THE RULE AGAINST PENALTIES IN CONTRACT: AN ECONOMIC PERSPECTIVE*

BY ANDREW HAM**

[*This article considers the arguments for the abolition of the penalty doctrine in the law of contract. The basic principles and most recent developments in the law are examined, then the tools of economic analysis are used to develop a series of propositions. The questions of optimal breach and efficient allocation of risk are considered. Also addressed are the issues of strategic behaviour by the parties and the relationship between unconscionable conduct and the rationale for the penalty doctrine. The author concludes that it is most efficient to allow parties on equal terms to reach agreements in regard to stipulated damages without interference from the courts.*]

INTRODUCTION

The general law over the centuries has evolved new doctrines, *ad hoc*, based on often vague and ephemeral judicial notions of equity and justice. This paper seeks to take one such doctrine and apply the basic principles of welfare economics to enunciate and test the implicit assumptions on which the doctrine is based in a systematic and rigorous way.

The law of contract is a vital part of business relations and the daily processes of negotiation and enforcement of agreements. By formulating clear rules as to what is and is not acceptable behaviour, relations between contracting parties are facilitated and the use of contractual instruments encouraged. The law relating to liquidated damages and penalties remains an unfortunate exception to this general rule.

In the course of drawing up an agreement, the parties may determine in advance the damages that are payable should one party breach the contract. Thus the expense and uncertainty of litigation are avoided and the parties can be sure that their interests are fully protected. If there is a breach, then the breaching party may challenge in the courts the previously agreed sum as 'excessive' and therefore a penalty which should not be enforced. This common law rule, the penalty doctrine, is the subject of this article. The discussion here is in the context of fully bargained, arms-length agreements between parties of comparable bargaining power. Stipulated or agreed damages are classed as liquidated damages when the court finds the amount to be a genuine pre-estimate of probable damage, thereby avoiding the difficulty of proving actual damage in an action for breach, particularly when the damage is uncertain. In this case, the courts will implement the parties' agreement without the need to prove actual

* This paper is based on the author's B. Com. Honours research essay submitted in October 1987. It has been updated to take account of recent developments in the law.

** B. Com. (Hons) (Melb.), currently student of law, University of Melbourne. The author gratefully acknowledges the advice and assistance of Dr Philip L. Williams in the preparation of this paper.

damage and regardless of the amount of actual damage that is provable. A stipulated sum is classed as a penalty when it is found to be in the nature of a threat, intended to prevent breach by establishing a greater incentive for performance. Such clauses are not enforceable, and only actual damages as determined by the court are awarded. Apart from the separate issue of whether such interference is justified at all, there is the further problem that the rules themselves are in a confused and contradictory state. Despite some recent decisions of high authority[1] there remains sufficient uncertainty in the law to make it worthwhile for the breaching party to challenge the agreement in order to avoid his or her obligations under the contract. Clearly, reform of the law relating to penalties in contract is essential to reduce this unnecessary litigation which defeats the purpose of pre-estimating damages in the first place.

The basic philosophy of the economic analysis of legal rules is that individuals are rational agents aiming to maximize their well-being and make the best agreement possible, subject to the constraints imposed upon them by the law and the other party. My argument is that only the parties themselves have sufficient information at the time of forming the contract to be able to formulate the optimal damages clause taking all relevant subjective and objective factors into account. Therefore, in commercial transactions which are fully negotiated between parties of comparable bargaining strength, the optimal rule is to enforce all stipulated damage agreements. The courts still have an important role, however, in ensuring that the contract is fully and fairly negotiated and that one party is not taking unconscionable advantage of the other.

The penalty doctrine, when it strikes down a stipulated damages clause, substitutes the court's own assessment of the loss due to breach. Thus there are two major legal issues; the criteria for enforcement of a stipulated damages clause and the limitations of the court's assessment of loss in comparison to that of the parties themselves.

THE LAW OF LIQUIDATED DAMAGES AND PENALTIES

When the parties have made provision in advance for damages payable on breach, the plaintiff may seek to recover this agreed sum rather than his or her actual loss. The law of liquidated damages seeks to reconcile the conflict between restricting compensation to the loss caused by breach in order to protect weaker parties and the principle of freedom of contract which decrees that the parties should be free to determine the consequences of breach for themselves. One of the effects of this compromise is that a clause may be a penalty even though actual loss is assessed by the court at a greater amount than that provided for.[2] Thus the larger amount may be recovered even though the clause would have been enforced if it had been found to be liquidated damages and to be intended to limit liability.

[1] *Esanda v. Plessnig* (1988) 166 C.L.R. 131; *AMEV-UDC Finance v. Austin* (1986) 162 C.L.R. 170.

[2] *Widnes Foundry v. Cellulose Acetate Silk Co.* [1931] 2 K.B. 393 (H.L.).

The current law of penalties began with the *Dunlop*[3] case, still a leading authority, which lays down a series of propositions that remain the foundation of the penalty doctrine today. The decision has not gone uncriticized, however.

> The historical perspective suggests that the rules in *Dunlop* reflect an assortment of notions; a jumble of historical curiosities which out of context provide no unitary rationale for invalidating stipulations for the payment of an agreed sum.[4]

'The modern law on penalties takes form from the time the courts chose to disregard the intentions of the parties as expressed on the face of the contract.'[5] The wording used by the parties is of marginal importance as subjective intention is no longer relevant, and the terminology used adds nothing to the objective question of whether the amount is a genuine pre-estimate of damage, although there is a *prima facie* assumption that the parties mean what they say. Whether or not a sum is a penalty or is liquidated damages depends on the intention of the parties, but this is a question of construction 'to be decided on the terms and inherent circumstances of each particular contract, judged as at the time of making the contract, not as at the time of the breach.'[6] Four rules were summarized by Lord Dunedin[7] to guide the court in determining whether the agreed sum is a genuine estimate of probable loss. They are as follows;

(a) the stipulated sum is a penalty if it is extravagant and unconscionable in amount in comparison with the greatest loss that could possibly follow from the breach.

(b) If the breach consists only of failing to pay a certain sum of money under the contract and it is agreed that a larger sum shall become payable in such a case, then this latter sum will be a penalty because the damage from breach is able to be exactly defined. Fixing a larger sum cannot be a pre-estimate of the probable damage resulting from failure to pay a smaller sum.

(c) There is a presumption (but no more) that a clause is a penalty if a single lump sum is made payable by way of compensation on the occurrence of one or more or all of several events, some of which may occasion serious and others mere trifling damage.

(d) It is no obstacle to the sum stipulated being a precise pre-estimate of damage that the consequences of breach are such as to make precise pre-estimation almost an impossibility. On the contrary, that is just the situation when it is probable that pre-estimated damages was the true bargain between the parties.

The basic principles of the *Dunlop* case have been developed by later decisions of the Australian courts, particularly in the context of finance leases and hire-purchase agreements. These cases have eroded the principle that relief should only be granted if the agreed amount were exorbitant or unconscionable. In the interests of increased certainty, the courts have struck down clauses because the

3 *Dunlop Pneumatic Tyre Co. v. New Garage* [1915] A.C. 79.

4 Muir, G., 'Stipulations for the Payment of Agreed Sums' (1985) 10 *Sydney Law Review* 503, 516.

5 *Ibid.* 508.

6 *Dunlop Pneumatic Tyre Co. v. New Garage* [1915] A.C. 79, 86-7 *per* Lord Dunedin.

7 *Ibid.*

agreed amount may possibly exceed the amount which a court would award for the breach in question. For example, in *Citicorp v. Hendry*[8], Priestley J.A. held a carefully drafted agreed damages clause in a finance lease for three cranes to be a penalty because there was a chance of a windfall gain to the financier if the agreement were terminated early in the 48-month term. The High Court began to reverse this trend in the cases of *AMEV-UDC v. Austin* and *Esanda v. Plessnig*, and called for a return to the basic principles to allow parties greater freedom in reaching agreements without interference.

> the concept is that the agreed sum is a penalty if it is 'extravagant, exorbitant or unconscionable' . . . This concept has been eroded by more recent decisions which, in the interests of greater certainty, have struck down provisions for the payment of an agreed sum merely because it may be greater than the amount of damages which could be awarded for the breach of contract in respect of which the agreed sum is to be paid. These decisions are more consistent with an underlying policy of restricting the parties . . . to the recovery of an amount of damages no greater than that for which the law provides. However, there is much to be said for the view that the courts should return to the *Clydebank Engineering* and *Dunlop* concept, thereby allowing parties to a contract greater latitude in determining what their rights and liabilities will be . . .[9]

In the *Austin* case Mason and Wilson JJ. said that this increased latitude will be mutually beneficial to the contracting parties. Certainty will be enhanced as parties may determine their rights and liabilities on breach or termination more precisely. Thus, they may compensate loss that is difficult or impossible to quantify or not recoverable at common law, and avoid costly and time consuming litigation. They emphasize that it is the essence of a penalty that an agreed sum is only a penalty if it is extravagant, exorbitant or unconscionable and therefore out of all proportion to the damage likely to be suffered as a result of the breach. They held that the test

> is one of degree and will depend on a number of circumstances, including (1) the degree of disproportion between the stipulated sum and the loss likely to be suffered by the plaintiff, a factor relevant to the oppressiveness of the term to the defendant, and (2) the nature of the relationship between the contracting parties, a factor relevant to the unconscionability of the plaintiff's conduct in seeking to enforce the term.[10]

Similarly, in *Plessnig's* case, on appeal from the Full Supreme Court of South Australia, Wilson and Toohey JJ. in the High Court reject the proposition accepted in the Full Court that the mere possibility of unfairness in the agreed formula by which the stipulated damages are calculated is sufficient to characterize it as a penalty. Such a result fails to allow the latitude that necessarily is associated with a genuine pre-estimate of damage.

Kirby P. and Mahoney J. A. in the New South Wales Court of Appeal made a number of comments on the unsatisfactory state of the law of penalties.[11] They agreed that there should be some consideration of the relative bargaining positions of the parties, and criticized the court's failure to recognize that a provision alleged to be a penalty, if otherwise enforceable in the contract, remained enforceable until relief was granted to the party aggrieved. Thus,

[8] [1985] 4 N.S.W.L.R. 1, 35.

[9] *AMEV-UDC v. Austin* (1986) 162 C.L.R. 170, 190 per Mason and Wilson JJ. This passage was approved by Wilson and Toohey JJ. in *Esanda v. Plessnig* (1988) 166 C.L.R. 131, 139.

[10] *AMEV v. Austin* (1986) 162 C.L.R. 170, 193.

[11] *Citicorp v. Hendry* [1985] 4 N.S.W.L.R. 1.

> in the present state of authority there is neither an appropriate basis to take into account the nature of the transaction and the relationship of the parties, nor is there means of providing partial relief. If the clause is characterized as a penalty it is unenforceable *ab initio* . . .[12]

LIMITATIONS ON THE DOCTRINE OF COMPENSATION

The second issue that arises is the court's quantification of loss. This is important for three reasons: First, it is this assessment that is the alternative to that stipulated by the parties, hence its perceived shortcomings may motivate the parties to negotiate their own pre-estimate of loss. Secondly, should the pre-estimate be struck down as a penalty it is the court's assessment that is substituted as the recoverable amount. Thirdly, in deciding whether an agreed sum is 'extravagant and unconscionable' in comparison with the greatest probable loss, the courts must assess what that loss might be by their own criteria.

In principle the expectation interest of the wronged party includes all pecuniary and non-pecuniary advantages which that party would have enjoyed if the party in breach had performed. However, it is rare for a court to award the full expectation as, in order to allow only a 'just and reasonable' sum to be recovered, it construes the promised advantage as narrowly as possible, excludes uncertain, remote, and non-pecuniary losses and requires the innocent party to take all reasonable steps to mitigate his or her loss.

One problem that has arisen is that of the causal link between breach and damage suffered. For example, in most credit contracts and hiring agreements, the goods are exchanged for a series of instalments representing the price or rental. Default on an instalment by the customer is only a breach of warranty and does not entitle the financier to bring the contract to an end. If financiers exercises their contractual right to terminate the agreement then strictly they are entitled only to recover the missed instalments as any greater losses are due to their own voluntary decision to end the agreement. The English Court of Appeal took this strict view in *Financings v. Baldock*[13] which was accepted by the High Court of Australia.[14]

The question then arises of what loss the court is to use in assessing whether an agreed amount is extravagant and unconscionable under the penalty doctrine. Brennan J. held in the recent *Plessnig*[15] case that no clear opinion had yet emerged in the High Court on this issue. He went on to consider *Austin*, and found that although loss caused by termination (*cf.* breach) of the contract was not to be included in the financier's damages, such losses should be taken into account in determining whether or not a pecuniary liability imposed by the contract is a penalty.

> In the light of these observations, I take the law to accept an incongruity in holding that an owner's damages at law for a non-repudiatory breach are limited to losses caused by the breach alone while holding that a clause which imposes a liability on the hirer [*i.e.* the customer] to pay the losses caused by exercise of a power to terminate a hiring upon breach is not a penalty. It may be

[12] *Ibid.* 23-24.
[13] [1963] 2 Q.B. 104.
[14] *Shevill v. Builders Licensing Board* (1982) 149 C.L.R. 620 and more recently in *AMEV-UDC v. Austin* 176, 186.
[15] (1988) 166 C.L.R. 131, 144.

> appropriate to reconsider this incongruity in some later case . . . For the moment I adopt, in common with the other members of the Court, the view that the owner's loss consequent upon the termination of the hiring for non-repudiatory breach is to be taken into account in determining whether the recoverable amount prescribed by cl. 5 is a penalty . . . assuming that the power to terminate . . . effectively exercised, the question is whether the amount of the hirer's liability . . . is extravagant or unconscionable compared with the greatest losses that could conceivably be proved to have followed the breach *and* the termination.[16]

So it would appear that while the courts are not yet ready to consider 'actual' loss, they are taking a more realistic and less technical approach than in the past.

CONTRACTS AND BREACH OF CONTRACTS

This change in direction cannot yet be said to be settled law, however, and much confusion remains in the application of the penalty doctrine. The law does need to be reconsidered; the question is, what does economics have to offer in such a reconsideration? First, the rationale for the concept of freedom of contract and the proposition that it is sometimes better if agreements are not kept must be considered.

Freedom of Contract

Proposition 1: Agents should in general be free to bind themselves in any freely bargained exchange which they might choose.

Freedom of contract is one of the principles that underlies contract law, yet its rationale is seldom examined. It is based on the standard assumptions of economic analysis. To simplify matters we assume that all benefits and costs can be measured in dollar values, including 'non-economic' considerations. It is individuals themselves who determine the dollar values to be placed on the benefits and costs of an agreement. Furthermore, individuals will maximize the difference between their benefits and their costs; utility maximization. It follows that left to themselves, parties on equal terms will reach the agreement that maximizes the benefits to both sides. Such an outcome is desirable because it is efficient; that is, from a given pool of resources, the net benefit to the overall community is maximized.

The problem of the efficient allocation of resources, the outcome of freedom of contract, is quite separate from the problem of the equity of their distribution which is a political decision. Legal intervention is justified if unfairness, dishonesty, or some other factor against public policy has led to a sub-optimal (*i.e.* inefficient) distribution of resources. 'Interference with freedom of contract is deemed justified when necessary to protect community interest not represented by the parties to a transaction: the law declares certain bargains void as in violation of public policy.'[17] The law will not intervene in voluntary agreements without good cause. The question that must be considered is whether the penalty doctrine is grounded in one such good cause.

[16] *Plessnig*'s case, (1988) 166 C.L.R. 131, 147-8 *per* Brennan J.

[17] Birmingham, R. L., 'Damage Measures and Economic Rationality: the Geometry of Contract Law' [1969] *Duke Law Journal* 49, 63.

Optimal Breach of Contract

Proposition 2: Breach is optimal if it potentially allows a Pareto-improvement over performance.

Contractual obligations are enforced by the law at least partly because of the belief that a promise has inherent moral force which should be recognized. Also, business certainty requires that promises should be kept. However, breach should be encouraged when, relative to performance, it yields a net gain to the parties (*i.e.* a Pareto-improvement). Circumstances frequently change between the time of formation and the time of performance. If the situation becomes such that the opportunity cost of performance to one party is greater than the other party's gain from the contract, then there is clearly a net sum gain if the contract is not performed. Encouraging the breach of such inefficient contracts will not, in principle, deter the formation of future contracts because the victim of breach is always fully compensated and hence is indifferent between damages and performance. The breaching party is able to fully compensate the innocent promisee for losses sustained due to the former's breach, and still gain him or herself. Thus breach will benefit both parties and is 'efficient'.[18] For example, a contract for the supply of rainforest timber is signed by a wholesaler and a retailer six months before the date agreed for delivery and in that time a worldwide moratorium on the logging of rainforests is ratified by all the relevant nations. If the ban is effective, supplies will dry up and prices will rise. The value of the timber to the wholesaler clearly becomes much greater than the price at which he is contracted to on-sell it to the retailer. The efficient result is for the wholesaler to end the contract and sell at a price reflecting the current market price, either to the original buyer or another bidder. The seller can give the buyer his full expectation under the original contract and keep the gain from the rise in price. Thus he is better off and no one is worse off from the breach of contract.

Whether this occurs in practice depends on how damages are assessed: to protect the innocent party's expectation interest and put him or her in the same position as if the contract had been performed; or to maintain the correct incentives and prevent inefficient breach by ensuring that the cost of breach to the party in default is as great as the cost of compliance. Thus damages should force the breaching party to include all the costs of breach to all parties in his or her decision to end the contract.

In most cases these two goals of compensation will coincide and will encourage optimal breach. By allowing recovery of an equivalent amount to the expected benefit lost when the contract is broken, the expectation measure of damages transfers the full cost of breach from the injured promisee to the promisor. Therefore a party has an incentive to break his or her promise only

18 Although Birmingham, *op. cit.* introduced the concept of 'efficient breach', the term itself is attributed by Macneil (Macneil, I. R., 'Efficient Breach of Contract: Circles in the Sky' (1982) 68 *Virginia Law Review* 947) to Goetz & Scott (Goetz, C. J. and Scott, R. E. Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach.' (1977) 77 *Columbia Law Review* 554).

when the gains from doing so are greater than the lost expectation of the innocent party which must be compensated. Breach will only occur when it results in an increase in the productive value of the resources in question. This is not always the case, as the optimal level of damages depends on the motive for breach.

Breach may be involuntary, to avoid greater loss, or it may be voluntary, motivated by greater profits (from a higher offer, for example) than those expected from performance. Thus, in a contract for the production of goods, costs may be uncertain until after the contract is made. In this case breach will be efficient if the seller's production costs exceed the value of the goods to the buyer. There may be uncertainty due to the chance of third-party offers; such as in the case of a contract transferring possession of existing goods. In the latter case there are two possibilities; (a) if bids are available only to the seller, the seller should not perform when the bid exceeds the buyer's expectancy, as above, and (b) if bids are also available to the buyer, it will always be efficient for the seller to perform because the buyer has the opportunity to resell to a higher bidder if there is one. Thus if there is performance the buyer benefits, or if there is a breach the seller benefits from the higher offer. However, unless the market is exceptionally well organized, buyers (a diffuse group) are less likely to have access to third party offers than are sellers who are specialists and dealers. Incentive-maintenance may dictate less comprehensive compensation for unavoidable breach than for voluntary breach in order to get the most efficient result. The significance of this will become apparent once we have considered the effect of damages on the allocation of risk.

DAMAGES, PENALTIES AND RISK

Damages Allocate Risk

Proposition 3: The measure of damages affects the parties' allocation of risk.

The standard of compensation employed will determine how the gains or losses from breach are distributed between the parties. As a corollary of the two motives for breach, there are two forms of risk to be allocated; market risk, that the market price will change, and casualty risk, that one party will be unable to perform his obligations due to unforeseen circumstances.

Under expectation damages, if all loss is compensated by the breaching party, then of course all risk is borne by that party. The courts limit the risk that must be borne to that which *inter alia* is reasonably foreseeable at the time of contracting. This, together with the other limitations on what the courts consider to be recoverable loss, means that some of the risk of loss must often be borne by the innocent party. This may or may not be the optimal allocation of risk, depending on the parties' relative attitudes to risk and whether breach was due to the choice of the breaching party or was unavoidable.

Proposition 4: It may be optimal not to compensate some losses, which may lead to a conflict between incentive maintenance and the desired allocation of risk.

Full compensation of all losses from breach, pecuniary and non-pecuniary, should only be allowed if breach results from the choice of the breaching party. This ensures the correct incentives for efficient breach. However, if breach is unavoidable due to some unforeseen event, the full cost of the breaching party's actions should not necessarily be borne by that party.[19] Some kinds of loss should not be compensated because although the victim has been 'injured', he or she cannot replace what has been lost with inferior substitutes, as none exist. This applies particularly to subjective, non-quantifiable, non-physical loss. Attempts to compensate such losses may actually make the plaintiff worse off if he or she is in effect being 'insured' for losses that cannot be replaced, and which they do not wish to have insured. Sellers will adjust prices to allow for the added risk they must bear even though buyers have no use for this extra insurance and do not wish to pay the premium.

This result creates a conflict between insurance and incentives as less than full compensation means the cost of breach to the seller is less than to the buyer, reducing the seller's incentive to perform. Although voluntary breach can be distinguished from involuntary breach, in the latter case a seller can usually take precautions to influence the probability of a random event (such as maintenance to prevent breakdowns). Such precautions cannot be observed by the courts in assessing what loss to compensate after the event. One solution to the conflict is for the buyer to compromise.[20] As damages are the only means available to influence reliability, a liquidated damages clause must be drawn up such that the buyer is insured, if necessary to excess, to maintain the seller's incentive to perform.

The Effect of Penalties

Proposition 5: 'Excessive' liquidated damages may prevent breach even when such breach would be optimal.

The penalty doctrine aims to prevent unfair recovery in excess of justifiable and quantifiable loss and to prevent inefficient performance through fear of a penalty when breach is preferable.

Economists[21] have developed models to show the ideal response to a change in circumstances assuming partial performance is possible. A penalty clause acts as a means of forcing inefficient performance as this is preferable (less costly) to paying a still larger penalty. So a penalty clause is analogous to an award of specific performance as both, whether by direct court order or by the deterrent effect of an '*in terrorem*' clause, result in performance of the contract that one party sought to breach causing the other party to initiate litigation. Specific performance is only awarded by the courts when ordinary money damages are

[19] Rea, S. A., 'Nonpecuniary Loss and Breach of Contract' (1982) 11 *Journal of Legal Studies* 35.

[20] *Ibid.*

[21] Goetz & Scott, *op. cit.*

inadequate, and cannot substitute for performance. As a penalty has the same effect, it can be argued that similar criteria should be applied to their enforcement. That is, even when a stipulated sum has been found to be a penalty, the courts should enforce it if there is evidence that damages as assessed by the court would be inadequate.

If enforcing a penalty amounts to *de facto* specific performance then other writers' analyses of the latter can be extended to apply more widely. It has been shown[22] that specific performance is better than other remedies in minimizing 'excessive breach' due to the courts' undercompensation of loss. That is, breach occurs too often (*i.e.* contracts that should be performed are breached) because the party breaking the agreement does not take full account of the losses he or she imposes on others. 'Excessive performance' (*i.e.* contracts that should be broken are performed) will not be a problem if both parties have access to third party bids because the buyer can ensure the efficient result by reselling to a third party who offers more for the goods than their value to the original buyer. Performance of the original agreement only affects the distribution of the gains from resale. Hence specific performance is always superior to damages, given our assumptions, and therefore penalties that amount to specific performance should be enforced.

THE COURT'S ASSESSMENT OF DAMAGES

The Costs of the Penalty Doctrine

Proposition 6: Shortcomings in the court's assessment of loss may cause an optimal, efficient liquidated damages claim to appear excessive.

For a number of reasons the court's assessment of loss to the plaintiff may differ from that actually suffered. This may mean that a stipulated damages clause that would fully compensate is not enforced, and an inadequate court award is substituted.[23] Thus the court's inference of unfairness in their attitude to agreed damages clauses for what they consider to be excessive amounts may not be warranted, and refusing to enforce such clauses may impose more costs than it removes.

Traditional damage measures provide quite adequate compensation in purely commercial transactions, as losses can be objectively evaluated either on the basis of lost profits from anticipated breach, or to reflect the difference between contract and market prices. The principles of compensation are harder to apply in two other possible cases. First, when the promised performance has a market value but the promisee attaches an additional idiosyncratic value to perfor-

22 Shavell, S., 'The Design of Contracts and Remedies for Breach' (1984) 99 *Quarterly Journal of Economics* 121.

23 The High Court has begun to recognize this problem in relation to at least some aspects of damage in *AMEV-UDC v. Austin* and *Esanda v. Plessnig supra* 5-7.

mance[24] which he or she has contracted for by paying a premium on the market price. Evaluated after breach, such subjective loss is regarded as too speculative and uncertain to be recovered in the courts. Of course, such a limitation is justified where the costs of establishing idiosyncratic value exceed the costs of an inaccurate measure of damage. Secondly, limitations are imposed on the recovery of consequential loss and even reasonably foreseeable loss that is too uncertain to quantify. This avoids the difficult process of quantifying uncertain loss, but at the plaintiff's expense. Even when the parties have attempted to resolve the uncertainty with a stipulated sum, as efficiency dictates, the penalty doctrine increases transaction costs again by relating this sum to the maximum uncertain loss possible before enforcing the agreement.

Parties who know they may be disadvantaged by such rules can unambiguously gain if liquidated damages clauses that compensate fully are always enforced. 'In the absence of evidence of unfairness or other bargaining abnormalities, efficiency would be maximized by the enforcement of the agreed allocation of risks embodied in a liquidated damages clause'.[25] Efficiency is enhanced through minimization of transaction costs if agreements negotiated *ex-ante* are enforced.

Apart from the limitations of compensation by the courts, such damages awards are not the best way of insuring against loss from breach because litigation is an inherently expensive and uncertain process. The seller already knows the probability of breach and could set the premium accordingly if he or she were to directly insure the buyer. A third party insurance company would have to devote significant resources to finding this probability out. Further, a seller may be able to influence the probability of breach for the best result. The penalty doctrine prevents parties from self-insuring idiosyncratic value like this.

More general enforcement of stipulated damage agreements would considerably reduce the time and expense of litigation. Proof of damage sustained, frequently a complex and expensive process, would no longer be necessary. In addition, if liability is not a major issue, then out of court settlements may be promoted due to the greater predictability of the outcome as the problem of valuation uncertainty has been removed.

The penalty doctrine undoubtedly increases transaction costs at the time of forming the contract. The threat of subsequent judicial review and non-enforcement of a clause means the parties must spend extra time and effort to frame their liquidated damages clause so that it fits the requirements imposed by the penalty doctrine and the courts will enforce it. These unnecessary costs are increased by the current uncertainty of the Australian law in this area. For example, transaction costs are increased by the court's refusal to allow the parties to specify a single damages provision for a range of contingencies. Parties cannot spread their risks by averaging potential loss over a number of transactions or a number of contingencies in one transaction. Each contingency must and will be

24 Goetz & Scott *op. cit.* 570.
25 *Ibid.* 578.

negotiated separately as long as the costs of such negotiations are less than the expected cost of litigation should a 'blanket' clause be challenged.

Finally, traditional measures of compensation are unsatisfactory because of their failure to distinguish breaches due to unexpected shifts in the profitability of agreements ('Voluntary Breach') from breaches in which there is no net gain to be allocated ('Involuntary Breach'). Under expectation damages the person in breach retains all the net gains from breach as the non-breacher is placed in the position he would occupy if he had received performance. These gains could be allocated between the parties (largely a simple case of wealth transfer) by a liquidated damages clause. The incentive to negotiate such a clause arises when the costs of negotiation are less than the expected costs of relying, in breach, on the standard damages rules of the courts. Compensation by the courts has a number of shortcomings, yet attempts by the parties to overcome the problems of uncertain loss, subjective valuation and the distribution of windfall gains are frustrated by the rule against penalties. Indeed, the penalty doctrine increases transaction costs in more cases than it reduces them, strengthening the case for enforcement of stipulated damages clauses. The argument is further bolstered by examination of the premise underlying the penalty doctrine that 'penalties' are 'unfair'.

The Presumption of Unfairness

Penalty clauses are assumed by the courts to be unfair or unconscionable because such an agreement could only have been reached by an abuse of the bargaining process. They assume that compensation in excess of actual loss is unjust and oppressive. This begs the question of whether the courts do in fact award full compensation and of whether such full recovery is always the best result. These issues have already been discussed but even if we accept the court's assumptions, there are problems.

First, the reasoning in decisions assessing a stipulated damages clause often comes close to conflicting with the fundamental principle that the law will neither inquire into the adequacy of consideration nor, as a rule, offer relief from what has turned out simply to be a bad bargain.

Secondly, the unfairness rationale is a *post hoc* judgment of the bargain which ignores the risk allocation made at the time of contracting. The entire bargaining process should be taken into account. The risk of breach should be recognized as a factor in determining the contract price. A premium may be paid by the buyer in order to get the seller to accept some of his or her risk in the form of an agreed damages clause for a sum greater than that the courts would award. To call an agreed remedy 'unfair' either neglects its role in shifting risk or implies inadequate consideration for the shift. Neither assertion is justified.[26]

Stipulated damages clauses are often used where the real loss, including the subjective elements, is difficult to assess in advance but a pre-estimate can be

[26] Kaplan, P. R., 'A Critique of the Penalty Limitation on Liquidated Damages' (1977) 50 *Southern California Law Review* 1055, 1071-2.

made which averages out the risks of over- and under-compensation. The question of whether a sum is coercive becomes irrelevant when the innocent party is indifferent between the money and performance. By inflating the measure of agreed damages the innocent party signals the other party that he or she values performance highly. By making a rational choice to minimize his or her own costs the breaching party will only breach when the cost of performance exceeds the agreed sum. To intervene at the time of breach is an unfair redistribution of contract risks and deprives one party of a benefit he or she contracted for.

OPTIMAL ALLOCATION OF RISK WHEN THE PARTIES' ATTITUDES TO RISK DIFFER

Stipulated Damages and Insurance

Proposition 7: It may be efficient to stipulate damages in advance to optimally insure the more risk-averse party.

As we have seen, a stipulated damages clause is the most efficient and reliable way for parties to share the risk of breach. One of the most important reasons for a stipulated damages sum exceeding actual loss is the differing attitudes to risk of the parties.

A party who is risk-averse is willing to pay more (as a matter of certainty) than the expected value of the risk in order to avoid the risk. For example, take a contract with a 50:50 chance of breach at some point before completion, and a potential profit of $100 to the buyer if it is performed and $0 if it is broken. The expected value of the agreement to the buyer is (0.5 × $100) + (0.5 × $0) = $50. Therefore if the buyer is risk-averse he or she will require a liquidated damages clause specifying more than $50 to ensure the contract is performed; just how much more depends on the buyer's degree of risk-aversion. In contrast, if the buyer is risk-neutral, he or she will require only the expected value as liquidated damages.

A risk-averse person will be willing to pay a premium included in the contract price in exchange for a liquidated damages clause that the courts would currently regard as excessive. Such an agreement reduces risk-bearing costs (*i.e.* the difference between the expected value of the loss and the largest amount the person forced to bear the risk would pay to avoid it) by transferring risk to the party more willing to bear it, but conflicts with optimal breach incentives because the breaching party must pay more in damages than the actual loss he or she causes. The effect of this can be seen by example.[27] Assume profit is uncertain, the price is paid in advance and there is no market for the good. If only one party is risk-averse then the other, risk-neutral, party should bear all the risk

[27] Polinsky, A. M., 'Risk Sharing through Breach of Contract Remedies' (1983) 12 *Journal of Legal Studies* 427.

of breach. A risk-averse buyer should always receive an amount equivalent to his or her benefit; a risk-averse seller should always receive the contract price, passing any gain or loss to the buyer. Therefore the expectation measure allocates risk optimally only if the buyer is risk-neutral, the seller is risk-averse, and only the seller has access to third party offers because the seller needs only to return the buyer's expectation to him and can keep any gains from breach. If both parties have access to third party offers then the risk-neutral buyer should be returned to the position he or she would have been upon reselling to the higher bidder. That is, the seller should pay him or her an amount equal to the third party offer. Because the buyer keeps the gains such expectation damages would only optimally allocate risk if the seller is risk-averse and the buyer is risk-neutral.

The seller is more likely than the customer to have access to other offers because he or she is a dealer and a specialist relative to the consumer. Before breach there is uncertainty and both parties have incentive to reveal their attitudes to risk and negotiate a clause that allocates the contract risk to the benefit of both parties (leaving aside the different problem of unfair conduct). After breach the situation is different and both parties, knowing the outcome of the transaction, direct their energies into gaining as big a share of the benefits as possible. Any binding executory contract will reduce uncertainty about the future. A fundamental part of making such agreements is allocating the risk of loss (or gain) between the parties should the agreement not be performed. The nature of the legal process is to intervene after breach when it is no longer possible to allocate risk and so the issue is ignored. The penalty doctrine prevents the parties from making any allocation that is significantly different from that of the courts.

Stipulated Damages as an Enforceable Interest

Most contracts can be characterized as consisting of a primary obligation, the subject of the agreement, and a secondary obligation in the form of a liquidated damages clause which, as we have seen, may exceed the expected value of performance. The primary obligation is protected by the court's award of compensatory damages, assuming this to be accurate. The promisee's interests in the promise to pay an agreed amount in excess of the court's award in the event of breach are not protected, on the basis that the bargain is actually for the base promise and not the secondary promise of damages.

This rationale must be rejected for two reasons. First, the measure of compensation used by the courts is often not fully compensatory. The perceived inadequacy of protection for a party's interests in the primary contractual obligation is the reason parties resort to liquidated damages in the first place. Secondly, an agreed damages clause is a protectable interest. Such a clause transfers the risk of undercompensation for breach to the breaching party. This transfer is reflected in the contract price because as the stipulated amount of recovery rises the promisor should demand a commensurate adjustment in the contract price to compensate him or her for the increased risk he is assuming. Therefore, as the promisee paid a premium for the higher than usual damages

clause, he or she has a substantial interest in seeing that the clause is enforced should the risky contingency eventuate, and that he or she gets the insurance bargained for. If the clause is not enforced the breaching party has in effect collected the premium without having to bear the risk. In other words, a risk-averse person having paid a premium on the contract price to avoid the risk of breach is entitled to, and has an interest in, the enforcement of the secondary promise to pay an agreed sum in the event of breach.

Full Information and Wagering

Proposition 8: When there is substantial bilateral knowledge of risks, enforcement of liquidated damages is efficient and generates lower transaction costs and better incentives than a court award as long as actual loss is not less than the liquidated amount. This proposition leaves aside the problem of conflict between the incentives for optimal breach and the allocation of risk by the parties.

Rational choice requires adequate information on the alternatives available. If the parties have unequal information at the time of contracting then intervention may be justified. Without full possession of all the facts by both sides, parties cannot properly allocate the risk of breach or 'price' the liquidated damages clause. The concept of reasonably foreseeable loss in the *Hadley v. Baxendale*[28] line of cases recognizes this. To be recoverable, the loss must either be injury normally to be expected to flow from the breach in question, or if there are special circumstances resulting in unusual loss, these circumstances must be known to both parties.

The penalty doctrine does not enforce agreed remedies on the basis of the information available to the parties at the time of contracting. The mere fact that damages are overliquidated is no reason to infer unequal access to information. Those who accurately forecast or under-estimate loss are just as likely to be aware of the risks as those who over-estimate it. The penalty doctrine cannot be justified on the grounds that the parties lacked sufficient information to effectively allocate the risk of breach.

One argument against the enforcement of stipulated damages agreements has been the possibility of such clauses acting as wagers. The vital difference between a wager and other contracts conditional on the occurrence of some uncertain event is that wagers create the risk which is transferred to the promisor rather than transfer a pre-existing risk (such as in a contract of insurance). Thus to wager on a liquidated damages clause would be to set it at an excessive amount and gamble that breach occurred. The law will not enforce a wager on public policy grounds because it is perceived as immoral and economically unproductive, a waste of the court's time. A number of writers[29] have considered the question of whether enforcement of overliquidated damages would encourage

[28] (1854) 9 Ex. 341; 156 E.R. 145.
[29] Fenton, J. P., 'Liquidated Damages as Prima Facie Evidence' (1975) 51 *Indiana Law Journal* 189, 195; Kaplan *op. cit.* 1073.

wagering and found that it would not. Fenton notes that the premium paid in the contract price for the liquidated damages clause is discounted for the chance that the clause will be unenforceable. If such agreements were almost always enforced, the cost of the wager would become prohibitive. A genuinely risk-averse person will not be deterred, however, as the premium will not exceed the value to the promisee of avoiding that risk. Kaplan views speculation on liquidated damages clauses as a highly risky investment, as all the factors affecting the breach decision are in the other party's control. Success would require either a breach and enforcement of the agreed damages clause, or inefficient performance when breach would have been more efficient.

Renegotiation as an Alternative to Breach

Proposition 9: Renegotiation may overcome the problem of inefficient performance and is an alternative to breach.

In considering the effects of the penalty doctrine on the efficient breach of contracts, it is often forgotten that the choice is wider than simply breach or performance. When enforcement of a penalty would lead to an inefficient result, there are incentives for the innocent parties to renegotiate their penalty rights. Thus, the parties can negotiate an efficient solution and divide the gains between them rather than allow all the gains. This breach to go to one party or another. This could be done, for example, by negotiating a release from the contract which allows transfer of production to a more efficient market. Efficiency is not affected by the distribution of gains. This depends on factors such as bargaining power. Even if there is no penalty clause the parties will engage in renegotiation if the gains are greater than their joint costs. For example, if specific performance has been granted, the actual result may be that a compromise is reached under the shadow of the court's decision so that another solution, preferred by all, is reached.

Proposition 10: Such renegotiation may not be successful due to strategic behaviour in the bargaining process.

There are problems in relying on renegotiation as it can significantly increase transaction costs due to strategic behaviour by the parties. Circumstances may change such that the promisee no longer wants performance and so has incentive to negotiate with the promisor to release him or her from their obligations (the efficient result) and split the gain between them. The social costs of non-cooperation then increase as there are increased incentives for the promisee to induce breach by demanding performance and then to collect a premium, the stipulated penalty, over and above actual loss.

The *incentive* to induce the other party to breach arises when the benefits of doing so exceed the costs, such as when a penalty clause provides for compensation in excess of actual damage. The potential breach-inducer also needs an *opportunity*, as detected inducement may lead to non-enforcement and the loss of

any gain from inducing breach. Thus such strategic behaviour is only a problem when it is difficult to detect, as for example when performance depends at least in part on the buyer's co-operation and assistance. Breach could be induced simply by withholding information at a critical time or by failing to cut 'red tape' thereby causing significant delays. There will also be increased transaction costs on the other side of the transaction as producers attempt to combat induced breach. This will entail both researching a potential inducer's previous record as a 'victim' of breach before entering a contract, and efforts to detect breach inducement during performance of a contract. Thus, while the parties can be relied upon to renegotiate if strict enforcement of an 'overliquidated' damages clause would lead to inefficiency, additional transaction costs are incurred due to the danger of strategic behaviour.

Other remedies for breach also impose transaction costs. Undercompensation of the promisee's loss by the courts will cause the promisor to breach too often (excessive breach) as the promisee is not required to bear all the costs of the action, so that he or she gains by breaching without improving efficiency. Similarly, specific performance results in 'excessive performance' when contracts that should be ended are performed at the promisee's insistence, even though his or her benefit is less than the cost to the promisor. If performance is substantially complete once the goods are produced, the potential cost of excessive performance is low as no further resources are required to simply transfer ownership, for example, in the transfer of title or the sale of finished goods. The potential cost of excessive breach is low when the buyer can readily obtain substitutes for the subject of the contract.

Efficiency applies only to the means by which parties reach a certain wealth distribution relative to other ways of doing so; wealth distribution has no consequences for Pareto-optimality. Therefore, the relative efficiency of damage measures can be gauged by their relative transaction costs.[30]

There is no reason for post-breach negotiation costs to be higher under a damages rule than under specific performance or a penalty.[31] If specific performance is ordered, the parties will negotiate the buyer's share of the extra profit gained from the deal between the seller (promisor) and a third-party buyer. If expectation damages are ordered they will negotiate over what the original buyer's (promisee's) expectation was. This can be difficult (and hence costly) to quantify if no market and thus no substitutes exist. Therefore specific-performance entails lower renegotiation costs than expectation damages. Specific performance and penalty clauses are thus preferable to ordinary damages and a 'penalty' which does not require litigation is more efficient than an order for specific performance, which does.

30 Macneil *op. cit.* 954.

31 Bishop, W., ('The Choice of Remedy for Breach of Contract' (1985) 14 *Journal of Legal Studies* 299) presents Schwartz' theory of general specific performance which compares damages and specific performance. His results can be applied to the penalty doctrine using the conclusion of Goetz & Scott that penalty clauses lead to full performance as that cost of the penalty exceeds the cost of performance.

Unconscionability and the Penalty Doctrine

In May 1988 the Law Reform Commission of Victoria issued a Discussion Paper[32] in which it was tentatively concluded

> that the present rule against penalties should be abolished. A more flexible rule should be enacted in its place. The courts should be given a discretion to set aside an agreed damages clause which is unconscionable in all the circumstances.[33]

A similar suggestion was made by Kirby P. in *Citicorp v. Hendry*.[34] However, there is the danger that the drawbacks of an overly restrictive rule will simply be replaced by a different set of problems flowing from an overly general rule. It is on this issue, providing guidance as to what constitutes unconscionability, that the Commission's views are open to criticism. They suggest that

> In deciding whether an agreed damages clause is unconscionable, the court should have regard to the extent to which the agreed sum exceeds the actual loss which is likely to be suffered as a result of breach, not merely the extent to which the agreed sum exceeds the damages which would be legally recoverable in the absence of the clause. Even if the agreed sum appears unconscionable as judged at the time of the contract, the court should not set it aside unless it is unconscionable (that is, extravagant or substantially disproportionate) when account is taken of the loss actually suffered by the injured party.[35]

Such a reform perpetuates one of the most basic objections to the penalty doctrine.

Penalty clauses are regarded as undesirable because the court assumes that such an agreement only came about through an abuse of the bargaining process, and that it is against basic justice to allow compensation in excess of actual loss because the benefits are all on one side. This justification is based on the presumption that the court award of damages is both fully compensatory and the moral standard by which justice is measured. It is inevitable that courts will invoke considerations relating to how fairly the bargain was negotiated. The question is whether one prefers this to be done covertly through the manipulation of technical rules or overtly through doctrines that make the real grounds for the decision explicit.

> It seems obvious that the latter course best serves the ends of constructive judicial lawmaking, on the one hand, and rational independent analysis and evaluation of the aptness of legal rules on the other. In addition, cost-efficient utilisation of judicial resources argues for decisions which provide maximum feasible guidance to other parties in their actions, through rules that have some generality of application.[36]

A number of classical defences, as well as the doctrine of unconscionability, have been developed by the courts to address the problem of unfairness in contractual relations. The defences of duress and fraudulent misrepresentation both examine the procedural fairness of the process by which the agreement was reached, not the actual substantive fairness of what was agreed. The former looks to the means by which consent was obtained, the latter looks to the conduct of the promisor.

32 Law Reform Commission of Victoria, Discussion Paper No. 10 *Liquidated Damages and Penalties*.
33 *Ibid*. para. 35.
34 N.S.W.L.R. 1, 23.
35 Law Reform Commission of Victoria, *op. cit.* para. 36.
36 Trebilcock, M. J., 'The Doctrine of Inequality of Bargaining Power: Post-Benthamite Economics in the House of Lords.' (1976) 26 *University of Toronto Law Journal* 359, 384.

> The process of formation provides the court with all the information it needs to allow the promisor to escape from performance under the contract, and the court itself can and must remain unconcerned with the substantive terms of the bargain.[37]

These rules will be inefficient and impose error costs if well-founded and deserving cases cannot establish the necessary standard of proof to make out the defence. If the error costs of legitimate contracts being defeated by undeserving defences are lower, then the rules should be relaxed.

Epstein[38] argues that the proper role of unconscionability is to protect against fraud duress and incompetence without demanding specific proof of any of them. To do this it looks at the subject matter of the agreements, the social positions of the parties who enter them and the perceptions of the stronger party. Freedom of contract does not require that all agreements are enforced. However the reasons for intervening and refusing to enforce an agreement must be either some defect in the process of contract formation, or some incompetence of the party against whom the contract is to be enforced. The doctrine of unconscionability (or any other, doctrine, by implication), should not allow the courts to set aside agreements when they find their substantive terms to be objectionable. Rather it should be used only to police the process whereby agreements are formed, so as to facilitate the setting aside of agreements vitiated by fraud, duress or incompetence.

What is Unfairness?

The penalty doctrine is an attempt by the courts to draw an inference of procedural unfairness from the substantive terms of the agreement. This is not legitimate.

> when the doctrine of unconscionability is used in its substantive dimension . . . it serves only to undercut the private right of contract in a manner that is apt to do more a social harm than good.[39]

The question arises, if the penalty doctrine is not the way to supervise negotiation, then what is?

In assessing the procedural fairness of a bargain, the absence of bargaining between the parties and the adoption of a 'take-it or leave-it' attitude by one party is

> evidence not of market power but of a recognition that neither producer, nor consumer interests in aggregate are served by incurring the costs involved in negotiating separately every transaction.[40]

The real measure of market power is the ability of a consumer if he rejects one supplier to turn to a workably competitive range of alternative sources. If a market is workably competitive, any supplier offering uncompetitive standard form terms will have to reformulate his package of price and contractual terms and conditions to prevent the loss of his business to competitors.

37 Epstein, R., 'Unconscionability: A Critical Reappraisal' (1975) 18 *Journal of Law and Economics* 293, 296.
38 *Op. cit.* 302.
39 Epstein, *op. cit.* 315.
40 Trebilcock, *op. cit.* 364.

Thus Trebilcock argues that the fairness of a bargain can be inferred not from the fact that there was bargaining and negotiation over terms between the parties, but from the fact that if the promisee decided to go elsewhere he had a real choice.

A second source of procedural unfairness arises because the weaker party is ignorant, in the sense that he has no information about, or understanding of, the content of his obligation.

This ignorance precludes reflection or a search for alternatives,[41] and results in unfair surprise which is a valid reason to intervene on legal and economic grounds.

> Where a contract is so worded or arranged that the supplier knows or should know that the other party does not understand its implications, and he knows or should know that the other party reasonably entertains other understandings as to its legal incidents, . . . to allow him to sign the contract without correcting those misunderstandings is tantamount to misrepresentation and thus conducive to suboptimal allocative decisions.[42]

Having found that there is no abnormal market power and no aberrations in the process of contract formation, the concept of substantive unfairness[43] poses 'conceptual problems' as almost by definition the outcome of such a process cannot be unfair.

It is my suggestion that the Commission's conclusions should be accepted. The penalty doctrine should be abolished and replaced by a general discretion to set aside agreed damages clauses when they are unconscionable. However, the doctrine of unconscionability must not be allowed to depart from its true purpose. Old habits die hard, and although it may remain as a secondary factor, the temptation to look at the substantive terms of the agreement should be resisted. The agreement reached by the parties should be enforced unless there is evidence that the bargaining process was defective in some way.

SUMMARY AND CONCLUSIONS

Contracts bring transactors together to their mutual benefit. In some cases changing circumstances may mean that a greater gain can be secured if the contract is broken. Liquidated damages are an attempt by the parties to pre-estimate, at formation of the contract, the amount of damages recoverable for breach. We have seen that a number of factors might induce the parties to negotiate a liquidated damages clause: (1) if expected damages are readily calculable but the parties believe that negotiation of a liquidated damages clause will save potential litigation or settlement costs should breach occur, (2) the expected damages are uncertain or difficult to establish and the parties may wish to eliminate the uncertainty of the amount that can be recovered and the expense of litigation to find it by setting liability in advance, (3) parties have different degrees of risk aversion and wish to negotiate liquidated damages so that the less risk-averse party at least partially insures the more risk-averse against loss from breach, (4) one party may attach unusual and subjective value to performance which would be rejected by a court as too fanciful to recover, (5) parties may

41 *Ibid.*
42 Trebilcock *op. cit.* 370.
43 Defined as judicially perceived non-equivalence of values exchanged by contracting parties.

desire a different allocation of the risks of and gains from breach than that which flows from the court's approach.

If the parties are risk-neutral they will only provide for a given contingency (for example, by stipulating a liquidated damages clause in the case of that contingency) if the adverse consequences of failing to provide for the occurrence of that event are sufficient to justify the sure costs of including such provisions and the expected cost of verifying the occurrence. Factors which must be taken into account are the difficulty of reaching agreement over an issue, the probability and magnitude of a particular loss occurring, the cost of verifying the occurrence of an event, and the importance of the transaction to the parties. The introduction of risk means that the allocation of the risk of loss due to this contingency must also be considered. In essence the issue is whether the costs of litigation and the risk of undercompensation exceed the costs of negotiating in advance.

The costs of negotiating a liquidated damages clause are greatly increased by the current uncertainty over what is required for a clause to be enforceable. Negotiation is more difficult as there are no clear guidelines and after breach there is almost always sufficient doubt for the breaching party to challenge the agreed sum on which the injured party seeks to rely. Much of this uncertainty is not due to the law itself but to the judicial practice of applying the law to take other, possibly unenunciated, factors into account — such as the relative 'culpability' of the parties — in reaching their decision. The test of an agreed sum as 'extravagant and unconscionable in relation to the greatest loss that could be suffered' is sufficiently wide to allow a range of interpretations in order to reach a result desired on grounds other than legal doctrine.

The penalty doctrine is a longstanding principle of the general law originally developed to combat unfair bargaining and extortionate 'penalty' clauses. However, with the development of the doctrine of unconscionability and the enactment of trade practices and consumer protection legislation, better ways have evolved to deal with such problems without the cost of striking out some fair and legitimate clauses. As with any attempt to correct a failure of the market, the most efficient approach is to address the problem (*i.e.* unfair clauses) as directly as possible. In this case, this means examining the bargaining process and the relationship of the parties (via the doctrines of duress and unconscionability) rather than looking merely at the end product, the agreed sum.

The adverse effects of the penalty doctrine have been discussed at length. It runs against the principle of freedom of contract not to enforce the fairly bargained agreements and risk allocations reached by the parties. The penalty doctrine may remove the need to prove unfairness, but it also makes it necessary to show what damage could have been suffered, which may be more difficult. Perhaps more importantly, the measure of damage used by the courts may not be fully compensatory. Expectation damages do not allow for the parties' attitudes to risk as they are calculated after breach. Some loss may remain uncompensated or loss may be compensated that should not have been. For example, it is settled law that in an agreement for the payment of money it is a penalty to fix a larger sum as damages for late or non-payment. This not only fails to allow for

consequential loss due to the breach but also does not allow for the allocation of risk. In short, the parties themselves are in the best position to know what value they want to protect and how to allocate risk. Further reasons to enforce liquidated damages are the conflicting roles it plays and the frequently neglected factor that such clauses are 'paid for' in the contract price.

The conflict between the roles of incentive maintenance and the allocation of risk arises because differences in attitude to risk mean that one party can insure the other so that he or she is indifferent between breach or performance. Thus the costs of the breach are not borne by the party responsible for imposing them. Again, the parties can resolve this best among themselves by partial insurance, through liquidated damages clauses, to compromise between the two roles according to their priorities. This is impossible for the court to achieve, after breach, because it lacks the necessary information and because after breach uncertainty is resolved and the parties have incentive to lie to maximize their gains.

In addition, as we have seen, parties wishing to be protected by an agreed damages clause in excess of what a court might award generally have paid a 'premium' on the contract price in return for this extra insurance. Hence they do have an interest in this secondary agreement to see that, if there is breach of the base agreement, they get what they bargained for and the clause is enforced. Currently, of course, such premiums are heavily discounted by the probability that they will not be enforced.

The conclusion to be drawn from all the foregoing considerations is that stipulated damage agreements should be enforced when the parties have comparable information unless there is evidence of duress or unconscionable conduct. That is, it can be shown that one party is at a serious disadvantage to the other and has been exploited by the stronger party in a morally culpable manner, resulting in a transaction that is not just improvident but overreaching and oppressive. This doctrine is a comparatively recent development in the law and addresses the problem much more efficiently and effectively than the penalty doctrine. The courts retain an important role in assessing fairness not only through doctrines such as these but also through their role in rendering agreements invalid in which one party is guilty of falsely alleging or inducing breach. It should be stressed that this conclusion is restricted to fully bargained exchanges between parties of comparable bargaining strength. Consumer transactions are a different situation, beyond the scope of this article, as for the most part they involve little or no bargaining and negotiation.

Any rule of damages will, 'influence resource allocation by affecting the probability that the parties will continue performance of an economically unjustified contract and by changing the way parties allocate the costs of covering various risks'.[44] The best and simplest way to minimize these costs is, subject to the aforementioned policy considerations, to leave it to the parties concerned.

[44] Barton, J. H., 'The Economic Basis of Damages for Breach of Contract' (1972) 1 *Journal of Legal Studies* 270, 282.