An Economic Analysis
of Consumer Redress Mechanisms

Alan A. Shapiro

En français: Analyse économique des moyens de recours
                       des consommateurs

Disponible au: Service des Communications
                       Consommation et Corporations Canada
                       Ottawa, Ont.
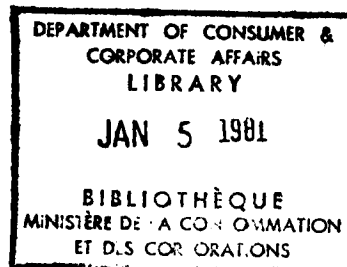   K1A 0C9

AN ECONOMIC ANALYSIS OF

CONSUMER REDRESS MECHANISMS

by

Alan A. Shapiro

January 1979
Revised March 1980

for

Consumer Research and Evaluation Branch
Consumer Bureau
Consumer and Corporate Affairs Canada

FOREWORD

In other legal, economic and behavioural literature, it
has been established analytically that consumers do not use
existing dispute resolution mechanisms and do not pursue many
complaints, in part because there are high transactions costs and
risks of additional losses which serve as disincentives to
pursuing claims even where the claim is well founded in law. The
conviction that this hypothesis is correct is sufficiently
widespread that much has been done to attempt to streamline the
court system and procedures and the citizen's access to justice.
The National Survey of Consumer Satisfaction, Dissatisfaction and
Complaining Behaviour undertaken by the Consumer Research and
Evaluation Branch provided evidence that only 1.07 per cent of
highly dissatisfied consumers turned to a lawyer for advice or
took other legally oriented action. Although this saves society
a large part of the potential cost of running a dispute
resolution system, it also suggests that the removal of
disincentives to using the court system, even if very successful,
might not lead to a high utilization rate of the law and public
dispute processing systems. Hence, more work must be done to
provide incentives for suppliers to give dissatisfied consumers a
satisfactory resolution to their complaint, among these
incentives being the provision to consumers of reasonable
recourse at law.

It is clear that access to the law concerning product
liability and many other matters ought to be improved. However,
in view of the persistent findings on the low propensity of
citizens to use the formal civil process, it is clear that
efforts to improve this process must merely be one of a number of
parallel efforts in related fields.

In the wealth of critical literature concerned with the
dispute processing system, although statistics of an economic
nature are presented, there is often lacking an analytic
framework for the critique. For this reason, we invited
Professor Shapiro, who has served as a Professor of Economics at
York University for many years and is a recent graduate of the
Osgoode Hall Law School, to provide such a framework, in
particular for application with respect to consumer redress.

Jonathan J. Guss,
Editor.

January 1980.

CONTENTS

INTRODUCTION

Do economists have anything unique to say about legal institutions and processes? The answer, it is submitted, is: "Yes, but ...". The open-ended qualification is intended to act as a warning to the effect that the economist can only deal with institutions when their goals are specified, and even then, only if the goals explicitly or implicitly have economic content.

In this paper our concern is with civil dispute resolution mechanisms, the circumstances under which they will be employed by disputants, particularly consumers, and the circumstances under which they will not be used either because the parties have satisfactorily settled the dispute themselves or because the plaintiff has abandoned his or her claim. In particular we shall be interested in the design characteristics of the resolution mechanism which contribute either to settlement or abandonment of the claim.

Our main concern however is not with whether consumer disputes are effectively resolved, or whether "justice" is done to and for the consumer. The economist is primarily concerned with resource allocation in society, and legal institutions only take on meaning to the economist insofar as they either contribute to or interfere with "efficient" resource use. In Chapter I we elaborate upon the concept of economic efficiency and the role of economists in dealing with questions involving societal resource use.

As suggested, to assess institutions, we must first determine their goals. In Chapter II we consider two theories of the civil process which, though not mutually exclusive, do stress different purposes or goals for legal institutions. It will be submitted there that only within the context of one of these theories can legal institutions and processes have economic relevance and hence be susceptible to economic analysis.

Chapter III analyses the relationship between producer decision making, particularly with respect to the allocation of resources to the elimination of defective products, and the availability of redress mechanisms for recipients of defective products. The alternative legal rules of producer/seller liability and consumer liability are compared with respect to their effects on efficient resource allocation.

Finally, in Chapter IV, we focus upon the actual conflict between two disputants, with particular emphasis on the consumer - producer/seller dispute, in order to determine which factors in the dispute and which characteristics of the dispute resolution mechanism shape the final outcome.

CHAPTER I

BASIC PRINCIPLES

1. Efficiency

        Economic analysis in general is concerned with the
concept of efficiency, and the economic analysis of legal
processes and institutions is no exception. It is important
however to distinguish at the outset between economic efficiency
(the concern of economists) and technical efficiency (the concern
of engineers, mechanics, time-work analysts, doctors, etc.).[1]
Technical efficiency is concerned with the physical relationship
between inputs and output. An automobile, for example, may be
said to be operating in a technically efficient manner when it is
attaining the maximum mileage per gallon of gasoline, all other
inputs being held constant (oil, tire wear, etc.). Similarly a
furnace which is producing the maximum heat output per fuel
gallon is operating in a technically efficient manner. A process
of production for a commodity may be said to be technically
efficient if, for a given bundle of inputs (capital, labour,
etc.), no higher flow of commodity output may be reached. Given
an acceptable definition of the output of a legal institution,
the same concept of technical efficiency may be applied. In
general, to assess technical efficiency, one must know the goal
(mileage, heat, output of widgets, "justice", etc.), and the
production function: the ways in which the inputs may be
combined to produce various outputs. Economists normally assume
the most technically efficient processes of production are
employed and are hence not overly interested in technical
efficiency per se.   That is to say, if the same quantity of
widgets[2] can be produced using either one machine and five
workers, or one machine and four workers, then the economist
simply assumes the most technically efficient process, one
machine and four workers, will be selected.

        Economic efficiency deals with "value" and its
maximization. There may be many technically efficient methods of
using capital and labour to produce one hundred widgets: one
machine and ten workers, two machines and seven workers, three
machines and five workers, etc. Once we are faced with given
prices (costs) for machines and workers, however, there will in
general be only one economically efficient method of producing

---

1. For a brief discussion of economic and technical efficiency,
    see R.G. Lipsey, G.R. Sparks, and P.O. Steiner, Economics
    (2nd ed.) (New York: Harper and Row, 1976), p. 187.

2. A widget is a universal commodity frequently encountered in
    economic treatises and lectures. It is believed to be a
    female gadget.

one hundred widgets, and that will be the least cost combination of capital and labour for the given output. Similarly, for the consumer, there may be many combinations of commodities which will yield a given state of satisfaction (well being, utility, etc.), but given the prices of the commodities, there will in general be one combination which is the least costly, and that will represent the economically efficient choice for generating that level of satisfaction. In both cases (the production of widgets and the production of consumer satisfaction) the selection of the economically efficient option maximizes value in the sense both of producing a given output at least cost and of producing the maximum output at a given cost.

Voluntary informed exchange between parties will be value enhancing, and hence economically efficient for the parties involved.[3] Clearly if one party considers it personally advantageous to part with one widget to obtain two gadgets, and the other considers it personally advantageous to part with two gadgets to obtain one widget, then by definition the carrying out of the exchange must leave both parties better off than prior to the exchange. Thus, if the parties are free to exchange or not to exchange and in fact choose to exchange, then the exchange must be value enhancing and hence efficient. The exchange will continue to be efficient if only one party is made better off (by the party's criteria) while the other party is indifferent (i.e., made neither better off nor worse off).

If the exchange is coerced rather than voluntary then there is no clear criterion by which the value enhancement potential of the exchange can be assessed: the parties could as well be worse off after the exchange than before it, or one party could be better off while the other party is left worse off. It is submitted however that if coercion was necessary to realize the exchange, then at least one party considered the exchange disadvantageous.[4]

---

3. If the parties themselves bear the full costs and benefits of the exchange, the exchange rate will be economically efficient for society as well.

4. This assumes that full costs and benefits of the exchange are borne solely by the parties involved. If there are substantial benefits which fall outside the arena of the exchange, however, a coerced transaction may be the only way of avoiding the "free rider" problem.

If the character of the exchange is fraudently misrepresented by one of the parties, then that exchange, though voluntary, will not be value enhancing. This statement must be viewed with caution, however. If a consumer knows that one in ten sales representations made to him will be fraudulent, and still at some price is willing to voluntarily enter that class of exchanges, then that class of exchanges is efficient, though not to the same extent as in the absence of misrepresentation.

A similar analysis is applicable to defective products received in exchanges. Again, if the consumer willingly at a price enters into a series of exchanges knowing that a certain proportion of the goods to be received will be defective, then the exchanges as a group must be efficient. Defects must be contrasted with misrepresentations, however, with respect to their optimal number. The optimal number of misrepresentations is zero. Misrepresentations involve a use of resources (investigation, insurance, prosecution, etc.) as well as diminishing the efficiency of the exchange, and above all, can be eliminated at zero cost in the sense of not requiring the use of economic resources for their elimination. Defective products on the other hand can only be reduced in number through the use of economic resources. There must come some point at which it is no longer economically efficient to expend these resources to obtain further reductions in the flow of defects.

Among those influences on the individual which lead him to evaluate a potential exchange as personally advantageous or disadvantageous are his wealth endowment and income flow.[5] For the same individual a given exchange could be transformed in terms of advantage by a change in wealth distribution between that individual and the community. Thus, unequivocal statements with respect to what is economically efficient for either an individual or the community are wholly dependent upon a given wealth/income distribution. It is necessary to keep this statement constantly in mind in order to avoid drifting into normative expressions based upon unstated assumptions with respect to the status quo wealth distribution.

In closing this discussion of economic efficiency, we may observe why economists are "against" theft. First, of course, theft leads to the use of resources by potential victims in their attempts to provide protection. This resource use is a dead weight loss, being socially unproductive. More important, in a

---

5. In theory, income may be seen as flowing from wealth, both human and non-human, and one need only deal in terms of wealth endowments when discussing questions of distribution.

theft situation there is no guarantee that the exchange will be value enhancing. Assume a person owns a watch which he values at $100, meaning that he would part with it for $100 or more. Assume a thief values that same watch at $150. There are two ways in which the watch may flow to the highest value user, through voluntary exchange for a sum between $100 and $150 and through theft. In terms of resources flowing to their most valuable use, each method of effecting the exchange leads to the same efficient result. The fact that the theft transfer leaves the victim worse off does not deny the fact that the watch <u>did</u> move to the party that valued it most, and that <u>in theory</u> the thief could compensate the victim for his loss (thus of course transforming the nature of the exchange) and leave both thief and victim better off. The problem of course is that we have no social or individual guide as to whether theft does transfer resources to parties who value them more than their previous owners. The economic rationale for punishment then is to deter potential thieves away from theft and entice them to make their valuations known through voluntary exchange. Only then will there be a guide as to whether the proposed exchange is value enhancing and hence efficient, or not.

2.   The Pareto Criterion

There are many different types of economic policies possible and many possible configurations of the economy as a result. Economists have generated a great deal of heat but only a little light in discussing the circumstances under which a policy or a move to another configuration can be said to lead to an improvement in the social (i.e., community) welfare.[6] Most criteria for assessing changes have foundered on the rock of interpersonal comparisons. Simply illustrated, it is not possible to state without recourse to an exterior value system that a change in the economy which leaves 99 per cent of its members better off and one percent of them worse off is a change for the better, i.e., that the social welfare has been improved.

The first suggested criterion (and for many economists still the only acceptable one) was the Pareto criterion,[7] which asserted that any change leading to at least one person

_____

6.   For an excellent critical discussion of the criteria for social welfare judgments, see W.J. Baumol, <u>Economic Theory and Operations Analysis</u> (2nd ed.) (Englewood Cliffs, N.J.: Prentice-Hall, 1965), pp. 375-380.

7.   Stated by the Italian economist and sociologist Vilfredo Pareto (1848-1923).

being better off and none worse off is an improvement in the social welfare. Similarly any change leaving some people worse off and none better off is a deterioration in the social welfare. Nothing may be said of a change which leaves some better off and some worse off.

Most observers would find this criterion, particularly its first two parts, to be acceptable but relatively weak in its usefulness since most policy changes do in fact leave some better off and some worse off. Such policy proposals cannot be judged by the Pareto criterion, largely because the criterion explicitly avoids the most critical issue in policy evaluation: inter-personal comparisons. In this writer's opinion the main use of the criterion has been in the theoretical demonstration that for given initial wealth endowments, a perfectly competitive economy, in the absence of externalities, interdependent utility functions, and increasing returns to scale, will yield an equilibrium solution with respect to allocation of commodities among consumers, allocation of resources among producers, and the composition of output, which is Pareto optimal, i.e., which cannot be improved upon while meeting the Pareto criterion. Putting it mildly, this may offer little guidance in the analysis of current economic and legal issues.

Some economists[8] have attempted to improve upon the Pareto criterion by suggesting that a change in the economy may be considered an improvement if those who gain are capable of evaluating their gains at a higher monetary value than the monetary value assigned to their losses by the losers, such that the gainers are potentially able to compensate the losers for their losses and still be better off than before the change. Actual compensation, however, as in our earlier example of the theft of the watch, need not take place. This modification of the criterion, though somewhat more useful, is still based upon the status quo wealth distribution, relying as it does on individuals' wealth related valuations of their gains and losses. Since most economic policies affect the community's wealth distribution, the modified criterion is of little more use than its parent.

A final adaptation suggests that a change which benefits some and harms others may be considered a social improvement not only if the winners are potentially able to compensate the losers, but if the compensation actually takes place as well. The reader is left to enumerate the social instances upon which such compensation has taken place.

---

8. Kaldor and Hicks, for example.

3. What Can Economists Do?

If the Pareto and Pareto-related criteria are likely to be of little use in real world policy analysis, what then can an economist do or say when faced with a socially perceived problem, i.e., a social situation which some group or individual (legislators, consumers, citizens generally, economists, clients, etc.) finds locally (i.e., for them) undesirable? The economist in such a situation can usually only trace consequences, leaving final assessment as to the change in social welfare to those entrusted with the responsibility for making social decisions. He may say, for example, that if policy option X_1 is exercised, the consequences will be Y_1; if X_2 is exercised, the consequences will be Y_2; etc. He may also recommend the elimination of one or more of the X_1 policy options if they are Pareto-inferior in terms of their results to some of the other options. On occasion it may be possible to recommend a single policy to the decision maker, but it will still require a political decision by the latter as to whether or not the change following from the policy represents an improvement in social welfare.

THEORIES OF THE CIVIL PROCESS

1. Theories Generally

It is impossible to observe any phenomenon, social or physical, without a theory - a conceptual framework within which one may classify stimuli. The number of stimuli emanating from any single phenomenon is infinite and only within a conceptual framework can one select and reject among them so as to construct a meaningful observation. Theories in this sense are purely functional. They may help us to explain a given phenomenon a great deal, somewhat, a little, or hardly at all. They cannot, however, ever be said to be "correct". The most that can be said about a theory is that it is the most useful among competing theories.

Different theories purporting to explain the same phenomenon may coexist. Thus a psychological theory of criminal behaviour may explain some aspects of that behaviour, while an economic or sociological theory may explain other aspects. Which theory is selected will depend upon which aspect of the phenomenon is of primary concern to the observer, and the observer is not obliged to function with a single theory. It should not be surprising, however, if psychologists employ psychological theories to explain the same phenomena for which economists employ economic theories.

What conceptual frameworks are available for looking at the civil process? In a social setting just what do the civil courts really do or accomplish? Certainly, conflict situations which come before the courts are judicially resolved while those that do not come before the courts, either voluntarily or by explicit denial of the mechanism to the parties, are resolved extra-judicially. But this is description, not analysis. In the larger social setting we may ask what courts accomplish by doing what they do. Two approaches have been suggested:[1] the conflict resolution - compensation model, and the behaviour modification - control model.

---

1. See K.E. Scott, "Two Models of the Civil Process", Stanford Law Review (27:1975), p. 937.

2. The Conflict Resolution - Compensation Model

The conflict resolution model sees the civil process as a desirable alternative to self-help remedies and retaliatory violence.[2] Society benefits through security and a reduction in resource use by the channeling of disputes through the court system and this positive externality justifies the provision of subsidized court services. The emphasis throughout in the model is upon the plaintiff - the extent to which he has been harmed and the compensation due him. As violence is considered more likely the greater the harm done to the individual plaintiff, the model would predict a bias in favour of readier access to the court system for larger claims. If the injured party is unwilling to launch an action, he is clearly unlikely to seek redress outside the court system and hence, the model suggests, there need be little concern with him. Similarly, a court system operating in the context of such a model would not look favourably upon third parties encouraging litigation or upon the assignment of claims, or for that matter upon the maintenance of actions on behalf of injured parties, consumers for example, by an administrative agency. The concern in the model is with the peaceful resolution of conflict and if there is no real manifestation of conflict, then there is little concern in the context of the model.

3. The Behaviour Modification Model[3]

Whereas the former model focussed on the plaintiff, the behaviour modification model is largely concerned with the social activity of the defendant, and the concern is not so much with the dispute at hand, but with the future behaviour of the defendant and other potential defendants. The civil process is thus seen as deterring or altering particular behaviour by imposing costs upon it. Compensation and pacification of the plaintiff are not of direct concern after the action is launched.

---

2. See A.M. Linden, "Faulty No Fault: A Critique of the Ontario Law Reform Commission Report on Motor Vehicle Accident Compensation", Osgoode Hall Law Journal (13:1975), p. 439.

3. See K.E. Scott, "Two Models of the Civil Process", Stanford Law Journal (27:1975), p. 938, note 11.

The model fits well with the economic view that legal rules with respect to liability can be explained as contributing to economic efficiency.[4] If an activity imposes costs which are borne by others - externalities - then the actor should be made to carry those costs. Thus one who through negligence injures another will be called upon to pay damages sufficient (in theory) to make his victim whole. The awareness of the full costs of negligence will provide an incentive for the actor to exercise precautions to avoid injuring others. Not every possible precaution will be taken however, only those that are economically efficient in the sense of reducing expected liability for damages by more than the cost of exercising the precaution. The social cost of negligence, being the sum of the cost of preventing accidents and the damages arising from those accidents that are not "worth" preventing, will thus be minimized.

A similar approach may be taken to those who breach contracts upon which others have relied. If the full cost of the breach is brought to bear upon the promisor, then only those contracts will be breached for which the opportunity cost (either the alternative return foregone by entering the contract in question, or the cost of performance) exceeds the return plus the damages for breach. In general, the internalization of externalities through imposition of legal liability works toward introducing the optimal level of deterrence - that level which minimizes the social cost of the activity.

In the absence of enforcement costs, the optimal deterrence level is attained by forcing actors to pay the full costs of their actions. It is not necessary in the context of the model, however, for those who are injured to be the recipients, i.e., for plaintiffs to be compensated. If damages assessed to the defendant were paid to the state or any other institution, the same level of deterrence would result. What is necessary is for the defendant to be faced with the full costs of each option when choosing among different behaviour patterns. Nevertheless, if liability is to be placed upon a defendant through the civil law process, the payment of compensation to the plaintiff is of critical importance as an incentive for the plaintiff to bring the action initially. Thus, although compensation is irrelevant for deterrence, it is a necessary aspect of the civil process, particularly in the absence of administrative agencies responsible for enforcing liability, or in those cases where agencies are inappropriate for ferreting out the cause of action initially.

---

4. See R.A. Posner, Economic Analysis of the Law (2nd ed.) (Boston: Little, Brown and Co., 1977).

4. Implications of the Choice of Model

The behaviour control model clearly lends itself as a conceptual framework to an economic analysis of the law. It does not preclude a concern with equity - it may in fact facilitate such a concern insofar as a court working within the parameters of this model is likely to be much more concerned with unjust enrichment than one working within the conflict resolution model.

If it is seen as desirable that those whose actions affect others be called upon to pay the full cost of such actions - not for reasons of equity, but for allocative efficiency - then problems may emerge if the rules of the civil process tend to lay an inordinate stress on conflict resolution. The conflict resolution model requires a clearly defined plaintiff sufficiently injured to be willing to bring an action. There are many cases however, particularly in the area of consumer protection, where a defendant's conduct may injure many, but each only slightly and not sufficiently to warrant any individual plaintiff being willing to incur the costs of an action in time, energy and money. Although the aggregate value of the damage may be great and public complaint may mount, in the absence of any administrative action the full costs of the activity will not be brought to bear upon the defendant. And should some plaintiff be willing to sue to recover his own damages, this amount will be small in relation to the total damage done and will prove no deterrent to the defendant. In short, misallocation of resources resulting from the divergence of social and private costs will continue.

One suggestion for dealing with the problem is to allow readier resort, particularly for consumers, to class actions. The problem of incentive remains however. If there is no clearly defined and identifiable class of plaintiffs, then (i) who is to bring the action, and (ii) how are the net damages (after costs) to be paid out? An obvious way to deal with the first problem is to allow a contingent fee as an incentive for counsel to bring the action - anathema in the context of the conflict resolution model. As for the second problem, in the context of the behaviour control model it really matters little what is done with what remains of the award after payment to counsel and to whatever identifiable plaintiffs do come forward.[5] What matters is that the defendant be faced with the full cost of his activities, even if this requires the enrichment of counsel.

---

5. See Daar v. Yellow Cab Co., 67 Cal. 2nd 695, 433 P. 2d 732, 63 Cal. Rptr. 724 (1967).

CHAPTER III

CONSUMER PROBLEMS - DEFECTS AND ACCESS TO REDRESS

Many pitfalls may await the consumer, including fraud, misrepresentation, danger, and defective products. In what follows, we shall deal only with defective products - products, loosely put, which do not perform according to the reasonable consumer's expectations. The analysis, particularly that dealing with availability of redress mechanisms, may be adapted without difficulty to deal with other consumer complaints.

1. The Optimal Rate of Defects

    (a) The Marginal Cost of Eliminating Defects

No process of production is infallible - error, whether human or mechanical, is omnipresent. The rate at which mistakes occur or, of more concern here, at which defective products pass into the sales stream, is subject to control, however. More or less resources can be devoted to the detection and elimination of defects. To reduce the rate of defective products is of course costly. We may safely hypothesize in addition, however, that it becomes marginally more and more expensive to reduce the defect rate more and more. For example, to reduce the rate of defective widgets from, say, 100 per day to 99 per day may require only a casual visual inspection of every tenth widget. To reduce the rate further from 99 to 98 defective widgets may require the casual visual inspection of every fifth widget, and so on. At lower levels of defects, redesign of productive processes and changed tolerance limits may be necessary to obtain but a slight reduction in the defect rate. Finally, it is not unrealistic to suggest that it is ultimately impossible to reduce the defect rate from one to zero (or put it another way, that an infinite expenditure of resources would be necessary to obtain a wholly defect-free production process). All of this may be summarized in the assertion that a producer faces an increasing marginal cost of elimination (MCE) of defects and this is portrayed in Diagram 1.

$

MCE

DIAGRAM 1

11
10
0

40   99   100   Defect Rate
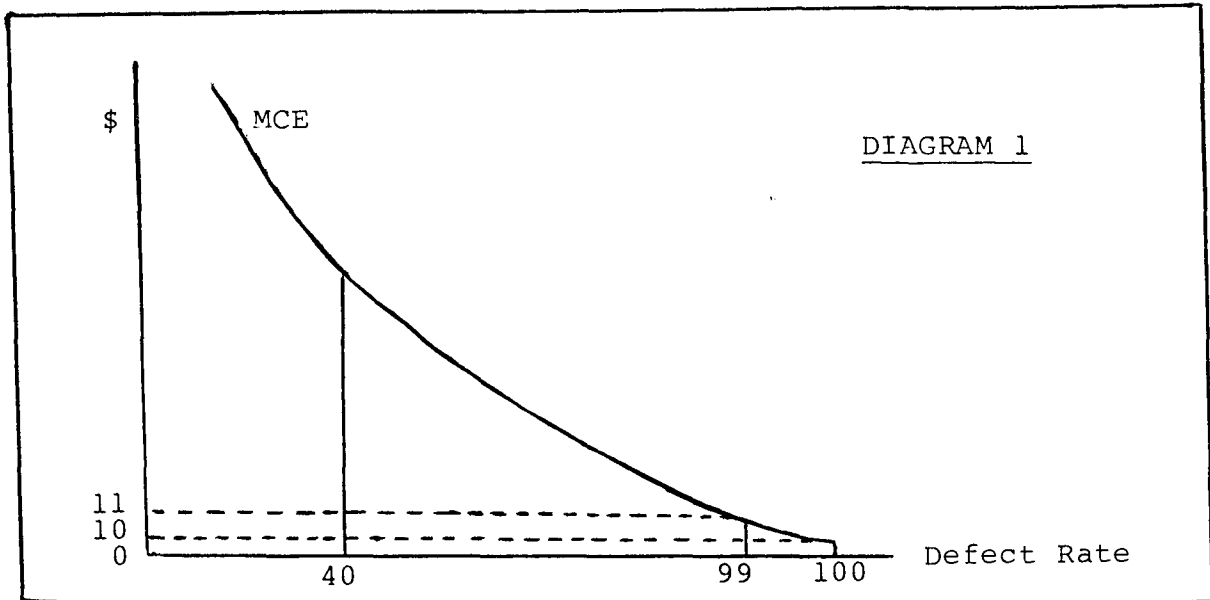
        The MCE curve is arbitrarily drawn to show that if
absolutely no attention (read resources) was devoted to detection
of defects, they would occur at the rate of 100 per period. To
eliminate the 100th defect, i.e., to reduce the defect rate to
99, would require an expenditure of $10. To then eliminate the
99th defect would require a further expenditure of $11, i.e., a
total expenditure of $21 would be necessary to reduce the defect
rate to 98, and so on. The cost of eliminating the marginal
defect per period is always greater than the cost of eliminating
the previous defect per period and the total cost of eliminating
the defects so far eliminated is the sum of the individual
marginal costs of elimination. The total expenditure on defect
elimination translates diagramatically into the area under the
MCE curve to the right of the current defect rate. For example,
if 60 defects per period are eliminated, thus reducing the defect
rate to 40 per period, the total cost of eliminating the 60
defects is the area under the MCE to the right of the vertical at
40.

        (b) The Marginal Benefit of Eliminating Defects (MBE)

        Why should a producer be willing to expend resources to
reduce the rate of defects? The obvious answer is that there
must be some benefit to be derived from such expenditure. The
actual marginal benefit to be derived from the elimination of
successive defects will depend on many factors including market
structure and legal rules of liability. For now let us assume
for purposes of example that the benefit from each defect
eliminated is constant. This is illustrated by the MBE curve in
Diagram 2. The total benefit derived will be the sum of benefits
from each defect eliminated and will be the area under the MBE
curve to the right of the remaining defect rate. Thus the extra

benefit to be derived from eliminating the D_2nd defect is fg and the total benefit from reducing the defect rate from D_0 to D_2 is gf times $D_0 - D_2$.

DIAGRAM 2
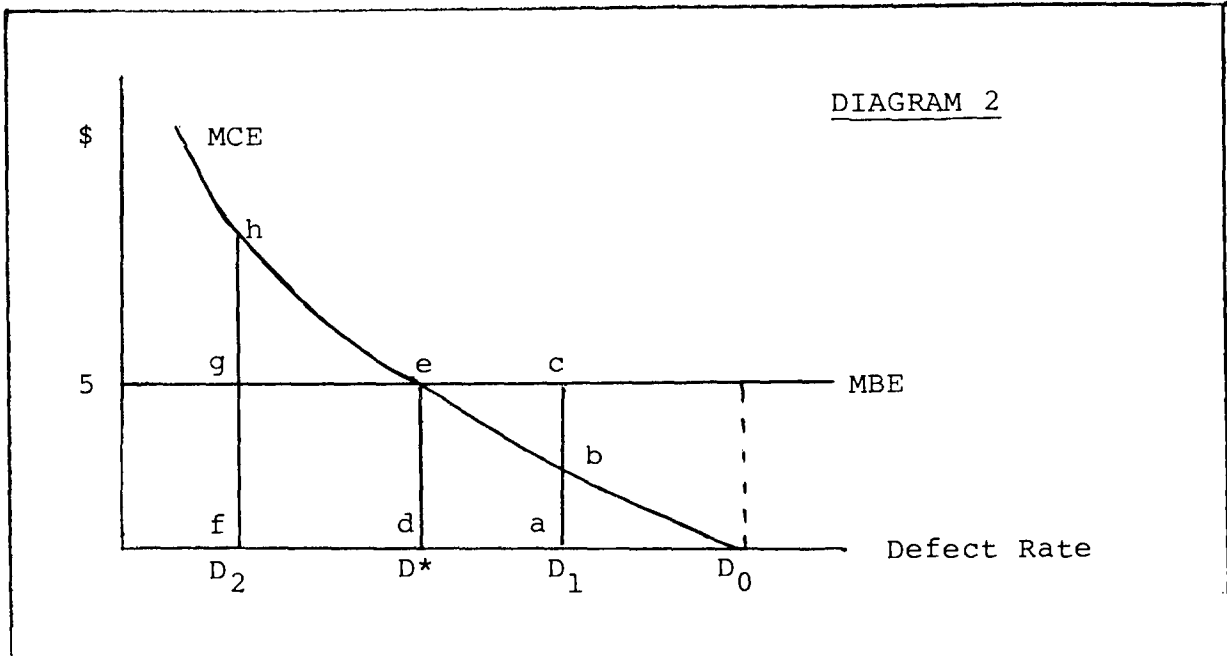
c)   The Optimal Defect Rate

It is in the interest of the producer to eliminate a defect if the cost of eliminating that defect is less than the benefit to be derived from the elimination of that defect. Similarly it is to the benefit of the producer to allow the marginal defect per period to occur if the cost of eliminating it is greater than the marginal benefit of elimination.   In terms of Diagram 3 the optimal defect rate will be D^*. The total prevention cost will be the area D^*eD_0 and this will be less than the benefit of eliminating $D_0 - D^*$ defects, this benefit being D^*e times $D_0 - D^*$.   The equilibrium defect rate will clearly be found where the marginal cost and benefit to the producer are equated.

d) The Optimal Defect Rate Under Different Liability Rules: The Coase Theorem

In a pathbreaking article published in 1960,[1] Ronald Coase established a proposition which has come to be known as the Coase Theorem. Put generally, the theorem states that if there are no costs of transacting between parties in an interacting activity (one wherein one party's behaviour has consequences – good or bad – for the other) then the legal assignment of liability (or benefit) will have no effect upon the allocation of resources – i.e., upon the level of the activity.

We may observe the application of the theorem in a simplified example involving a producer and consumers of potentially defective products. But first some simplifying assumptions must be made, the most dramatic of which is that there exist no transactions costs between consumers and producers. This means in effect that consumers as a group can costlessly deal in one voice with producers and vice versa. For this example we also assume that a defective product is incapable of being repaired. Therefore the cost to whoever must bear it is the replacement price of the product. A defective product is thus valueless junk in the hands of whichever party it is left with on the full completion of exchange. Finally, we assume that consumers and producers have full information with respect to defect rates, though not of course with respect to which particular items will prove defective.

Consider first the case where consumers who have bought defective products (now junk) have no legal remedy. In other words, liability for defective products is placed upon consumers... caveat emptor. The consumer has paid the full price of the defective product and has, in effect, received nothing. The full price is retained by the producer, who provides no in-house redress.

Consumers are not powerless, however. Consider the situation portrayed in Diagram 3. In the absence of any other consideration, the defect rate would be D_0. For a very small additional payment over price P from consumers, however, the producer should be willing to reduce the defect rate to $D_0 - 1$. The payment need be only slightly in excess of the producer's marginal cost of elimination of the D_0th defect. The benefit to consumers will be the full product price (otherwise lost) less the small payment to the producer. As long as the MCE is less

---

1.   R. Coase, "The Problem of Social Cost", Journal of Law and Economics (3:1960), p. 1.

than the product price, it is in the interest of consumers and producer for consumers to pay the producer a sum slightly in excess of the producer's MCE for the producer to eliminate that marginal defective product. Thus for example if the defect rate is D_1 and consumers do nothing, they will lose ac on each defective product and their total loss will therefore be ac times D_1. However, by offering the producer slightly in excess of ab they can persuade the producer to reduce the defect rate to $D_1 - 1$. Their saving, by eliminating the D_1th defect is thus ac – ab, or bc. Consumers can in this fashion persuade the producer to reduce the defect rate to D^*. It is not in consumers' interest to try to buy a further reduction in the defect rate below D^*, since the required payment, a sum slightly in excess of the producer's MCE, would be greater than the loss the consumer would otherwise sustain. D^* thus becomes the equilibrium defect rate, and the price line becomes in effect the producer's MBE schedule since that reflects the payment from consumers which the producer foregoes if he chooses to allow the marginal defect.

DIAGRAM 3
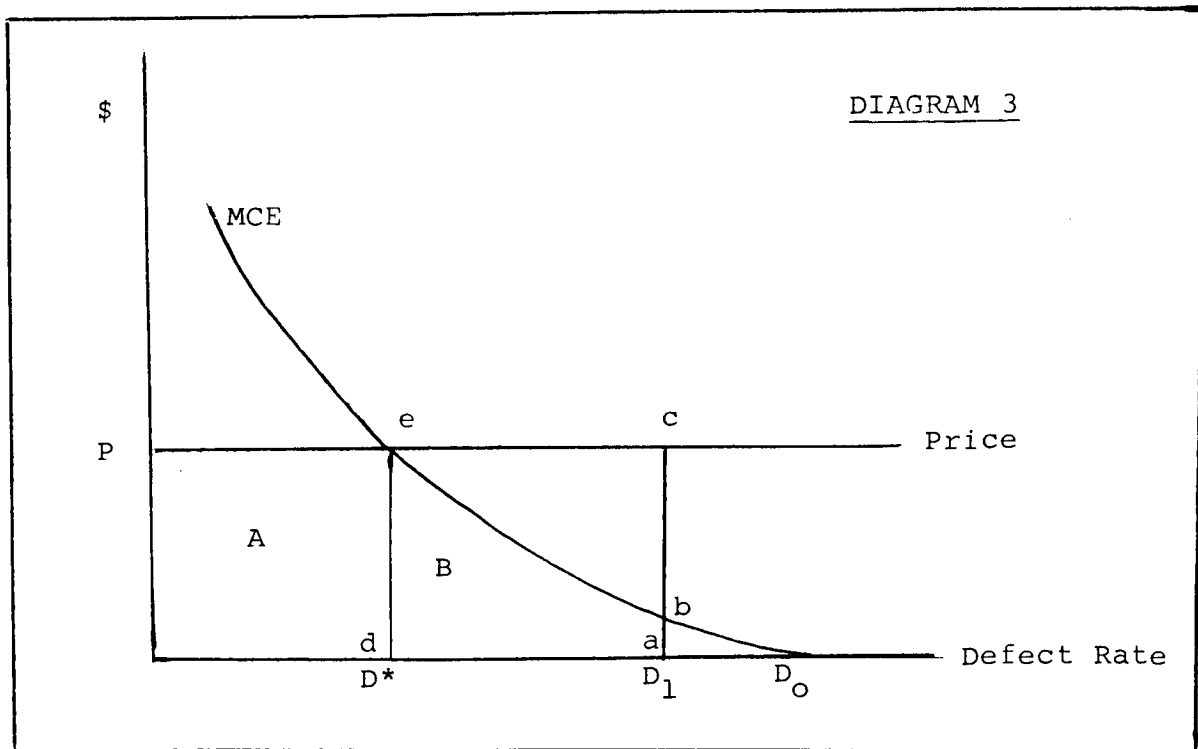
It should be noted here that in the context of this example, all consumers have an interest in paying producers to reduce the defect rate to D* _prior_ to any purchases. After the fact however, once a consumer is left with a defective product, no payment to the producer can avoid or reduce that particular loss.

Finally, we may note the total cost to society of defective products. As stated earlier, this cost is made up of two parts, the cost arising out of the actual defect rate, and the cost of the measures taken to prevent further defects per period. In this case, the cost of the realized defects will be P times D*, or the area of the rectangle A, and the prevention cost will be the area under the MCE curve designated B, or deD_0. If consumers have no remedy, then they alone bear the full social cost, A + B.

Now let us consider a full reversal of the liability rule. The consumer now has the right to return the defective product for a full refund of the purchase price from the producer. Furthermore, this legal right is assumed to be costlessly enforceable. As before, a defective product is valueless in whosever hands it may come to rest. Clearly again, the equilibrium defect rate will be D*. On each defective product the producer will lose the full cost of production including profit - i.e., the producer will lose the price of the product for each product that is defective. The price line thus again becomes the schedule reflecting the marginal benefit of eliminating successive defective products per period. The producer will reduce the defect rate as long as each marginal defect eliminated costs less to eliminate than the benefit derived from its elimination. At D* there will be incentive to neither expand nor decrease the defect rate. As before, the total social cost of defective products will be A + B, only in this case the cost will be borne wholly by the producer.

The Coase Theorem conclusion that assignment of liability is irrelevant to resource use in the absence of transaction costs is thus confirmed in our defective product case, the defect rate and total social cost of defects being the same regardless of whether producer or consumer bears the cost. It is also evident that in the absence of transactions costs, an externality - a cost or benefit generated but not faced by the producer - cannot occur. Even though in the case of full consumer liability the consumer alone pays for the elimination of defects down to D*, the producer is still confronted by his own cost of elimination schedule, since that represents his opportunity cost of defect elimination, and by the price line, which in the neighbourhood of D* represents the producer's opportunity cost of forebearing to eliminate a defect. In short, regardless of the liability rule, all costs and benefits in this example are

internalized to the transaction, thus guaranteeing that the socially optimal quantity of resources will be devoted to defect prevention.

(e) Repairable Defects

The analysis is not greatly affected by dropping the assumption that the defect cannot be repaired. We may first note that both producers and consumers are capable of rendering repairs to a defective product and that the party who is able to repair at the least cost will vary with the product and the nature of the defect. If our other assumptions are left intact then only the lower repair costs are relevant. In what follows, we simply assume the producer can repair the product at least cost and that the marginal cost of repair (MCR) is constant, as in Diagram 4.

DIAGRAM 4
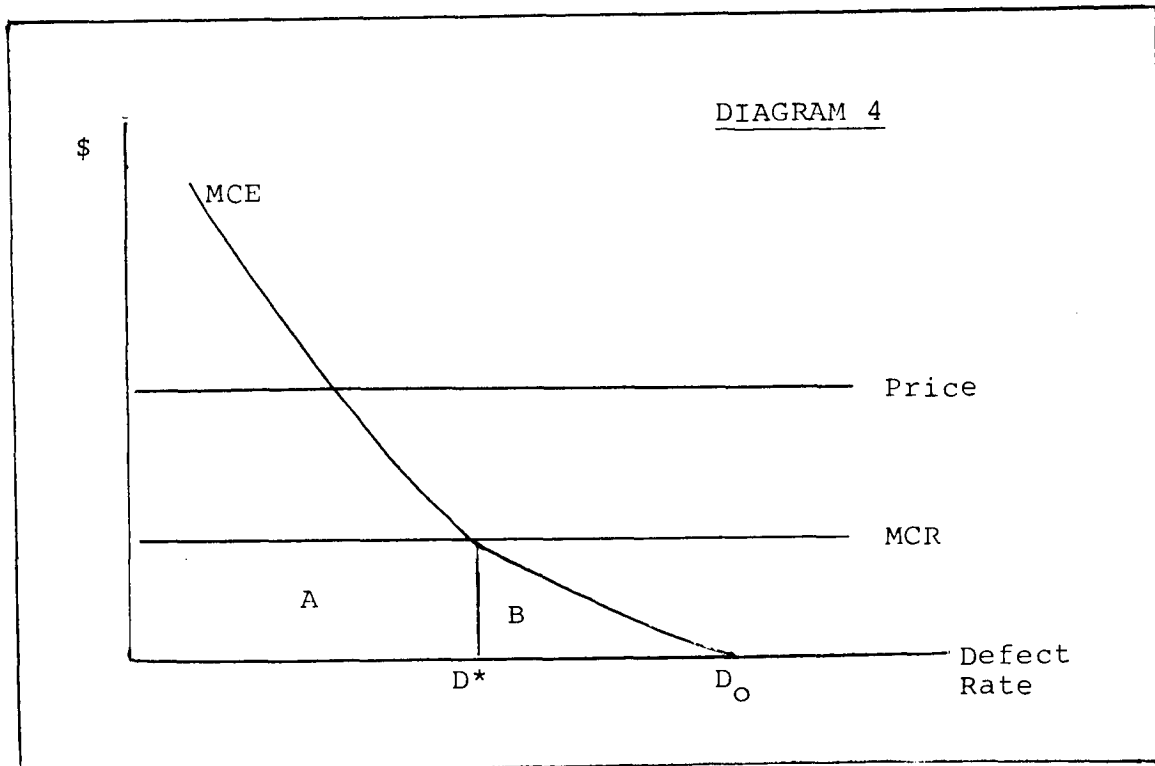
Again, we have two possible liability rules: producer liability wherein the producer is obliged to repair or replace,2 and consumer liability wherein the producer is under no legal or voluntary obligation to supply redress.

If the producer is liable, it will be in his interest to reduce the defect rate from D_0 to D^* as that is his least cost action. He will then provide repairs on the D^* remaining defective products. The total social cost of defective products will be the repair costs on the D^* defects, being rectangle A , and the cost of eliminating $D_0 - D^*$ defects, being the area under the MCE schedule designated B. In this case, A + B falls wholly upon the producer. D^* is the optimal as well as equilibrium defect rate as it represents the lowest social costs of defects.

The defect rate will still be D^* when we reverse the liability rule and have consumers wholly without legal redress. Consumers will pay producers slightly in excess of producers' MCE to reduce defects to D^* and thereafter will pay producers to repair the D^* defective products at slightly in excess of producers' MCR. Again, total social cost of defects will be A + B, now wholly borne by consumers and this is, as before, the lowest possible social cost.

We may note that the equilibrium and optimal defect rate, will be <u>greater</u> when repairs are possible (regardless of who is the least-cost repairer) than when they are not repairable. In terms of the previous diagram, had repairs not been possible the defect rate would have been determined by the intersection of the price and MCE lines and a rate of defects lower than D^* would have resulted. The explanation for the higher defect rate when defects can be cured for less than their cost of production, is that a repairable defective product represents lower social costs than a non-repairable one, and the lower the cost of repair, the less the social concern with defects. Thus if a defective product could be repaired by a consumer, say, by simply adjusting a nut or bolt, then neither producer nor consumer would be much concerned and the optimal defect rate would be close to D_0 in terms of Diagram 4.

---

2. If a repaired product is indistinguishable from an originally non-defective product, it is assumed that both the consumer and producer will be indifferent between repair and replacement.

(f) Distributional Consequences

Though D* may be the same under our assumptions so far regardless of the liability rule chosen, the distributional consequences will clearly vary: producers will pay A + B when they are liable and consumers will pay it when they are liable. Little will be said now as to the equitable choice with respect to cost bearing. It should be borne in mind however that the lay image of a defective product as representing "sloppiness" or lack of concern by producers is of little help in allocating costs. Defective products are a necessary by-product in a normal process of production. Their number can be reduced only by the further expenditure of resources.

(g) Transactions Costs

In a frictionless world, the rule of liability is irrelevant. But when transactions costs exist, costs so high as to prohibit effective voluntary negotiations between consumers and producers, the choice of liability rule will determine the allocation of resources to defect prevention.

Consider first the situation for non-repairable products. If the producer of a defective product is fully liable, and if by "liability" we mean that the rule is costlessly enforceable, then the equilibrium defect rate is D^*, as in Diagram 5, and this is the same defect rate as in the case of zero transactions costs. If, however, the consumer is liable, then in the presence of prohibitive transactions costs, there will be no incentive for the producer to take any preventive action and the defect rate will be D_0, much in excess of the optimum. The net social loss of defect rate D_0 rather than D^* will be the benefits foresaken by operating at D_0 rather than D^*, which is the area A + B, less the costs of prevention not incurred by operating at D_0 rather than D^*, which is the area A . The net social loss is thus the area B .

DIAGRAM 5
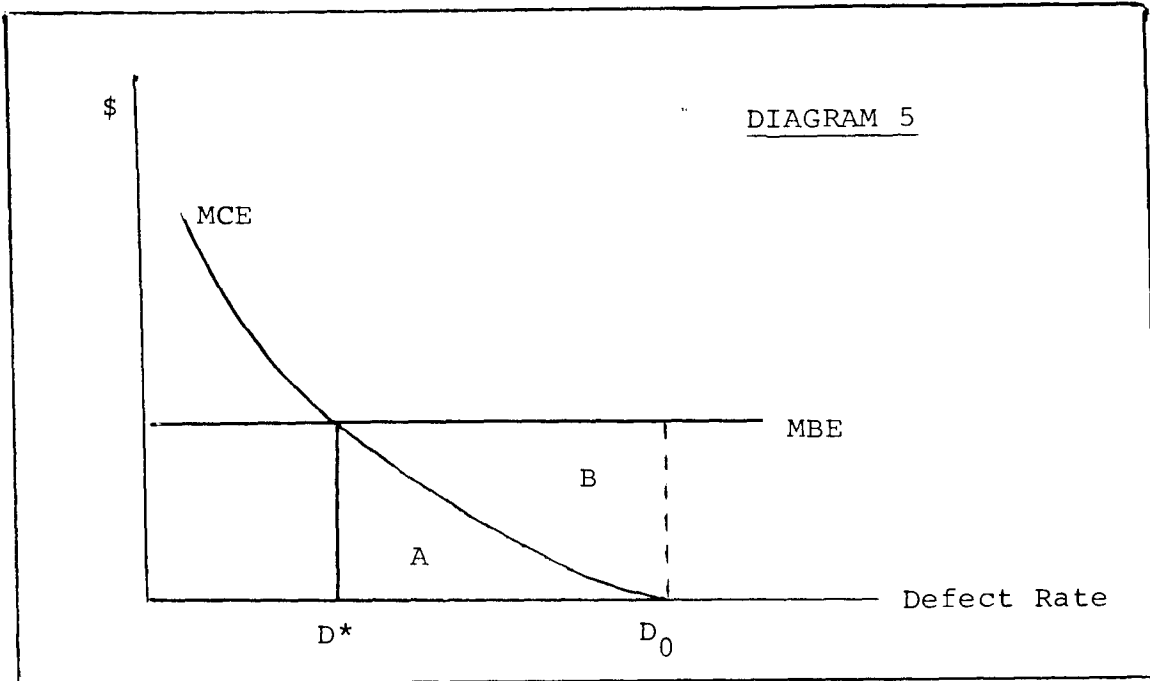
Thus in the absence of repairability, in the presence of transactions costs which prohibit bargaining between producers and consumers, the liability rule which yields the optimal social outcome is that of producer liability. Note, however, the importance of the assumption that the liability rule is costlessly enforceable. If legal redress is available and costless to the consumer, then generally the producer will have an incentive to provide the redress directly. If, in spite of the statement of the liability rule, the costs to the consumer of enforcing liability are prohibitive, then effectively the rule becomes one of consumer liability with the consequent social costs.[3]

---

3. "... (the) focusing of attention on the creation of new doctrines and laws has tended to obscure the need for a thorough examination of the adequacy of the mechanisms through which these new rights are to be made effective." Eovaldi, T.L., and Gestrin, J.E., "Justice for Consumers: The Mechanisms of Redress", Northwestern Law Review (66:1971), p. 281. Eovaldi and Gestrin also quote Ralph Nader on the danger of confusing "verbal and symbolic momentums - such as new toothless laws with no funding and enforcement - with true progress" and Betty Furness, at the time Special Assistant to the President for Consumer Affairs, on the problems of "name only" bills, p. 281, footnote 3.

The situation is less clear when both producers and consumer can repair defective products but at different costs. Diagram 6 illustrates the situation wherein the producer has the lower costs of repair.

DIAGRAM 6
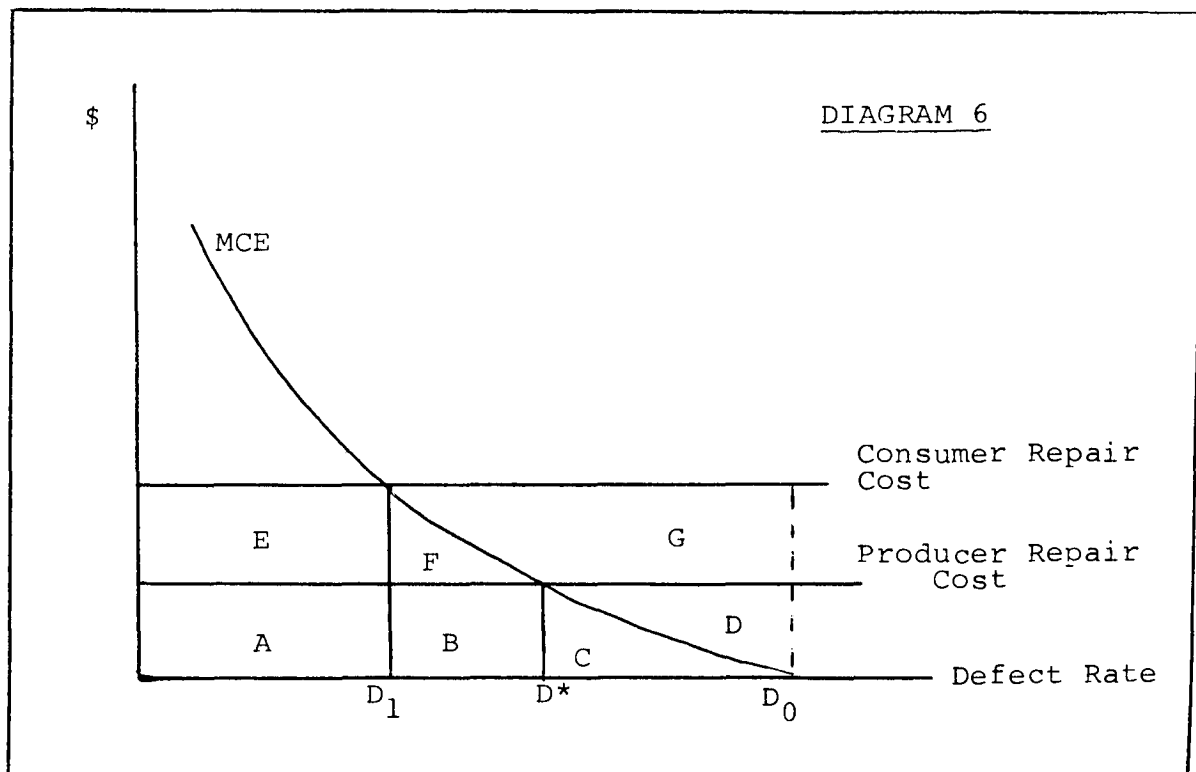
If the producer is held liable, then the optimal defect rate, D^*, will result. If, however, the consumer is liable for defects and there is no scope for negotiation whatsoever, then the producer will face no incentive to prevent defects, D_0 defects will occur and the consumer will be liable for D_0 times the repair costs per item. In terms of the diagram, the extra social costs incurred are represented by the area E + F + G + D. If we allow for ex post negotiation between the producer and the individual consumers, i.e., negotiations not as to prevention, but as to who is to pay for and perform the repairs, then consumers would be able to pay producers their lower costs to repair the defective products. This will have no effect on the non-optimal defect rate D_0 but it will bring the excess social cost down to the area D. We may here again conclude that when producers have the lower costs of repair, only the rule holding producers liable for defective products will, in the presence of prohibitive transactions costs, lead to the optimal defect rate.

If consumers have the lower repair costs, then <u>neither</u> rule in the presence of prohibitive transactions costs <u>will lead</u> to the optimal rate of defects. This is illustrated in Diagram 7.

DIAGRAM 7
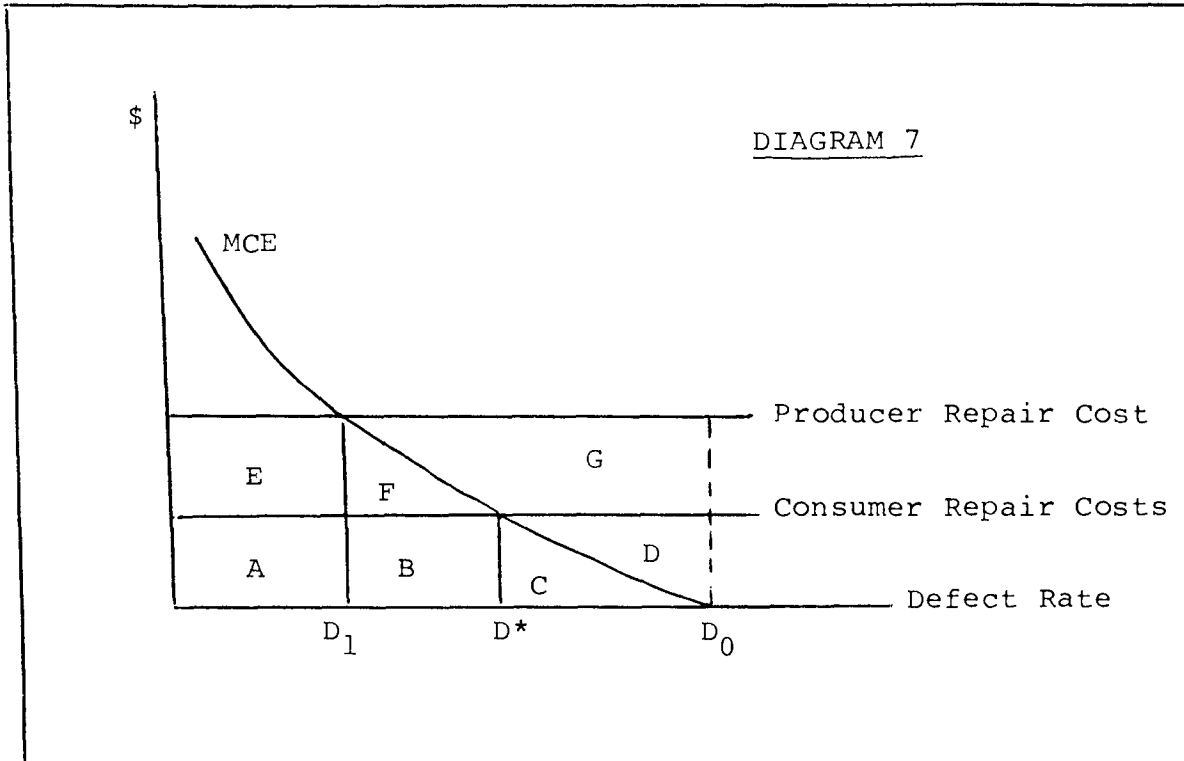
The socially optimal defect rate, that which minimizes the sum of prevention and repair costs, is D^*. If producers are liable, however, <u>their</u> optimal defect rate will be D_1, since beyond D_1 it is cheaper to eliminate defects and below D_1 it is cheaper to repair them. A defect rate of D_1 represents an excess in social costs over that of D^* by an amount designated by the area E + F .

If the transactions costs assumption is varied, however, to allow <u>ex post</u> payments by producers to consumers to allow consumers to repair the products, then D^* can be attained, it being cheaper to eliminate $D_0 - D^*$ defects than to repair them and cheaper to pay consumers to repair the remaining D^* defects than to eliminate any of them.

If consumers are liable for defects, given this cost structure, then producers will have no incentive to reduce defects, the defect rate will be D_0, and consumers will expend A + B + C + D on repairs, which is greater by the area D

than if, with zero transactions costs, they could pay an amount equal to area C to persuade producers to restrict the defect rate to D*.

The choice of liability rule appears to be clearly in favour of holding producers liable throughout. The case is unarguable if the remedy in question is restitution or replacement. Where the appropriate remedy (in terms of resource use) is to repair the product, if the producer has the lower repair costs, producer liability will still yield the optimum defect prevention. Even if consumers have lower repair costs than producers, holding producers liable is more likely, given the ease with which ex post payments from producers to consumers may be made, to lead to the optimum rate of defect prevention.


(h) Time and Trouble Costs

We have so far assumed that the social cost of a defective product is either the price paid for the product if it is unrepairable, or the repair costs, whichever is lower. These are almost certain to underestimate the true cost, however, and if the liability rule holds the producer liable only for the price or repair cost, a higher than optimal defect rate will result. In short, the producer will not be faced with the full social cost of his activities.

Initially, the full cost of the product to the consumer is the money price paid plus the opportunity cost of searching, comparing products, gathering information, and so on. The information obtained by this expenditure of resources is not rendered useless when the product turns out to be defective, however. Though the consumer is left with, say, a piece of junk, his stock of information is still worth as much as before, if not more.[4] Should he be compensated by the return of the purchase price, by replacement of the product or by the repair of the product, he need not again incur the information costs. There is, therefore, no reason to compensate the recipient of a defective product for his information gathering costs.

---

4. His information stock with respect to defective products may be enhanced by the actual receipt of a defective product. If so, the change in the value of the stock should be subtracted from his compensation.

Of more concern are the "time and trouble" costs incurred after discovery of the defective nature of the product. To the extent that the redress received by the consumer fails to compensate him for his time and energies associated with the process of obtaining relief, the telephone calls, visits to the seller, negotiations, anxiety, etc., the schedule showing the marginal benefit of eliminating defects which the producer faces will be below its social value. In other words, the private marginal benefit of eliminating a defect will be below the social marginal benefit. This is illustrated in Diagram 8 where the price line reflects the payment to the consumer either in the form of return of the purchase price or replacement of the product, and the P + T line is price plus the opportunity costs of the after-discovery time and trouble.

DIAGRAM 8
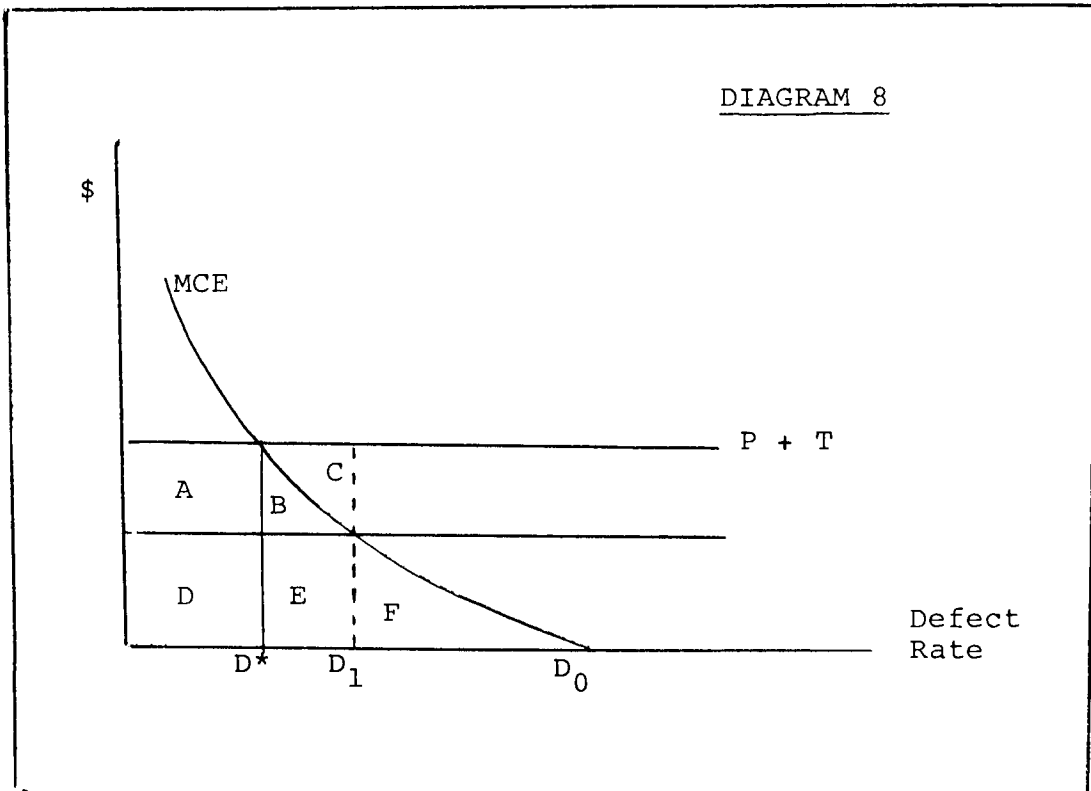
The socially optimal defect rate is at D*. This may be seen by considering, under a rule of full consumer liability with zero transactions costs, how far consumers would be willing to pay the producer to reduce defects. It is in the consumers' interests in such a situation to pay the producer's marginal cost of eliminating defects as long as that MCE is less than the cost to the consumer of the defective product. This yields the defect rate D*. However, if the compensation to the consumer is only the return of the purchase price or replacement of the product, then defect rate D_1 will result, the consumer will be liable for an amount equal to the area A + B + C, and the total resources expended by both consumers and producers will be greater by the equivalent of area C than at D*.

If the product is repairable, then the optimal defect rate will be determined by the intersection of the producer's MCE and the schedule reflecting the sum of repair costs plus time and trouble costs.

2. Access to Redress

A rule of producer liability for defective products is only as good as the degree of access the consumer has to enforcement mechanisms. In the extreme case of zero access, the rule is in effect converted to one of full consumer liability and this, coupled with prohibitive transactions costs with respect to ex ante negotiations between consumers and producers, must lead to inadequate expenditure by producers on defect prevention and a higher than optimal defect rate. (Henceforth we assume prohibitive transactions costs with respect to consumer-producer negotiations over the defect rate.) It is tempting to suggest that no legal rule of liability can exist without the availability of an enforcement mechanism. The error in this may be observed simply by considering the legal rule of consumer liability for defective products. Assuming payment is prior to possession by the consumer, no enforcement mechanism is required.

(a) The Cost of Access

We shall consider the cost of using the mechanism as the only barrier to access to the means of enforcing the liability rule. Subsumed under "cost" are all the attendant expenses, pecuniary and non-pecuniary, of gathering information with respect to the enforcement mechanism and the substantive and procedural rules, hiring appropriate personnel, paying necessary fees, documenting the case, attending the court or other agency, enforcing the judgment, and so on. Also included in costs are any psychological impediments faced by the consumer in the use of

the enforcement mechanism, e.g., communication problems in dealing with personnel within the legal bureaucracy, fear of or lack of trust in authority, and so on. Finally, there is included the uncertainty associated with judicial or other determination.5

Most of these costs, if not explicitly pecuniary, can at least in theory be imputed. Such is the case for example with any input of time. It is arguable whether this can be done with psychological barriers, however. These barriers may nevertheless be treated qualitatively in the sense that we may treat an increase in the complexity of the process as an increase in costs (even though the time input may remain unchanged) or an increase in difficulty in communication as an increase in costs, or the provision of a court document translated from legal terminology into everyday language as a decrease in costs, and so on.

(b) Demand for Judicial Redress

The lack of use of a judicial redress mechanism by a particular class of complainants, e.g., consumers, does not imply that that class has been effectively excluded from access to the mechanism. If all substantive legal rules were clear both in terms of their content per se and their application to particular disputes, then there would be no need to resort to the use of the mechanism, particularly if the cost of access to it were zero. No producer would deny voluntary redress to a consumer if it was evident that (a) the liability rule unequivocally favoured the consumer, and (b) the consumer could enforce the rule costlessly. In such a situation the number of disputes judicially resolved is no guide to accessibility.

The situation changes, however, when there are uncompensated costs associated with the use of the judicial redress mechanism. Such costs are clearly a barrier to use of the mechanism. In many consumer dispute situations the total cost (pecuniary and psychological) of using the mechanism will more than offset the full value of the grievance, particularly for lower valued items. This will remain the case even if some part of the costs are compensated for the successful consumer litigant.

---

5. The next chapter will deal in some detail with the consequences flowing from the uncertainty of outcome in a judicial determination.

From the producer's perspective, the presence of costs to the consumer of using the judicial redress mechanism appears as a reduction in the marginal benefit incurred from the elimination of defects. Knowing that a proportion of consumers receiving defective products will be dissuaded from seeking official redress, the producer will have an incentive at least temporarily to refuse and generally delay voluntary redress. If the proportion of the total that ultimately turns away from seeking both official and voluntary redress is t, the marginal benefit will be 1 - t times what it would otherwise be. This is reflected in Diagram 9, wherein P is the marginal benefit schedule if all defects are redressed, and the schedule (1 - t)P is the marginal benefit to the producer after allowing for "discouraged" consumers. The marginal benefit schedule is, in effect, discounted by producers to the extent that consumers do not pursue their remedies.

DIAGRAM 9
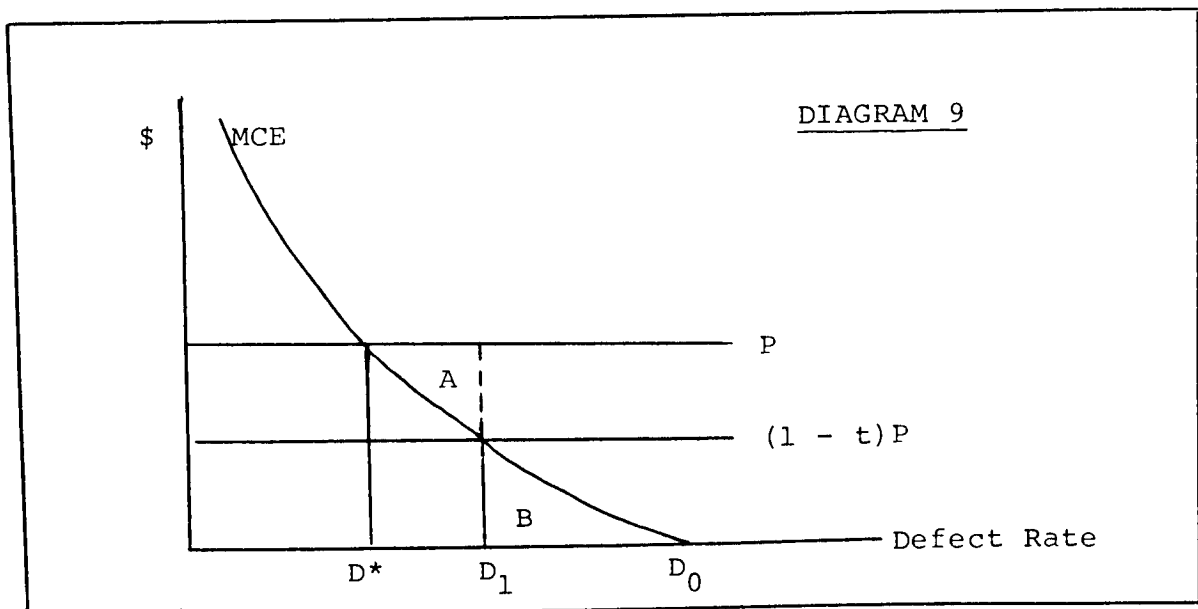
Faced with an effectively lower marginal benefit schedule, producers will be led to a higher defect rate, D_1. Actual redress "voluntarily" provided will be D_1 times $(1 - t)P$ and resources equivalent to area B will be expended in the elimination of defects. Consumers who are dissuaded from seeking redress will bear the cost D_1 times tP. The total social cost will be the sum of the redress cost borne by producers, the "absence of redress" cost borne by consumers, and the cost of resources devoted to defect elimination. These costs add up to a total which is in excess by the amount indicated by area A over the total social costs at D^*.

In addition to implying greater social costs due to a higher than optimal defect rate, the cost barriers with respect to access to the redress mechanism effectively convert the liability rule from one of full producer liability to one of partial producer and partial consumer liability, with those consumers having the higher costs (financial and psychological) now liable.

(c)   Access Cost Subsidies

A public subsidy of court costs does not reduce social costs per se. It merely transfers them from particular consumers to the public at large. Were this the only consequence, the merits of such a subsidy could only be assessed in the political arena. But of course, this is not the only consequence. From the perspective of the behaviour control model the justification for institutional subsidy rests on its effects for deterrence, in this case, the reduction in defects to their optimal number. In general, any reduction in "system costs" implies a reduction in t, the proportion of grievors unable or unwilling to fully pursue their remedies. A reduction in t implies a reduction in the defect rate, a consequent reduction in social costs, and hence an increase in social welfare.

We may note that the subsidy is expected to be more relevant with respect to lower cost consumer items. Litigation costs per unit of service are absolute, not proportional (subject of course to the different monetary jurisdictions of the small claims court, county court, and supreme court). The absolute nature of the costs implies that they form a smaller percentage offset to consumer recovery for higher valued consumer grievances and are hence less of a barrier to the pursuit of judicial remedies for those grievances. This being so, we would expect t, the proportion of consumer complainants who fail for one reason or another to pursue their legal remedies, to be inversely related to the value of the commodity. Thus, subsidizing court costs for all consumers implies subsidizing some who would be prepared to pursue a judicial remedy in any case. From a

societal perspective, nothing is gained by this aspect of a subsidy, the social purpose of the subsidy being to induce otherwise reluctant consumers to pursue their complaints.

A general subsidy should be contrasted with alternatives for accomplishing the same end. One obvious measure would be to subsidize on an inverse absolute scale, with the amount of subsidy in absolute terms declining with the increase in the magnitude of the claim. Another alternative, one more cognizant of the psychological costs associated with judicial redress, is to devise redress institutions which are themselves less costly to operate and which are aimed specifically at small scale consumer complaints.

### (d)   The Social Goal of a Procedural System

Three sources of real costs have been identified: the resource waste (or repair cost) associated with defective products, the resource cost associated with the prevention of defects, and the cost of the redress institution. From an economic perspective, we may state the goal of a procedural system to be the minimization of the sum of the three costs.

3.   Consumer Surplus Analysis of the Social Loss

To this point, the sole problem emerging from unremedied defective consumer products has been the social cost of the higher than optimal defect rate resulting from the divergence between the social marginal benefit schedule and the private marginal benefit schedule facing the producer. Nothing has been said of the magnitude of the loss borne by the consumer and it may have appeared that this cost is simply the price of the product times the number of defective products for the simple case of wholly unrepairable products, or the repair cost times the number of defective products for repairable products. This would however be an underestimate of the consumer loss, and to illustrate that it is necessary to introduce the concept of consumer surplus.

(a)   Definition of Consumer Surplus

Consider a normally shaped downward-sloping demand curve for a product (widgets) as illustrated by DD in Diagram 10.
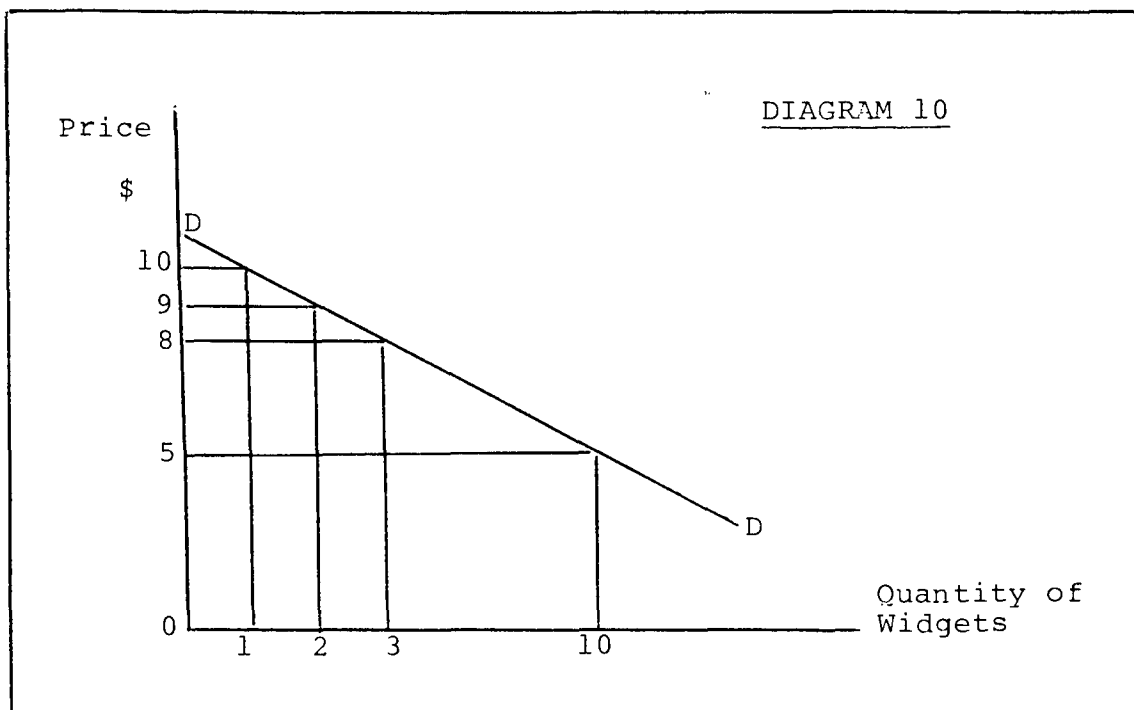
The demand curve indicates the quantity demanded by the community at each possible price of the product. Simply put, quantity demanded increases as price decreases for two reasons: first, existing purchasers tend to purchase more of the product, and second, new purchasers are induced to purchase. an extra unit of the product if the extra or marginal utility derived from consumption of the product is greater, even infinitesimally, than the utility of the dollars given up to purchase the product. If, as generally assumed, the marginal utility associated with successive units of the commodity consumed decreases, then a lower price will induce greater consumption. This is indicated in Diagram 10. If the price per widget is $10, then only one widget will be demanded, the value of the utility associated with the very first widget consumed being just in excess of $10. At any higher price no widgets would be demanded, for the value of the dollars that would have to be given up would be in excess of the value of the utility to be gained.

If the price per widget falls to $9, then clearly the first widget is still worth purchasing in each period, but also a second widget per period, yielding additional utility just in excess of $9 is worth purchasing. In this situation, a total of $18 for two widgets will be expended. But the total value of the

utility gained from consuming two widgets will be $10 from the first and $9 from the second, for a total of $19, for an excess of utility gained over that lost by giving up the purchase dollars of $1. Similarly it would take a price of $8 each to induce consumers to purchase three widgets per period. Total expenditure would be $24, but the value of total utility gained would be $27 (10 + 9 + 8) for an excess of utility over cost of $3. This excess is known as "consumer surplus" and represents, for each item purchased, the excess of the value of the extra or marginal utility gained from that item over the cost of the item. In terms of our previous diagram, if the market price of widgets is $5 per widget and 10 widgets are being consumed per period, then a consumer surplus is being gained on each of the first nine items consumed. On the tenth item, the value of the dollars given up to purchase that item is approximately equal to the value of the utility gained for that item.

In diagrammatic terms, the value of the total utility gained from consuming Q_0 items per period at a price of P_0 each, will be the area under the demand curve up to Q_0. In Diagram 11 this corresponds to the area A + B. The value of the dollars given up to purchase Q_0 items, $P_0 Q_0$, is the area denoted by B in the diagram. The difference between the two, area A, is thus the consumer surplus per period arising out of the consumption of Q_0 items per period at price P_0 each.

Price

$

DIAGRAM 11

D

A

P_0

B

D

Quantity

Q_0
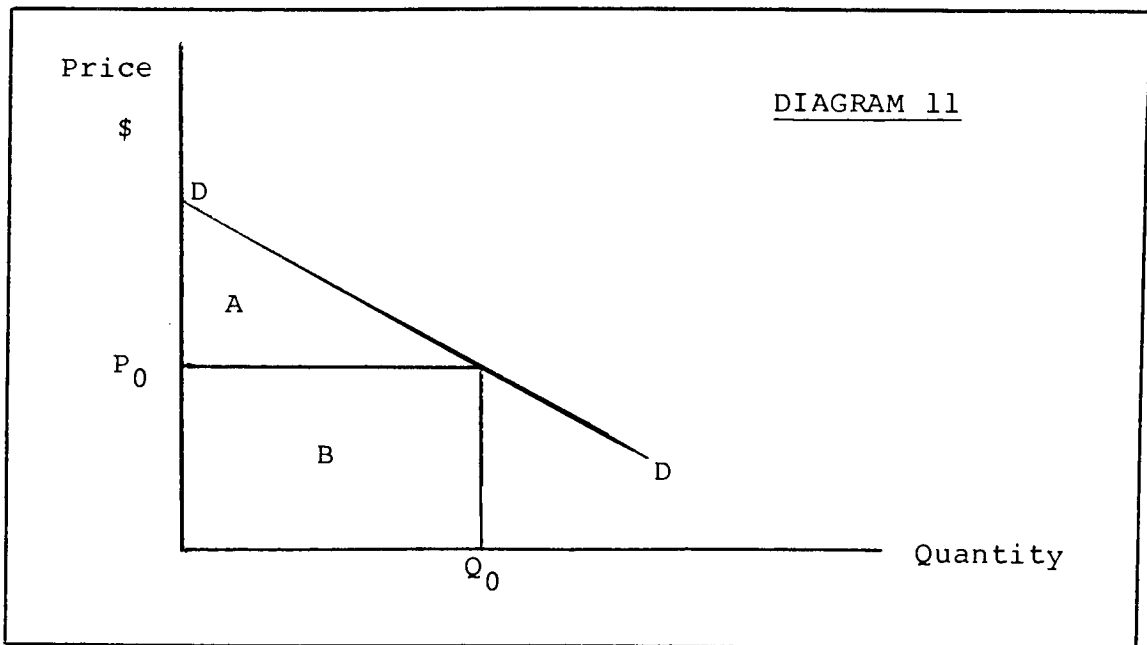
If the price of the commodity increases, fewer items will
be consumed per period and the consumer surplus will decrease. In
Diagram 12, Q_0 is purchased at price P_0, yielding a
consumer surplus of A + B + D.   If price rises to P_1 and
quantity demanded falls to Q_1, then the total utility will
decline to A + B + C (from A + B + C + D + E), total cost will
be P_1Q_1 or B + C, and the consumer surplus will have
decreased from A + B + D to A . Of the amount by which the
consumer surplus has decreased, B is simply transferred to
producers, implying on that account no clear diminution of social
welfare. There is however no "recipient" of the value
corresponding to area D. This is lost by consumers and gained by
no one. Hence area D is known as the "dead-weight loss"
associated with a price increase.
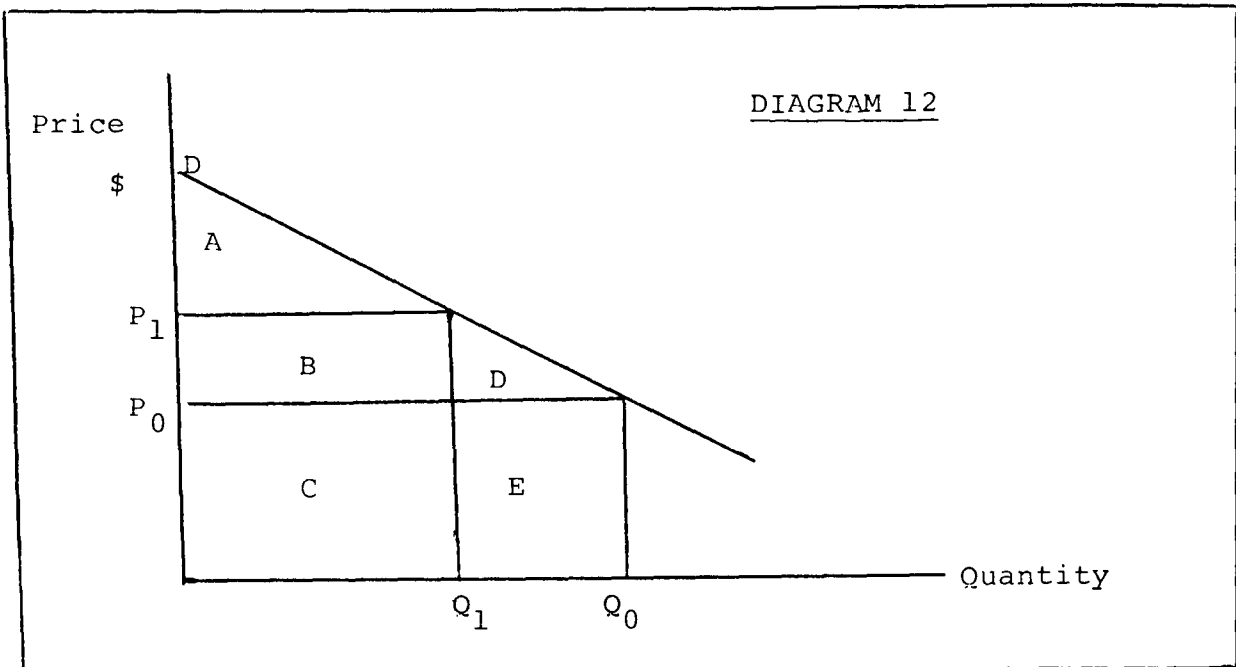
(b)   Consumer Surplus and Defective Products

Assume that the demand curve for widgets, reflecting the
utility to be derived from consuming widgets, is as shown in
Diagram 13.   If the price is P_0, Q_0 will be the quantity
purchased. Let us assume however that a proportion of Q_0 is
defective, so that the consumer receives only Q_1 usable
widgets. It is also assumed that the consumer is unable to
obtain redress for the defective products. This being the case,
the total utility derived from widgets will be the area under the
demand curve up to Q_1, and not up to Q_0, i.e., A + B + C
rather than A + B + C + D + E.

Although only Q_1 usable widgets are received, Q_0 widgets are paid for at price P_0 for each. Hence the effective price for each <u>usable</u> widget is $P_0(Q_0/Q_1)$, shown here as P*. The total cost can thus be shown as $P*Q_1 = P_0Q_0$ and is equivalent to the area B + C. The consumer surplus, the difference between the value of total utility and total cost is therefore equal to area A, and the loss borne by the consumer due to the unredressed defects is equal to the area B + D. Area B is transferred to producers and area D is the dead weight loss to the society.

DIAGRAM 13
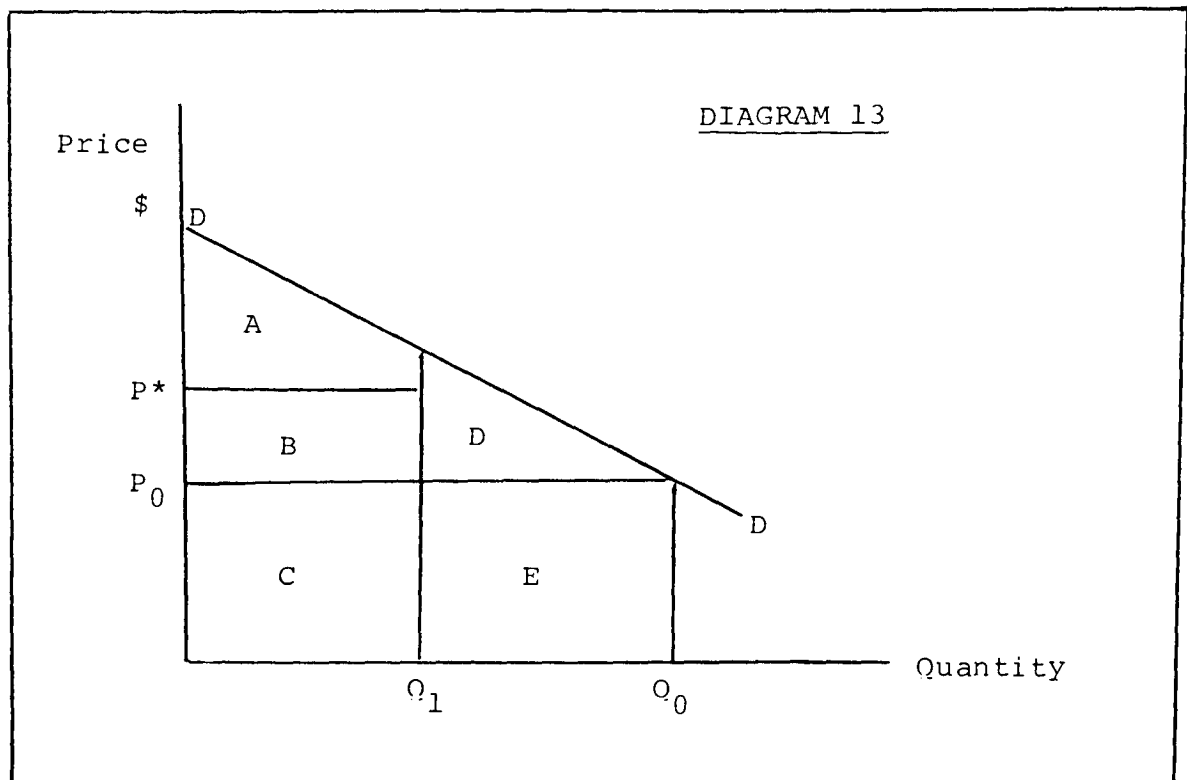
It will be noted that the loss to consumers is greater than merely the money cost of the defective goods. The quantity of defective goods is $Q_0 - Q_1$ and their cost may be expressed either as $P_0(Q_0 - Q_1)$, i.e., area E, or as $Q_1(P* - P_0)$, i.e., area B. The decrease in consumer surplus is greater however, being areas B + D. Thus the loss to consumers is equal to the cost of the defective goods plus the dead weight loss. The first is a transfer to producers, but the second, the dead weight loss, is a loss to society.

Consumers may improve slightly on their position as shown in the last diagram. If the point P^*Q_1 is not on the demand curve, they may improve their welfare by moving to a new point, P^*Q^*, which is on the demand curve. Such a move implies awareness of the defect rate $(Q_0 - Q_1)/Q_0$ and the consequent implied higher price, P^*. After such a move, consumers would be purchasing $Q^*(Q_0/Q_1)$ items in total at a price of P_0 each, Q^* of them would be usable and the implied price per usable item would be P^*. Though this will be an improved position than if consumers did not adjust to the implied higher price, there will still be a dead-weight loss (though smaller than before) and the total loss to consumers will be greater than the money value of the defective goods, P_0Q^*.

CHAPTER IV

LITIGATION AND SETTLEMENT

1. Introduction

        Lawyers are concerned with the outcome of litigation;
economists consider why litigation occurs in the first place.   In
a world of perfect foresight civil litigation over contracts,
negligence liability, property rights and so on would be
irrational (assuming parties received no pleasure from the
process of litigation itself). In such an unreal world all
costs, judicial determinations and awards would be known with
certainty. Assuming no transactions costs to act as barriers·to
settlement, settlement rather than litigation would always
result. A transaction in a claim would result with both parties,
as is the case in any voluntary transaction, being the better off
for it. As Professor Atiyah has described it:[1]

        A settlement is a business bargain in which the
        plaintiff sells his claim ... for what he can
        get and the buyer buys for as little as he has
        to pay.

        Under perfect foresight, however, no bargaining would be
necessary: offer and acceptance prices would be identical.

        The unreal may act as a guide to the real, and this
perfect world suggests that litigation occurs primarily because
of uncertainty.

        The purpose of this chapter is to analyze the decision
process and criteria with respect to the choice between
settlement and litigation. After establishing a model to
describe the general case, we will consider whether consumer
disputes fit well within that model and if not, what
modifications are necessary. Considerably more mathematics will
be used in this chapter than in the previous chapters, but a
non-technical introduction will be presented.

----

1. P.S. Atiyah, Accidents, Compensation, and the Law (2nd ed.)
    (London: Weidenfeld and Nicolson, 1975), p. 277.

2. Factors Affecting the Choice Between Settlement and Litigation

(a) The Problem Generally

Consider a civil dispute between a potential plaintiff (or claimant) and a potential defendant. It is irrelevant whether the dispute lies in tort or contract. Two general solution possibilities exist: the dispute may be voluntarily settled between the parties (settlement), or the matter may be decided by a third-party-imposed solution (adjudication).2 Binding arbitration and judicial determination (litigation) are the most common examples of the second type of solution. In what follows we assume the choice to lie between settlement and litigation.

For the rational non-vindictive plaintiff there will be some minimum settlement offer from the defendant which is acceptable in exchange for his claim. To be acceptable, the offer must leave the plaintiff at least as well off as his expected net return after litigation would leave him. In determining his minimum acceptable settlement price, the plaintiff will have to take into account the probability of his success at trial, the likely award of damages if successful, the cost of bringing the action, the degree of indemnification for costs if successful, and the costs associated with the alternative to litigation, i.e. settlement costs. All of these factors will have to be assessed in the light of the plaintiff's attitude toward risk. He may be a risk averter, a risk preferrer, or one who is wholly indifferent to risk. Finally, should settlement not be reached and litigation ensue, the plaintiff will have to determine his optimal use of legal services in the contest.

The defendant faces a similar decision requirement. There will exist some maximum settlement offer to the plaintiff which, if paid, would leave the defendant as well off as he would be were he to pay the expected net damages and costs arising from litigation. In determining his maximum offer, the defendant must consider his assessment of the probability of the plaintiff succeeding, the likely damages award against him in the event of the plaintiff's success, his costs of defending the action, the proportion of the plaintiff's costs he will be required to bear in the event of the plaintiff's success, his costs of settling rather than litigating and his attitude to risk.

---

2. We exclude from consideration random decision-making processes, e.g., coin flipping, or alternatively view this as a form of voluntary settlement.

If the plaintiff's minimum settlement price is equal to or less than the defendant's maximum settlement price, the parties will settle out of court rather than litigate. If the plaintiff's price is less than the defendant's, then a contract zone will be established. In this case the exact settlement price will be determined by the relative bargaining strengths of the parties.[3] Any change in the factors affecting plaintiff's and defendant's minimum and maximum settlement prices which increases the plaintiff's price and/or decreases the defendant's price will increase the likelihood of litigation, while any change in the factors which lowers the plaintiff's price and/or raises the defendant's will increase the likelihood of settlement. Hence, it may be noted, actions generally to lower court costs and create easier access to judicial remedies may in some circumstances increase the rate of litigation and lower the settlement rate. More will be said of this in the context of consumer dispute resolution. We turn now to a consideration of the individual factors affecting plaintiff's and defendant's settlement prices.

(b)   Factors Affecting Settlement Prices

(i)   Attitudes Toward Risk

A more technical treatment of risk aversion will be presented later. For now we will simply say that one who is a risk averter will accept a smaller sum certain than a larger sum uncertain; or, if paying, would prefer to pay a larger sum certain than a smaller sum uncertain. In short, risk as to the final outcome is, for the risk averter, an undesirable feature which he would pay to avoid.

Consider for example the expected value to a person who is offered the opportunity to participate in a coin toss wherein he will receive $2 if the coin comes up heads and zero if the coin comes up tails. The expected value of the outcome is the sum of the possible outcomes, each weighted by the probability that that particular result will occur. If the coin is unbiased, there is a .5 probability of heads and a .5 probability of tails. The expected value of a single toss is thus $1. A risk neutral person would be willing to pay up to $1 for the opportunity to play the game, or alternatively, would be indifferent between receiving a sum certain of $1 or playing the game with its expected value prior to playing of $1.

---

3.   For a discussion of the notion of "bargaining strength", see J.G. Cross, The Economics of Bargaining (New York: Basic Books, 1969).

A risk averter however, though certain of the probability distribution of the outcomes, finds that the risk associated with lack of knowledge as to what the actual outcome of the toss will be is distasteful. Though the expected value of the outcome is $1, a risk averter would only pay a sum less than $1 to play the game or, alternatively, would prefer to receive a dollar with certainty rather than a 50 per cent chance of winning $2 and a 50 per cent chance of winning zero.

A risk preferrer is of course one who values the risk associated with the game and hence would pay more than $1 to play the game or, alternatively, would rather toss the coin than receive $1 certain.

In technical terms, a certain outcome has a zero variance while the equivalent expected value where there is more than one possible outcome has a positive variance. The greater the range of the possible outcomes around the expected value, the greater the variance. Thus a coin toss paying $1.50 on heads and $.50 on tails still has an expected value of $1, but has a smaller variance than does the $2 - zero game. The smaller the variance, the less the risk.

(ii) Probabilities of Success and Failure

An increase in the plaintiff's subjective evaluation of the probability of his success in litigation will increase the likelihood of litigation, while an increase in the defendant's subjective evaluation of the probability of the plaintiff's success will enhance the prospects for settlement. The greater the convergence of the probabilities, i.e., the greater the degree of agreement between the parties as to the outcome of litigation, the greater the likelihood of settlement.

The more relevant information is shared between the parties, the more will the parties' probability estimates converge. Hence, the process of pre-trial discovery, though justified in terms of the elimination of surprise at trial, may in this context be seen as a process contributing to settlement insofar as it tends to eliminate error in the evaluation of the parties' probabilities of success.

The clarity and applicability of the substantive law on the point in issue will affect the parties' probability estimates. The greater the clarity and specificity of the substantive law, the less the variance will be for each party between the subjective estimate of the probability of success and the "true" probability. Similarly, the more the specific fact situation can be said to come within the substantive legal rule, the less the errors in the probability evaluations and the

greater the convergence of the two estimates. Services provided by lawyers or paralegal personnel may be viewed as aiding the parties in clarifying the substantive law and in interpreting their fact situation in its context, thus encouraging settlement.

(iii) The Damages Award

If both parties agree on the likelihood of the plaintiff's success and both furthermore agree on the likely damage award, then, assuming neither party is a risk preferrer, the magnitude of the damage award is irrelevant and the parties will settle. If the award is agreed upon but the probabilities vary, then the likelihood of litigation is increased. A necessary though not sufficient condition for litigation in such a situation is that the plaintiff's estimate of his probability of success be greater than the defendant's estimate of the plaintiff's probability of success. This being the case, both plaintiff's and defendant's settlement prices will increase at the rate of their respective probabilities as the size of the damages award increases, and thus the plaintiff's minimum acceptable price will increase faster than the defendant's maximum offer. It may be noted that a situation where plaintiff's evaluation of the probability of his success is greater than the defendant's evaluation of the probability of the plaintiff's success is not a sufficient condition for litigation by itself, because of the existence of uncompensated costs of launching and pursuing the action.

Among risk averse parties, the greater the uncertainty associated with the likely damage award the greater the likelihood of settlement. A risk averse plaintiff will tend to reduce his minimum acceptable settlement price as the variance (or riskiness) associated with the expected damages award increases, while a risk averse defendant will increase his maximum offer as the variance of the expected award increases. Both changes will contribute to a greater likelihood of settlement.

The larger the claim in relation to the party's total wealth the greater will be the risk facing the party, for the greater will be the variance in that party's wealth, given success or failure. For a given likely award, if the plaintiff is the poorer party a settlement is more likely than if the defendant is the poorer. In the general case there is no reason to expect plaintiffs to be poorer than defendants or vice versa. In consumer dispute cases, however, there are obvious reasons for assuming consumer plaintiffs to have less net wealth than business defendants. It is also not unreasonable to posit that consumers as a class are more risk averse than businesses.

The greater the time period necessary to complete a court action, the greater the likelihood of settlement. The longer the time period before an award is received or paid, the less will be its present value, assuming both parties have a positive discount rate. As suggested above, the smaller the award, the greater the likelihood of settlement. Delay will have a more pronounced effect in that class of cases where plaintiffs have higher discount rates than defendants, for the present value of the plaintiff's award will be deteriorating as delay increases at a faster rate than the present value of the payment required in the future from the defendant. It is submitted that personal injury cases (injured plaintiff versus insurance company) and consumer disputes (individual versus business) may fall within this class.

(iv)   Litigation Costs

The greater the costs of litigation, the greater the likelihood of settlement. Litigation costs are taken here to include not only explicit fees for court and legal services, but also the implicit cost of time and energy expended in pursuit of the action. Plaintiff and defendant need not and are not likely to face identical costs. Nevertheless, any increase in litigation costs to the plaintiff will lower his minimum settlement price while an increase in litigation costs facing the defendant will raise his. Either or both effects will contribute to a higher settlement rate and this conclusion is not affected by partial or even complete indemnification of the successful party's costs by the unsuccessful party.

(v)   Settlement Costs

Just as there are costs associated with litigation, there are also likely to be costs involved in the process of arriving at a settlement between the parties. Again, these costs are taken to include both pecuniary and non-pecuniary elements. It appears safe to assume that in most civil disputes, settlement costs are less than litigation costs. An increase in settlement costs will raise the plaintiff's minimum settlement price and lower the defendant's maximum offer, thus leading to an increased likelihood of litigation.

(vi)   Indemnification for Costs

Normally there will be an award of costs after the trial decision is rendered and in most cases costs will follow the event, i.e., the losing party will be called upon to recompense the victor for part of his costs in the action. This is of

course in addition to the loser's own costs. It is assumed in all cases that the award of costs will not fully indemnify the victorious party, even in the rare case of costs being awarded on the scale of "costs between a solicitor and his own client", for in no case will there be an attempt to indemnify for time, trouble and anxiety costs.

An indemnification rule leaves the successful party better off and the unsuccessful party worse off than each would be without such a rule. In so increasing the variance of the expected outcome from litigation, the rule operates to encourage settlement among risk-averse plaintiffs and defendants. The rule thus performs not only its conventional function of discouraging idle litigation but also acts to influence the choice between litigation and settlement in favour of the latter. In the face of two risk-neutral parties it is not possible to state in the general case whether the rule raises or lowers the settlement offer prices. Only in the special case of equal expected costs and equal subjective probabilities can it be said that both parties' prices will change by equivalent amounts, with each price rising if both agree that the plaintiff is more likely to win, and both prices decreasing if both agree that the plaintiff is more likely to lose.

It has been pointed out that another consequence of the indemnification rule is to reward the party that correctly predicts the outcome of litigation and penalize the party that fails to predict the outcome. This will lead to greater incentives to correctly predict the probabilities of success and failure, a greater convergence of the probability estimates by the parties, and hence to a higher settlement rate.[4]

(vii)   Use of Legal Resources

We have so far treated each party's probability estimate as an exogenous determination. In fact the probabilities, and possibly the damage award as well should the matter proceed to trial, are likely to be in part functions of the quantity of legal services employed. The rational party will thus purchase legal services as long as the marginal contribution to his expected trial outcome from the last unit of legal services exceeds the extra cost of that unit. Each unit of legal services purchased may be taken to increase the probability of that

---

4.   R.A. Posner, "An Economic Approach to Legal Procedures and Judicial Administration", _Journal of Legal Studies_ (2:1973), p. 399.

party's success and/or to vary the expected damage award in that party's favour. As with most productive inputs, however, diminishing marginal productivity may be assumed such that each legal resource unit adds less to the expected value of the trial outcome (or, in the case of the defendant, reduces the expected value of the trial outcome by less) than did the previous unit. Assuming the marginal cost per unit of legal resources is not decreasing, an equilibrium level of resource use may be reached at which marginal benefit equals marginal cost.

The larger the expected award, the greater will be the productivity per unit of legal resources. Put simply, this suggests that the greater the expected award, the more each party will spend on legal resources.

The analysis above assumes each party determined its legal resource use independent of the other party. If this is so, each party may reach its optimal resource use position as described. Problems similar to those involved in duopoly pricing arise however if (a) each party bases its expenditure on legal resources in part upon what the other party does, and (b) the productivities of each party's legal resources are interdependent – i.e., the productivity of one party's resources is in part dependant upon the quantity of legal resources employed by the other party. Optimal resource use for each party will then depend on the other party's observed or predicted behaviour. There may well not be an equilibrium solution for the parties' resource use in such a situation.

If there is no way of determining each party's optimal resource use under conditions of interdependency however, then the subjective probabilities will also be indeterminate and hence it will be impossible to solve for the precise settlement condition in terms of offer and acceptance problems. In what follows, we shall assume simply that a party's expenditures on legal resources are not a function of the other party's expenditure.

3. <u>A Formal Settlement – Litigation Model</u>[5]

(a) <u>Wealth States</u>

Judicial determination of a dispute will alter the wealth holdings of each party. The magnitude of the change in each party's wealth will depend upon the damages awarded, the quantity and unit cost of legal resources employed by each, and the costs indemnification rule in the jurisdiction. Let W_p and W_d be the plaintiff's and defendant's predetermination wealth states, D be the damage award, r the cost per unit of legal services (assumed the same for each party), and R_p and R_d the quantities of legal resources employed by plaintiff and defendant. With respect to indemnification, we assume that the successful party will be reimbursed for some fraction, z, of his legal resource costs ($0 \leq z \leq 1$). Thus, if the plaintiff is successful, he will receive zrR_p from the defendant, while if he is unsuccessful he must pay zrR_d to the defendant in addition to paying his own costs.

Four different wealth states may thus be delineated, two for each party, one for successful and the other being for unsuccessful litigation. Let W_{pn} be the wealth state of the plaintiff if he wins and W_{pl} if he loses, and similarly W_{dn} and W_{dl} for the defendant winning or losing. The wealth states are thus:

(i) plaintiff wins:

$$W_{pn} = W_p + D - rR_p + zrR_p$$

(ii) plaintiff loses:

$$W_{pl} = W_p - rR_p - zrR_d$$

(iii) defendant wins:

$$W_{dn} = W_d - rR_d + zrR_d$$

(iv) defendant loses:

$$W_{dl} = W_d - D - rR_d - zrR_p$$

---

5. This section is built in part upon the work of Landes, Gould and Posner. See W.M. Landes, "An Economic Analysis of the Courts", <u>Journal of Law and Economics</u> (14:1971), p. 61; J.P. Gould, "The Economics of Legal Conflicts", <u>Journal of Legal Studies</u> (2:1973), p. 279; and R. A. Posner, "An Economic Approach to Legal Procedures and Judicial Administration", <u>Journal of Legal Studies</u> (2:1973), p. 399.

b) Expenditure on Legal Resources

Let the plaintiff's subjective probability of his success be P_p and the defendant's subjective probability of the plaintiff's success be P_d. These probability estimates will not, however, be independent of the quantity of legal services purchased by each party. We may assume that the greater the quantity of services purchased by the plaintiff the greater will be <u>both</u> P_p and P_d while the greater the quantity purchased by the defendant the lower will be both P_p and P_d. Hence we may write

$$P_p = P_p(R_p, R_d; Z_p)$$

$$P_d = P_d(R_p, R_d; Z_d)$$

where Z_p and Z_d are factors other than quantities of resources which influence P_p and P_d. Furthermore

$$\frac{\partial P_p}{\partial R_p}, \frac{\partial P_d}{\partial R_p} > 0; \text{ and } \frac{\partial P_p}{\partial R_d}, \frac{\partial P_d}{\partial R_d} < 0.$$

We may expect the quantum of damages awarded also to be a function of the quantity of legal resources employed - a positive function of R_p and a negative function of R_d. Hence, we write

$$D = D(R_p, R_d; T)$$

where T merely represents other factors affecting damages, and

$$\frac{\partial D}{\partial R_p} > 0; \quad \frac{\partial D}{\partial R_d} < 0.$$

Finally, indemnification of costs to the successful party will increase linearly and proportionately with that party's expenditures for legal resources. Thus, if I is the amount of indemnification paid by the losing party to the winning party,

$$I = zrR_p, \quad \frac{\partial I}{\partial R_p} = zr$$

if the plaintiff is successful, and

$$I = zrR_d, \quad \frac{\partial I}{\partial R_d} = zr$$

if the defendant is successful.

(i) The Plaintiff's Optimal Expenditure on Legal Services

For the plaintiff, there is a probability P_p of being left with W_{pn} and a consequent probability of $1 - P_p$ of being left with W_{pl}. Though both W_{pn} and W_{pl} are expressed in money terms, they may not indicate what their "value" is to the plaintiff in terms of some sense of satisfaction, well being or utility. Thus, to use a familiar example, an additional dollar of wealth for a poor person may have far greater utility than an additional dollar for that person were he very rich. To indicate the utility of the possible wealth states facing the plaintiff, we write $U_p(W_{pn})$ and $U_p(W_{pl})$. Given the probability of each wealth state arising and hence the utility associated with that wealth state arising, the expected utility of litigation (before the fact) may be written

$$E(U_p) = P_p U_p(W_{pn}) + (1 - P_p) U_p(W_{pl}).$$

The plaintiff is assumed to allocate expenditures to the purchase of legal services so as to maximize the expected utility of litigation. Assuming diminishing marginal productivity of R_p with respect to raising P_p and D, a point must be reached at which the diminishing (though positive) marginal benefit of R_p is just equal to the marginal cost of R_p. At this point $E(U_p)$, the expected utility from litigation, will be maximized. Mathematically, we maximize $E(U_p)$ with respect to R_p. Thus,

$$\frac{\partial E(U_p)}{\partial R_p} = P_p U_p'(W_{pn}) \cdot \frac{\partial W_{pn}}{\partial R_p} + U_p(W_{pn}) \cdot \frac{\partial P_p}{\partial R_p}$$

$$+ (1-P_p) U_p'(W_{pl}) \cdot \frac{\partial W_{pl}}{\partial R_p} - U_p(W_{pl}) \cdot \frac{\partial P_p}{\partial R_p} = 0.$$

From the definitions of W_{pn} and W_{pl} it is observed that

$$\frac{\partial W_{pn}}{\partial R_p} = \frac{\partial D}{\partial R_p} - r + zr$$

and

$$\frac{W_{pl}}{R_p} = -r$$

and hence, the optimal allocation of resources to legal services occurs when

$$\frac{\partial P_p}{\partial R_p}[U_p(W_{pn}) - U_p(W_{p1})] + \frac{\partial D}{\partial R_p} \cdot P_p U_p'(W_{pn})$$

$$= r[(1-z) P_p U_p'(W_{pn}) + (1-P_p)U_p'(W_{p1})]$$

where

$$\frac{P}{R_p}\{U_p(W_{pn}) - U_p(W_{p1})\}$$

is the marginal benefit of increasing the probability of winning by the employment of additional legal resources;

$$\frac{\partial D}{\partial R_p} \cdot P_p U_p'(W_{pn})$$

is the marginal benefit of increasing the expected damages award by the employment of additional legal resources;

$$r(1-z)P_p U_p'(W_{pn})$$

is the marginal cost given the plaintiff's success, times the probability of his success; and

$$r(1-P_p)U_p'(W_{p1})$$

is the marginal cost given the plaintiff's loss, times the probability of his loss.

The left handside of the expression thus represents the marginal return or benefit to expenditures on legal services and is positive, while the right-hand side of the expression represents the marginal cost of legal services and is also positive. On the assumption of diminishing marginal productivity of R_p, both $\partial P_p/\partial R_p$ and $\partial D/\partial R_p$ decline as R_p increases, and hence the marginal benefit schedule is declining as in Diagram 14.

DIAGRAM 14

marginal cost
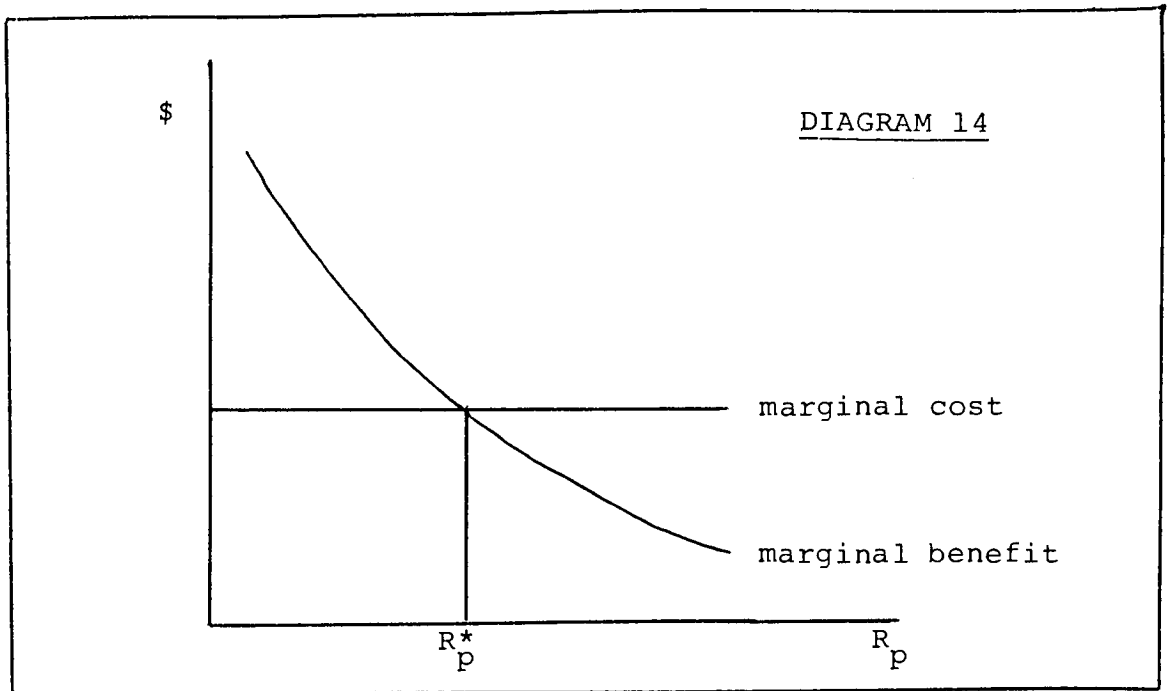
marginal benefit

It may be noted that a change in either the initial wealth endowment of the plaintiff, W_p, or the scale of damages, D, will shift the marginal benefit and cost schedules.

To determine the consequences of a change in the intitial wealth of the plaintiff, we differentiate the marginal benefit and marginal cost expressions, previously obtained, with respect to W_p. Thus,

$$\frac{\partial MB}{\partial W_p} \qquad \frac{\partial P_p}{\partial R_p} \{U_p'(W_{pn}) - U_p'(W_{p1})\} \qquad\qquad (<0)$$

$$+ \quad \frac{\partial D}{\partial R_p} \cdot P_p U_p''(W_{pn}) \qquad\qquad (<0)$$

and

$$\frac{\partial MC}{\partial W_p} = r\{(1-Z)P_p U_p''(W_{pn}) + (1-P_p)U_p''(W_{p1})\} \qquad\qquad (<0)$$

Hence, both the marginal cost and marginal benefit schedules shift down, and no unequivocal solution is possible with respect to the change in the use of legal resources as the wealth state changes.

The procedure is repeated to determine the effects on the MB and MC schedules of an increase in the "stakes", i.e., an increase in D.

$$\frac{\partial MB}{\partial D} = \frac{\partial P}{\partial R_p} \cdot U_p'(W_{pn}) \qquad\qquad (>0)$$

$$+ \frac{\partial D}{\partial R_p} \cdot P_p U_p''(W_{pn}) \qquad\qquad (<0)$$

and

$$\frac{\partial MC}{\partial D} = r\{(1-Z)P_p U_p''(W_{pn})\} \qquad\qquad (<0)$$

In this case, as with changes in W_p, an increase in D will shift the marginal cost curve down. The effect on the marginal benefit schedule is indeterminate, being capable of shifting in either direction. Thus, though this analysis cannot give us definite answers with respect to the use of legal resources when the initial wealth endowment grows or when the scale of damages grows, we can say that there is a greater likelihood of increased resource use in the latter than in the former case.

(ii)   The Defendant's Optimal Expenditure on Legal Services

The analysis of the defendant's optimal expenditure on legal services parallels that for the plaintiff and will not be presented in detail. The defendant's utility will be maximized at the point where

$$\frac{\partial P_d}{\partial R_d}\{U_d(W_{dl}) - U_d(W_{dn})\} - \frac{\partial D}{\partial R_d} \cdot P_d U_d'(W_{dl})$$

$$= r\{P_d U_d'(W_{dl}) + (1-Z)(1-P_d)U_d'(W_{dn})\}$$

Here again, the left-hand side of the expression represents the marginal benefit of expenditure on legal services while the right-hand side represents the marginal cost. As with the plaintiff, the defendant's expenditures will be determined by the productivity of legal services in lowering the probability of the plaintiff's success (P_d) and their productivity in lowering the potential damages award.

(c) Risk Aversion

Risk aversion was briefly referred to earlier and will here be elaborated upon somewhat. An individual is risk averse if he prefers a riskless payment of $X over an equivalent expected value of $X involving risk. Thus, for example, a risk averse person would choose $99 certain over the opportunity to participate in a venture with a 10 per cent chance of receiving $900 and a 90 per cent chance of receiving $10. The expected value of the return from the venture is $99, yet the risk with respect to the outcome makes it less attractive than the same sum certain. For a risk neutral person, there would be complete indifference between the opportunities, while a risk preferrer would prefer the venture over the sum certain.

The expected value of a risky outcome for which the probabilities are known is the value of the sum of the outcomes, each weighted by the probability of its occurrence, or

$$E[0] = P0_1 + (1 - P)0_2$$

The expected utility of a risky venture may be written as the sum of the utilities from each outcome, each weighted by the probability of its occurrence, or

$$PU(0_1) + (1 - P)U(0_2)$$

If a person is risk averse, he will prefer the sum certain equal to the expected value, rather than the risk venture with the same expected value. Hence, for such an individual

$$U(P0_1 + (1 - P)0_2) > PU(0_1) + (1 - P)U(0_2)$$

This inequality is the definition of a concave function of the type shown in Diagram 15.[6]

---

6. J.P. Gould, "The Economics of Legal Conflict", Journal of Legal Studies (2:1973), p. 281.

DIAGRAM 15

U(O)

O
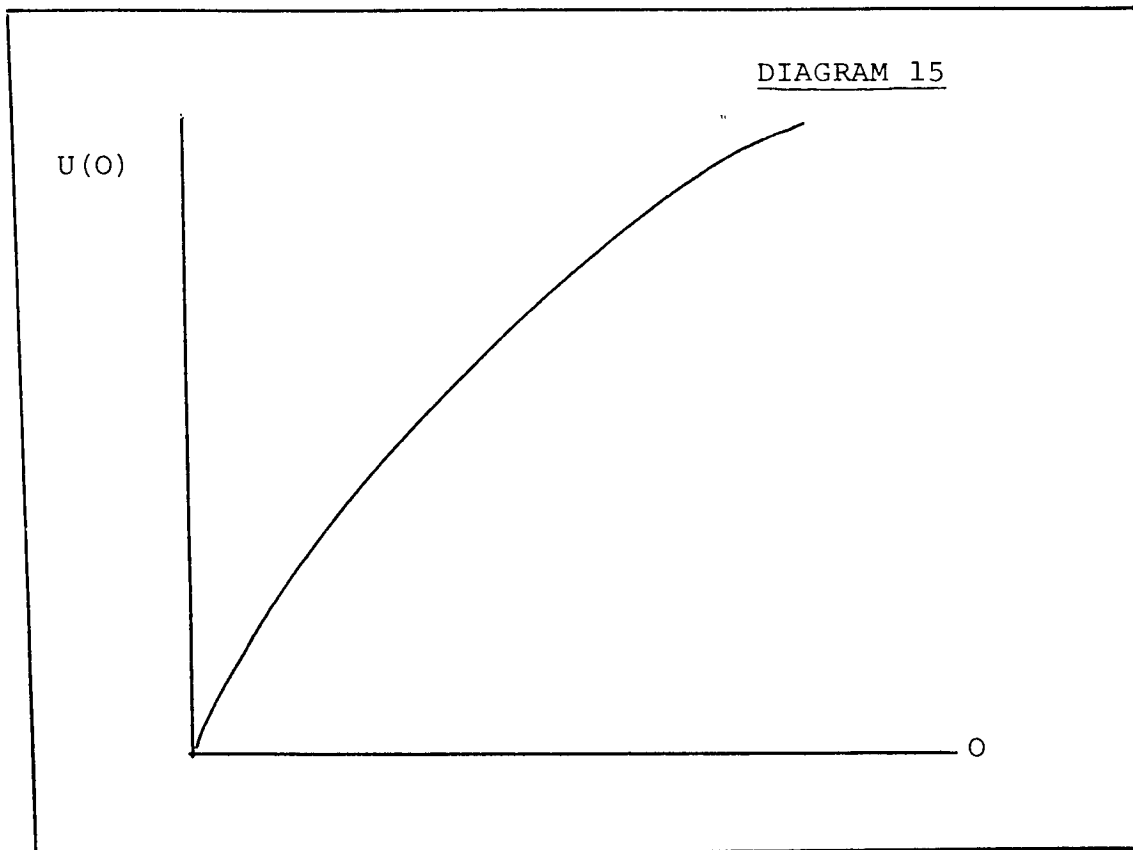
But this function reflects diminishing marginal utility with respect to 0, or $U''(0) < 0$. Hence, if we consider the outcomes to be different wealth states, one who displays diminishing marginal utility with respect to wealth is risk averse, and one who is risk averse will display diminishing marginal utility with respect to outcomes affecting wealth.

As stated, one who is risk averse will prefer a sum certain to an equivalent expected value from a risky venture. It follows then that there exists a smaller sum certain at which the individual will be indifferent between that sum certain and the expected value of the risky venture. If the outcomes are defined as different wealth states, and W_1 and W_2 are the possible wealth states arising from the risky venture, and W_0 is the smaller sum certain at which the risk averse individual is indifferent between that sum and the expected value of the venture, then

$$U(W_0) = PU(W_1) + (1 - P)U(W_2)$$

and

$$W_0 < PW_1 + (1 - P)W_2$$

It follows then that one who is risk averse is willing to pay a positive sum in order to avoid risk, the sum being

$$PW_1 + (1 - P)W_2 - W_0$$

Litigation is a risky venture for both the potential plaintiff and the potential defendant. If both parties are risk averse, then by definition each will be willing to pay a price in order to avoid the risk, the maximum sum for the plaintiff being the difference between the expected value, given his evaluation of the prospects, arising from the litigation (i.e. his expected monetary utility) and the lower sum certain yielding the same utility.

If both parties are risk averters, and if both share the same probability estimates and the same predictions as to the magnitude of damages, then clearly a settlement is possible by which both parties will be better off than if they go to trial. The presence of court costs may provide a further incentive to seek settlement rather than trial, but it is not a necessary condition for settlement. As long as one party is risk averse and the other is not a risk preferrer, the prospect of settlement is present.

Before turning to the analysis of the conditions for settlement, we inquire as to the determinants of the degree of risk aversion, or in terms of the previous discussion, the determinants for a risk averse individual of the premium he would be willing to pay to avoid the risk associated with the venture. Essentially this represents an inquiry into the nature of an individual's utility function where the arguments in the function are outcomes or wealth states and risk to be borne. Our concern is with the variation of utility with risk.

Consider the possible outcomes from a trial. Assume of course that there is a probability of a party being successful. Assume also that the damages award is unknown although the probability structure describing the likelihood of different damage outcomes is known to the party. The possible awards and the probability of their occurring are shown by curve I in Diagram 16.

Probability of
particular award
occurring

DIAGRAM 16

II

I

Possible
Damage
Awards
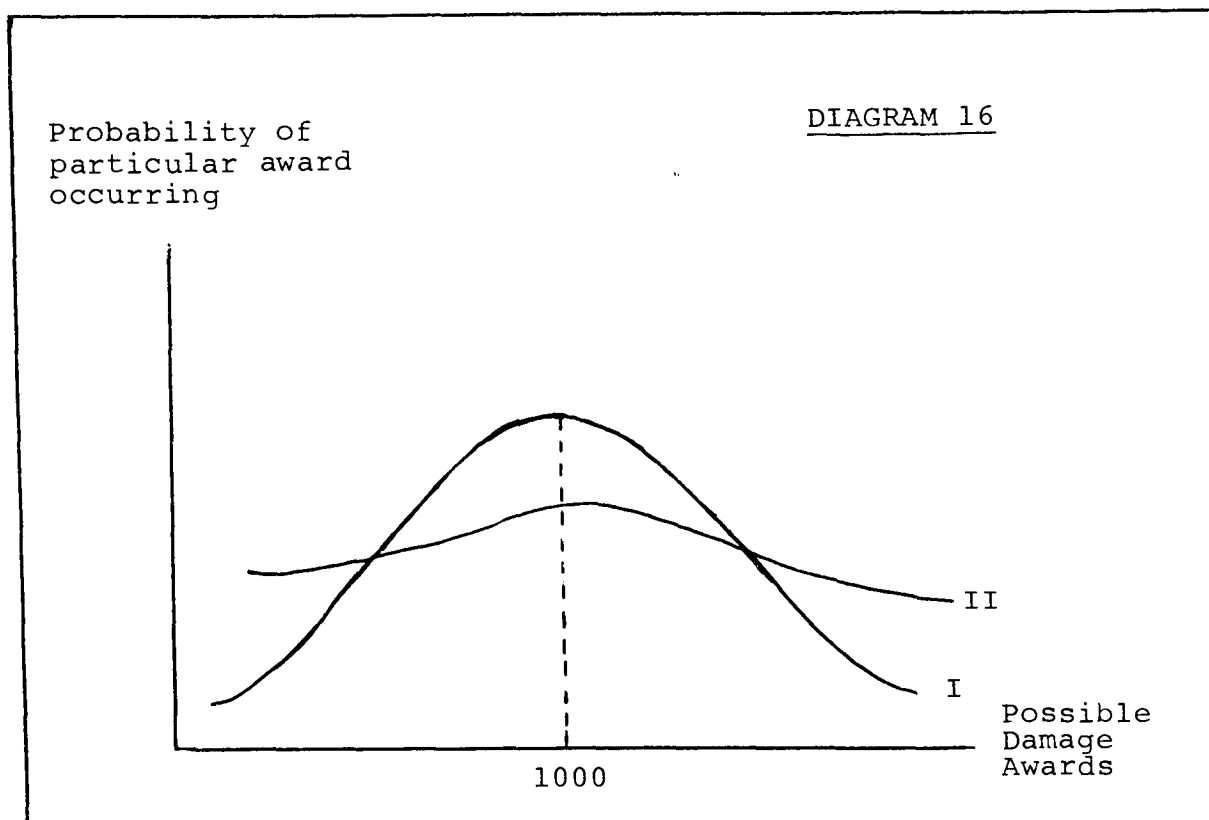
1000

Here, the expected value of the award is $1000 but this
is only the mean or the "most likely" award, a variety of other
awards being possible. The probability of a particular award
occurring is shown on the vertical axis. The probability of an
award diminishes the greater the difference, positive or
negative, between it and $1000. Though the mean or expected
value of the award is $1000, following our analysis of risk
aversion a plaintiff would be willing to accept a smaller sum
certain to avoid the risk associated with the actual outcome.
Similarly a defendant would be willing to pay in excess of the
mean or expected value to avoid the risk.

Consider now the distribution of possible awards shown by
curve II. The mean or expected value is still $1000. In this
case however the distribution is much more "spread out" with
higher and lower values more probable than with the distribution
of curve I. It is submitted that the more spread out the
distribution, the greater the risk associated with it. In the
least risk case there would be a probability of one of an award
of $1000 arising, there would be no variation around this mean,
and consequently no associated risk - it would be in effect a sum
certain of $1000. As the distribution flattens out, the risk
grows. As the risk grows, however, the premium that each party
is willing to pay to avoid risk presumably grows, the plaintiff
being willing to accept a smaller sum certain as the risk grows,

and the defendant to pay a greater sum certain as the risk grows. Thus, the greater the risk of litigation, in the sense of a greater variation of possible outcomes, the more likely the parties will "pay" to avoid litigation.

(d)   Settlement Prices: Conditions for Settlement

(i) The Plaintiff's Minimum Settlement Price

Whether he goes to litigation or accepts an offer in settlement from the defendant, the plaintiff's wealth state will vary from its original value. If the settlement sum accepted is S, then, abstracting from any costs associated with reaching a settlement (which costs are quite distinct from court costs), after settlement the plaintiff's wealth state will be:

$$W_{sp} = W_p + S$$

The rational plaintiff will be willing to accept any sum which leaves him as well off or better off than he would be if he litigated. His expected utility from litigation was shown to be

$$E(U_p) = P_p U_p(W_{pn}) + (1 - P_p)U_p(W_{pl})$$

It follows that his minimum acceptable settlement price must leave him no worse off than would litigation. If the plaintiff's minimum settlement price is denoted as S_p, then his wealth state after settlement will be $W_{sp} = W_p + S_p$ and the utility associated with his wealth state must, by definition of S_p, be equal to the utility of litigation. Hence,

$$U(W_p + S_p) = P_p U_p(W_{pn}) + (1 - P_p)U_p(W_{pl})$$

and following our discussion of risk aversion, if the plaintiff is risk averse, then

$$W_p + S_p < P_p W_{pn} + (1 - P_p)W_{pl}$$

and the more risk averse the greater the inequality.

If we consider the limiting case of a risk neutral plaintiff, then

$$W_p + S_p = P_p W_{pn} + (1 - P_p)W_{pl}$$

Substituting $W_p + D - rR_p + zrR_p$ for W_{pn} and $W_p - rR_p - zrR_d$ for W_{pl}, and solving for S_p, we obtain

$$S_p = P_p[D + zr(R_p + R_d)] - rR_p - zrR_d$$

(ii) The Defendant's Maximum Settlement Price

A similar analysis may be conducted to determine the maximum amount the defendant will be willing to offer to settle the plaintiff's claim before turning to litigation. On settlement, the defendant's wealth state will be $W_{sd} = W_d - S$. The maximum S for the defendant will be that which leaves him indifferent between settlement and trial of the action. Denoting this maximum S_d, it can be solved from the utility equivalence expression.

$$U(W_d - S_d) = P_d U(W_{dl}) + (1 - P_d)U(W_{dn})$$

and for a risk neutral defendant will be

$$S_d = P_d[D + zr(R_p + R_d)] + rR_d - zrR$$

(iii) The Condition for Settlement

For settlement to take place, the plaintiff's minimum settlement price must be less than or equal to the defendant's maximum settlement price:

$$S_d \geq S_p$$

or,

$$P_d[D + zr(R_p + R_d)] + rR_d \geq P_p[D + zr(R_p + R_d)] - rR_p$$

Consider first the situation wherein $P_d = P_p = 1$, i.e., both parties are completely certain that the plaintiff will succeed at trial for an award of D and partial cost indemnification. In this case, $S_p = D - (1 - z)rR_p$, which is less than D by the amount of the plaintiff's cost for which he is not indemnified, i.e., $(1 - z)rR_p$. Thus, though completely certain of his success and his award, the plaintiff still stands to benefit via settlement if some part of his costs are uncompensated. Similarly, the defendant, though certain of the plaintiff's success at trial, will be willing to offer an amount greater than D to avoid trial, i.e., $S_d = D + rR_d + zrR_p$. Here the maximum premium above D is composed of the

defendant's full costs, plus the portion of the plaintiff's costs which the defendant will be called upon to pay. Any settlement price which falls between S_p and S_d will leave both parties better off than if they had proceeded to trial.

If both parties agree upon D, then a sufficient (though not necessary) condition for settlement is $P_d \geq P_p$, i.e., that the defendant's assessment of the likelihood of the plaintiff's success not be less than the plaintiff's assessment of that probability. Should the probabilities be the same, the plaintiff's minimum price will still be below the defendant's maximum offer by the amount $rR_d + rR_p$, and we may expect settlement to occur.

Even should P_p be greater than P_d, it is still possible for settlement to occur, though the gap between S_d and S_p will be smaller. Settlement is still possible because of the existence of the cost elements rR_p and rR_d. Eventually of course a sufficient differential between P_p and P_d will eliminate the settlement range and litigation will become inevitable.

We have assumed the limiting condition of risk neutrality in this section so far. If both parties are risk averse then the range in which settlement may take place will be greater, and the greater the risk aversion, the greater the effects upon S_p and S_d. Risk preference, on the other hand, narrows the settlement range and if even one party is a risk preferrer to a sufficient degree, then even though the parties may have the same predictions of the plaintiff's likelihood of success and the same prediction as to D, settlement may be impossible.

(iv) The Costs of Settlement

We have specified the costs associated with trial, but have so far assumed that the transaction costs involved in reaching a settlement are zero. This of course need not and generally will not be the case. If C_p and C_d are the plaintiff's and defendant's costs respectively in reaching a settlement, then their wealth states after settlement will be

$$W_{ps} = W_p + S - C_p$$

$$W_{ds} = W_d - S - C_d$$

This in turn will affect their settlement prices, which will now be

$$S_p = P_p[D + zr(R_p + R_d)] - rR_p - rzR_d + C_p$$

$$S_d = P_d[D + zr(R_p + R_d)] + RR_d - rzR_d - C_d$$

Thus the existence of positive settlement costs will raise the minimum acceptable price for the plaintiff, and lower the maximum settlement price for the defendant. The effect of each price change is therefore to make settlement less likely.

Once settlement costs are introduced, $P_d \geq P_p$ is no longer a sufficient condition for settlement. To see this, we may again write out the full condition for settlement, $S_d \geq S_p$ and after rearranging terms this emerges as

$$(P_d - P_p)[D + zr(R_p + R_d)] \geq -r(R_p + R_d) + (C_p + C_d)$$

Now, $P_d \geq P_p$ is only a sufficient condition for settlement if $r(R_p + R_d) \geq (C_p + C_d)$, i.e., if total trial costs exceed total settlement costs. Generally, settlement costs work to encourage litigation.

Assuming joint settlement costs are less than trial costs, then an out-of-court settlement is Pareto-optimal under all conditions regardless of probability estimates, wealth states, degree of risk aversion and so on. Settlement involves a resource saving to society of $r(R_p + R_d) - (C_p + C_d)$, regardless of the relative bargaining strengths of the parties. How this saving is divided between the parties is of course likely to be related both to bargaining skills and to the nature of the estimated probabilities, but even if one party captures the entire saving, the result will be Pareto-superior to litigation.

(v)   When Will the Plaintiff Quit?

The analysis so far may appear to suggest that even if both P_p and P_d are very low, as long as they are equal or $P_d > P_p$, a settlement is possible. This is not necessarily the case, however. To the two possible courses of action so far considered, settlement or litigation, must now be added a third, the renunciation of the entire claim by the plaintiff.

For a given D, and for given expenditures on legal resources by the plaintiff and defendant, there must exist some probability of success greater than zero below which it is no longer worthwhile for the plaintiff to pursue the action. The minimum probability is that at which the expected value of the trial outcome leaves the plaintiff no better off than he was prior to commencing the action.

If W_p is the wealth state of the plaintiff prior to commencing the action, and $E(U_p)$ is the expected value of the trial outcome, then the claim will be renounced at the point where $E(U_p) = U(W_p)$, or

$$P_p W_{pn} + (1 - P_p)W_{pl} = W_p$$

for the risk neutral plaintiff. Substituting for W_{pn} and W_{pl}, and solving for P_p, we find the minimum P_p, P_p^{min} at or below which the claim is not worth pursuing,

$$P_p^{min} = \frac{rR_p + zrR_d}{D + zrR_p + zrR_d}$$

At this point, if the plaintiff's self-evaluated probability of success equals P_p^{min}, his settlement price will equal zero, i.e., he will accept a "price" of zero to renounce his claim.

A necessary condition for the claim to be pursued is clearly that the minimum "quit" probability be less than one, which means that we must have

$$D + zrR_p > rR_p$$

or,

$$D > (1 - z)rR_p$$

Thus a minimum condition for any claim to be pursued by a plaintiff is that the actual judgment expected, D (not to be confused with the expected value of the judgment, P_pD), must be greater than that element of the litigation costs which will remain uncompensated in spite of success.

The reader is reminded that for convenience we have been working under the assumption of a risk-neutral plaintiff. If we assume that the plaintiff is, as seems likely, risk averse, then for the plaintiff to be neither better off nor worse off after litigation than before litigation, we must write

$$P_p W_{pn} + (1 - P_p)W_{pl} = W_p + e$$

where e is a risk premium reflecting the fact that for a risk averter, to be exactly as well off in terms of utility after risky litigation as before requires some positive monetary gain to compensate for the risk of litigation. Solving this expression for P_p^{min}, we obtain for a risk averter

$$P_p^{min} = \frac{rR_p + zrR_d + e}{D + zrR_p + zrR_d}$$

Thus the greater the degree of risk aversion, the greater the likelihood the claim will be renounced.

The minimum "quit" probability will vary inversely with D, i.e., $\partial P_p^{min}/\partial D < 0$. The larger the "true" claim, i.e., that which is expected on success rather than that which is merely claimed, the less important do trial costs become and the lower will be the "quit" probability. A very large claim may thus be worth pursuing even though the probability of success is slight.

As the size of the claim shrinks, the "quit" probability rises. This may have strong implications for consumer claims and we will shortly return to this theme.

The "quit" probability can be shown to vary directly with both R_p and r (i.e., $\partial P_p^{min}/\partial R_p > 0$; $\partial P_p^{min}/\partial r > 0$). It was shown earlier that legal resources will be utilized up to the point at which their marginal benefit equals their marginal cost. This point provides the optimal resource use given that the plaintiff continues with the action. It may, however, yield a least loss position rather than maximum gain. In such a situation the plaintiff may be expected to drop the claim and achieve zero rather than a loss in utility. This then is what we may expect when, given the optimal quantity of resources, R_p, the derived "quit" probability is greater than the plaintiff's evaluation of the probability of his success.

Finally, we turn to the question of how P_p^{min} varies with z, the proportion of costs for which the successful party is

indemnified. A priori, the relationship between P_p^{min} and z

is ambiguous,[7] but the nature of z and of R_p, may merit

greater attention at this point. We have so far assumed R_p
to be legal resources in the strict sense, and z to be that
proportion of the cost of these resources that the successful
party would receive if "costs followed the cause". If however,
we adapt our notion of R_p to include "time, trouble, energy
and anxiety" costs, and if, as is the case, these are not
normally compensated for on success in the action, then the true
z facing the plaintiff with respect to his own total costs, will
be lower than that suggested by the "legal" z arising from the
costs award. And this true z facing the plaintiff, z_p, will
vary with the input of non-legal resources, while the z relating
to the defendant's costs remains unchanged.

It will therefore prove of some assistance when writing

the expression for P_p^{min} to distinguish, for the plaintiff,

between z_p and z_d, the latter only being determined by

the decision with respect to costs, the former being determined
by the costs decision plus the plaintiff's input of non-legal
resources. Thus,

$$P_p^{min} = \frac{rR_p + z_d rR_d}{D + z_p rR_p + z_d rR_d}$$

and P_p^{min} will vary inversely with z_p, meaning that for

a given R_p, the greater is the proportion of uncompensated

non-legal cost elements, the greater will be the "quit"

probability, P_p^{min}. Again, this appears highly relevant if

the plaintiff is a consumer.

We close this section merely by noting again that a
necessary condition for the claim to be pursued is that P_p^{min}
be less than one, implying $D > (1 - z_p)rR_p$, and this will
clearly be more difficult to meet the greater is $1 - z_p$, the

---

7. The two will vary directly or inversely depending upon
   whether DR_d is greater or less than $rR_p(P_p + R_d)$,
   and this relationship does not appear susceptible to
   economic/legal interpretation.

proportion of total resources expended by the plaintiff for which compensation is not provided if the plaintiff is successful.

(vi) Summary of Factors Affecting the Settlement-Litigation Decision

By definition, if the plaintiff's minimum settlement price is less than the defendant's maximum settlement price there is scope for a mutually advantageous bargain: the sale of the plaintiff's legal claim against the defendant. The elements that enter each party's settlement price are each party's estimate of the probability for success of the claim, the size of the expected judgment if the plaintiff is successful, the resources to be employed in litigation by each party and the unit resource price, and the costs of reaching an out-of-court settlement.

To the extent that the parties have access to a common stock of information, the greater will be the likelihood of settlement. Common information should lead to a convergence of opinions with respect to each party's probability estimate and judgment estimate. In the extreme case of equality of opinion, only a situation wherein settlement costs are greater than trial costs will prevent settlement. It follows that any legal or social institution which helps eliminate error, i.e., which aids the parties in moving to the "correct" probability, aids settlement.

Trial costs encourage settlement, settlement costs discourage it. It should be borne in mind however that "settlement" may include renunciation of the claim completely. Trial costs may be so large, particularly in relation to the size of the award expected upon success, that the plaintiff's settlement price becomes zero or negative, i.e., the claim is given up for nothing. To the extent that a part of the plaintiff's total litigation costs are uncompensated, there is a further likelihood of "settlement" at a zero price. We note here that normal settlement costs are not likely to arise in the special settlement case of claim renunciation.

To the extent that the parties are risk averse, there will be a greater likelihood of settlement. Again, however, risk aversion by the plaintiff may work to lower the plaintiff's settlement price to such an extent that the claim is simply renounced.

If the defendant's estimate of the plaintfff's probability of success is greater than the plaintiff's estimate, then the greater the award if the plaintiff is successful, the greater will be the likelihood of settlement. If the probability

relationship is the other way around, then the greater the
predicted award, the greater will be the likelihood of
litigation. Generally, however, the greater the predicted award,
the more likely is the claim to be pursued in some fashion
whether by litigation or by settlement at a non-zero price.

(vii)   Two Institutional Influences on the Factors

Reference has been made to the idea that settlement is
more likely when both parties are operating with the same or
similar information. Two institutional characteristics may be
mentioned which contribute to the sharing of information.

The first of these is the process of information exchange
which in normal civil disputes takes place through pleadings,
examination for discovery, and demands for production of
documents. In this way each party to some extent becomes aware
of the other's case, thus contributing to a convergence of
probability estimates and a more realistic view of the likely
award at trial. This, it is submitted, goes a long way to
explain the very low percentage of disputes actually litigated.

The second institutional characteristic is the role
played by legal precedent. Litigation uses more resources,
normally, than those used in the process of settlement, and
hence may be considered an inefficient method of dispute
resolution. On the other hand however the process of litigation
generates legal precedents and these contribute to greater
certainty as to the substantive law applicable to a given fact
situation. In the absence of a sufficient stream of precedents,
or in a situation of largely random decision making by courts,
there would be a far greater likelihood of individual probability
estimates varying from the "true" probability of success, and
hence a greater likelihood of litigation when settlement would be
more efficient. Precedents and stare decisis thus aid in the
convergence of probability estimates. It is clearly desirable
that some cases be litigated - those in areas where the law is
unclear or where changing relative costs require changing legal
rules. What the optimal stream of litigation is, however, will
not be investigated here.

4.   The Special Case of the Consumer

The settlement-litigation model was developed generally
without attaching particular or unique features to either
plaintiff or defendant. When we come to consider the consumer in

the context of the model, the parties become drastically less symmetrical.

We are concerned here with the consumer as plaintiff or potential plaintiff in an action relating to "defective" goods. We will not in this paper deal with consumers as defendants, which would be the case in debt collection actions.

One observes few court actions being brought by consumers over defective products, and it is frequently suggested that consumers very often fail to obtain voluntary redress from sellers or manufacturers. In the terminology of this paper, it appears that consumers frequently renounce their claims against sellers - i.e., settle for zero - in spite of having valid claims in relation to defective products. We now consider whether the general settlement-litigation model can help us to understand or explain this phenomenon.

It will be recalled that the more risk averse the plaintiff and the more risk he faces, the lower will be his minimum acceptable settlement price, and the higher will be the minimum probability of success below which he will not pursue the action.

There are a number of reasons why we may expect consumers to face greater uncertainty and have less ready access to knowledge with respect to their legal rights and the validity of their claims. The greater variance associated with their probability and award estimates suggests greater risk than that faced by, for example, business litigants. The greater the proportion of a party's total wealth which is represented by the claim, the greater will be the risk faced by the party. A diversified portfolio of assets reduces risk, while the more the portfolio approaches a single-asset portfolio the greater the risk, in terms of the variance of possible outcomes, becomes. It is submitted that in consumer conflicts generally, the value of the claim represents a higher proportion of the consumer plaintiff's assets than it does of the business defendant's assets.

The settlement-litigation model suggested that the greater the plaintiff's litigation costs in relation to the possible award, the higher the "quit" probability, and similarly the greater the proportion of these costs that are not compensated in a successful action, the higher the "quit" probability. It would appear that the average consumer plaintiff finds himself in both situations. It has already been suggested that consumers have less ready access to legal information and legal services, which in turn implies higher costs per unit of legal services purchased than those faced by businesses.

Another cost problem faced by consumers is that their optimal resource use[8] may be smaller than the minimum amount they must employ if they are to utilize the court system. Regardless of the size of the claim, a certain amount of time must be spent in court, certain fees must be paid, and so on. These fixed costs may not only be greater than the optimal costs in relation to a small consumer claim, they may be greater in many instances than the claim itself. The necessity of "purchasing" a discrete quantity of services larger than that actually required by the claim may, in the case of small claims, be sufficient by itself to raise the "quit" probability to unity.

Further, it is submitted that a large proportion of a consumer's potential litigation costs will be non-compensable. First, psychological costs - anxiety, anger, frustration, etc. - will never be compensated in a successful action. Second, many explicit costs, such as those relating to transportation, baby-sitting, time off work and so on, are rarely if ever compensated by courts. Third, most jurisdictions place relatively low limits on allowed costs in small claims courts,[9] the venue most likely to be employed in consumer disputes. Although this limit may benefit the large number of consumer debtors who find themselves defendants in small claims courts, it can only discourage consumer plaintiffs from seeking redress there. In short, the proportion of uncompensated costs facing consumer litigants is likely to be much higher than that facing litigants generally and this, it is submitted, helps account for the high quit rate apparently observed.

The smaller the claim the higher the "quit" probability, and if the predicted award is equal to or greater than the uncompensated costs even a case that is certain to succeed in court will not be brought. It is submitted that a large number of consumer claims fall into this category.

It was suggested earlier that if settlement costs are less than trial costs, there is a net benefit to society in terms of resource saving if the parties settle and this benefit is obtained regardless of which party captures the saving. Can the

---

8. See Chapter IV, section 3(b).

9. P. Sigurdson, Small Claims Courts and Consumer Access to Justice (Ottawa: Consumer Research Council Canada, 1976), p. 55.

same be said if there is a zero price settlement due to plaintiff's renunciation of his claim, particularly when the plaintiff is a consumer in receipt of defective goods? The answer is that there is a benefit in terms of resources saved when plaintiff "quits", but not necessarily (or likely) a <u>net</u> benefit to society. A high "quit" probability for consumers effectively abrogates the legal rule of producer-seller liability for defective products. In Chapter III it was demonstrated that in the absence of such a rule, i.e., under a regime of consumer liability, the number of defective products per period will be in excess of its social optimum and a social loss will be incurred. It is submitted that this loss in terms of magnitude is of a large order in relation to any resource savings via renunciation of claims.

But it is further submitted that if the appropriate legal rule holding the producer liable is restored by in some way providing access for the consumer to lower cost adjudicative mechanisms, this will not necessarily lead to a large increase in the use of legal resources. Although more consumer cases are likely to proceed through the court or other adjudicative system, we may also expect to see a higher rate of settlement at non-zero prices, and the lower the court or adjudicative costs and the clearer the liability rules both in general and in relation to particular products, the higher we may expect settlement prices to be.

It is not submitted that it is possible to make the courts or other dispute resolution mechanisms available for all claims no matter how low in value. A consumer resolves the problem of a defective book of matches by throwing it away and buying another. He may seek an alternative brand but if all producers face a consumer liability rule, all will have a higher than optimal defect rate. The consumer with but minor annoyance does not question the acceptance of a zero price settlement in such a situation and clearly the costs of any conceivable dispute resolution mechanism would be far in excess of the value of the product. Nevertheless, although the individual loss is trivial, the social loss via higher than optimal defect rates may be significant and is certain to be so when aggregated over a large number of low cost items. To deal with this problem, it is submitted that it is necessary to look outside of traditional individual plaintiff-defendant resolution schemes. Easier class actions, consumer association actions, and public agency orders are possible solutions, but evaluation of these is beyond the scope of this paper.

CHAPTER V

IN CONCLUSION

A purpose of this study has been to examine those factors influencing complainants, particularly consumer complainants' use of adjudicative forums in the pursuit of their complaints. In conclusion, we are not in a position to identify a forum with unique characteristics which will provide the consumer with a realistic alternative to abandoning his complaint.

Different forums have been suggested in various proposals. These include variants (or combinations) of modified small claims courts, consumer courts, consumer boards or agencies (acting in the consumer arena as labour relations boards act in the arena of labour conflict), mediation facilities, and arbitration facilities. Given the wide variety of consumer claims extant it is unlikely that a single mechanism will serve all purposes. It is quite possible, for example, that for claims of large value, say over $1000, modified small claims courts may prove satisfactory. It is difficult to envisage a courtlike structure being of great use to the consumer, however, for claims of, say, less than $100.

Any successful mechanism must involve four partially related elements: low costs, simple procedures, speed, and publicity. It must be stressed however that from the economist's perspective, the object is not to entice consumers into the forum, but rather to provide a realistic alternative to the consumer so as to persuade sellers/producers to voluntarily redress consumer grievances. The economist's concern, it will be recalled, is not with justice or fairness, though these traditional values can be shown to have economic content, but rather with resource allocation and misallocation. The rule of producer or seller liability for defective products was shown to be socially optimal in most situations. But such a rule can only be effective if the consumer has a forum in which to enforce it. It is thus only superficially paradoxical to state that the ready availability of an inexpensive, speedy, easy to use redress mechanism will ultimately reduce the need to fall back upon such a mechanism.

DATE DUE
DATE DE RETOUR

JUL 1 5 1999

CARR McLEAN   38-296