

§A.5 Where there is consistency between observed and theoretically optimal liability law, can it be said that observed law is explained? There are plainly a variety of factors going beyond the ones studied in this book that shape the law -- the power of interest groups, the particular histories of legal institutions, the opinions of influential
jurists and commentators, widely held notions of fairness, and so forth. The mere fact of consistency between observed law and what is here identified as theoretically optimal, therefore, hardly means that observed law is explained. For example, in the brief discussion in Chapter 2 of justifications actually given for use of strict liability, I noted that little mention is made of incentives; the use of strict liability cannot be primarily attributed in a direct way to a perception that otherwise the number of accidents would grow too large. Nevertheless, contemplating the matter further, it is more than possible that if we interrogated people on the question why they thought strict liability fair and desirable, we would be able to elicit from them statements indicating the importance of generating incentives to reduce participation in risky activity. It is also likely, in my opinion, that the incentives created by strict liability would be recognized explicitly, and perhaps given prominence (in legislative hearings, in the press, if not in judicial opinion), if, under the negligence rule, the number of accidents were to become high in some area of activity.

In any event, there are many examples of rules that are clearly understood to include among their major justifications those discussed in this book. For instance, the rule holding the award an injurer must pay to the losses a victim will suffer if he takes reasonable steps to mitigate damages was observed to be desirable in the analysis because it gives victims incentives to limit the magnitude of their losses.
This is also recognized by commentators to be an important reason for the rule.6

The conclusion, therefore, is that the interpretation of consistency between theoretically optimal and observed legal rules will depend very much on the rule in question. Sometimes consistency will signify very little or reflect mainly an implicit connection with the theoretical considerations studied here. Other times consistency will reflect a close and explicit connection.

§A.6 Of what importance is the omission from the analysis of income distributonal considerations? This omission appears to be of small importance from the normative perspective. On the one hand, there already exists a social institution with an overtly redistributive function, namely, the income tax system. It reaches virtually all individuals and is, by comparison to the liability system, administratively inexpensive.

On the other hand, there are serious problems with the liability system as a device for accomplishing the redistribution of income. First, it does not reach all individuals, only those who are involved in litigation. Second, legal rules and decisions are likely to affect in a uniform way groups that are quite dissimilar in their need for money or ability to pay awards. The group of victims of automobile accidents, for instance, includes a wide spectrum of individuals by income type, as does the group of negligent injurers. To take another example, a corporate defendant is
really comprised of its stockholders, and that group -- who may enjoy ownership through pension funds or life insurers' investments -- will usually include a diversity of individuals, many of whom will not be well off. Third, liability awards are unlikely to be coordinated with redistribution carried out by other means, such as rent control and food stamp programs. Yet awards would have to be linked to litigants' participation in these programs for redistribution to be appropriate. In all, the liability system appears to be an ineffective device for redistribution of income. Consequently, had I introduced income redistributational considerations into the analysis, and had I posited a socially desirable income tax structure, I would have concluded, in the main, that redistributational considerations should not influence the choice of liability rules.8

From the descriptive perspective, the omission of income redistributational considerations seems somewhat more important, but I do not believe that significant redistribution of income is taking place through the tort system. While there may be a tendency today towards generosity in deciding cases and in making awards where the accident victims are poor and the defendants are well to do or are firms, it is also true that these defendants will usually have engaged superior legal counsel. Moreover, firms may be able to raise prices to cover liability costs, dulling the redistributive effect of their liability (such as it is; recall the remark about stock ownership). Finally, since the principles of tort law do not
allow liability to be based openly on the wealth positions of litigants, the opportunities for redistribution are limited.

§A.7 Of what normative significance is the omission from the analysis of consideration of principles of fairness? One must be cautious in assessing the normative significance of this omission for two reasons. First, principles of fairness may to an important degree encapsulate the goals of social welfare maximization studied here, mainly because adherence to the principles often leads to behavior that reduces risk. An obvious example is that adherence to the principle that it is unfair to hold parties liable unless they acted negligently will lead parties to act non negligently. To determine the independent importance of a principle of fairness, one must therefore be careful not to count the consequences for behavior of adherence to the principle.9

Second, in evaluating the importance of adherence to a principle of fairness, one must take into account the fact that a liability insurer, not the injurer himself, will often be paying the award decided by a court, that the recipient of the award may not be the victim but rather his subrogated insurer, and that litigants may be faceless corporations rather than individuals with moral selves. These factors may well alter, and attenuate, the significance of adherence to notions of fairness for many readers.

In any event, readers can modify the conclusions reached here in light of the values they do attach to principles of fairness.10
B. Purpose and future of tort liability

$B.1$ What is the purpose of tort liability? The answer to this question depends, in the first place, on the interpretation given to the word "purpose." Suppose that by the purpose of tort liability, we mean what most participants in the legal system (or some wider class) say the purpose is. Then I would have to admit that the purpose of tort liability prominently includes the compensation of victims. It is my impression that the great majority legal scholars, lawyers, and judges would state that compensation of victims is an important purpose of the tort system.¹¹ (And this is not altogether surprising, since the issue of immediate concern in a legal dispute is generally whether the victim will be compensated by the injurer, not whether deterrence or other broad social goals will be furthered.)

An alternative interpretation of "purpose" goes to the difference that the presence of the liability system makes to actual outcomes. According to this functional definition, which I now adopt, compensation of victims cannot be said to be an important purpose of the liability system, since in its absence, victims would probably be about as well compensated as they now are (certainly they could be). Compensation would be accomplished by private and social accident insurance.¹² The main difference the presence of the liability system does make, therefore, is that it creates incentives towards safety. This, then, must be said to be the chief purpose of the liability system today.¹³
Historically, however, it is worth noting that the purposes of tort liability were different. Before the emergence of insurance markets, the tort system furnished victims a source of compensation that often would not otherwise have been forthcoming. Tort law thus served to an important degree the dual purposes of compensation and deterrence. Moreover, before there was a differentiated criminal and tort law, a significant purpose of the making of money payments for harm was the maintenance of social order. Without the system of money payments, private vengeance would often have followed the doing of harm. Evidently, the purposes of tort liability have changed over the years.

§B.2 What should be the purpose of tort liability? The principle justification for use of tort liability today should not be compensation of victims, because this can be accomplished with our well developed and comparatively cheaply operating insurance system. If the goal is only to compensate victims, there will be relatively little need, the reader will remember, to obtain the information the tort system requires: the cause of a victim's losses, the identity of the injurer, and facts about the injurer's behavior. (Of course, this is exactly why the insurance system is administratively less costly than the tort liability system.)

Hence, if tort liability is to be employed in some area of accident, the major justification should be to provide incentives towards safety.
Whether tort liability can be so justified in an area of accident is a question that will often merit careful consideration, in view of the opportunity to employ safety regulation and other approaches for controlling risk, and in view of the administrative cost of the liability system. This question will often merit careful consideration especially because, as I stressed, the private incentives to use the liability system may be very different from the social. Victims may quite rationally decide to bring suits even where the resulting incentives to reduce risk are small and do not warrant society's bearing the administrative costs that are generated.

§B.3 What is the future of tort liability? While in certain areas of accident, regimes of insurance, perhaps coupled with enhanced safety regulation, may turn out to replace the tort system,¹⁶ I believe that there will remain an important role for tort liability. Considerable scope probably exists for reducing the administrative costs of the tort liability system through its simplification. Moreover, in the discussion of alternative methods for the control of risk, we saw that the tort system possesses unique advantages where a regulatory or other social authority will not be expected to have good information about risk or the occurrence of harm, and where the deterrent inherent in liability will not be seriously weakened by injurers' inability to pay for harm or the possibility that they will escape suit. It is likely that
these conditions will hold, and continue to make tort liability socially valuable, in a large domain.