

LIQUIDATED DAMAGES, PENALTIES AND THE JUST COMPENSATION PRINCIPLE: SOME NOTES ON AN ENFORCEMENT MODEL AND A THEORY OF EFFICIENT BREACH

CHARLES J. GOETZ* and ROBERT E. SCOTT**

INTRODUCTION

For more than five centuries, strict judicial scrutiny has been applied to contractual provisions which specify an agreed amount of damages upon breach of a base obligation.¹ Although the standards determining the enforceability of liquidated damage clauses have developed novel and labyrinthine permutations,² their motivating principle has remained essentially immutable. For an executory agreement fixing damages in case of breach to be enforceable, it must constitute a reasonable forecast of the provable injury resulting from breach; otherwise, the clause will be unenforceable as a penalty and the non-breaching party will be limited to conventional damage measures.³

The historical genesis of this principle sheds some light on its original rationale. Relief against penalties was one of the earliest exercises of equitable interference, having developed during the fifteenth century when the common law had no adequate machinery for trying cases of fraud.⁴ At a time when legal rules permitted double recovery through the sealed penalty bond,⁵ as well as other recovery grounded in fraud, a presumption by the

* Professor of Law, University of Virginia School of Law; A.B., 1961, Providence College; Ph.D., 1964, University of Virginia.

** Professor of Law, University of Virginia School of Law; B.A., 1965, Oberlin College; J.D., 1968, William & Mary; S.J.D., 1973, University of Michigan.

The authors wish to acknowledge the significant contribution of Mr. Charles G. Cofer, B.A., 1974, Duke University; J.D., 1977, University of Virginia School of Law, in the preparation of this article. Mr. Cofer is primarily responsible for the material concerning the historical background of the penalty rule and also assisted in the preparation of the footnotes.

1. See 5 W. S. HOLDSWORTH, A HISTORY OF ENGLISH LAW 293 (1924). Relief from agreed damage provisions was premised on the notion that it was against "conscience" that a person might recover damages exceeding the loss which he suffered. Undoubtedly, one underlying fear was that fraud or other unconscionable conduct was involved.

2. See text accompanying notes 21-28 *infra*.

3. The rationale for designating a particular liquidated damages provision a penalty has as many formulations as there are treatments of the doctrine. For a representative sampling, see 5 A. CORBIN, CONTRACTS § 1054 (1964); G. GRISMORE, PRINCIPLES OF THE LAW OF CONTRACTS § 234 (rev. ed. 1974); C. McCORMICK, HANDBOOK ON THE LAW OF DAMAGES 599-608 (1935); RESTATEMENT OF CONTRACTS § 339 (1932); 5 S. WILLISTON, A TREATISE ON THE LAW OF CONTRACTS § 775A (3d ed. W. Jaeger 1961); Brightman, *Liquidated Damages*, 25 COLUM. L. REV. 277 (1925).

4. See 5 W. S. HOLDSWORTH, *supra* note 1, at 292-93. According to Holdsworth, the rules of equity at the mid-point of the fifteenth century were very vague. Combined with the turbulence of the country, these vague rules "gave rise to many interferences of the chancellor on the grounds of equity and conscience." *Id.* at 278. Such broad interferences may have been prompted by perceived defects in common law rules. *Id.* at 278-87. Since there were no legal rules available to relieve against unconscionable bargains, equity filled the gap. *Id.* at 292.

5. It was not uncommon that a debtor "for further security" would bind himself for double the amount of the actual debt. "[H]e might pay the debt, and still the creditor, armed

early equity courts that liquidated damage provisions carried an unusual danger of oppression and extortion would have seemed well justified.⁶ In addition, the promisor faced information barriers greatly increasing the risk of overestimating his ability to perform.⁷ Consequently, the equity courts apparently refused enforcement when either actual or presumptive evidence of unfairness indicated that recovery would result in an "unjust, extravagant or unconscionable *quantum* of damages in case of a breach."⁸

The common law courts soon usurped and subtly altered the developing penalty rule, invalidating the agreed remedy in any case where it specified a significantly larger amount than conventional damage recovery.⁹ Applying the principle of "just compensation for the loss or injury actually sustained"¹⁰ to liquidated damage provisions, courts have subsequently refused enforcement where the clause agreed upon is held to be *in terrorem*—a sum fixed as a deterrent to breach or as security for full performance by the promisor, not as a realistic assessment of the provable damage.¹¹ Thus, attempts to secure performance through *in terrorem* clauses are currently declared unenforceable even where the evidence shows a voluntary, fairly bargained exchange.¹²

with his sealed writing, could collect the full sum named in the deed; for the common law received such evidence as conclusive." Barbour, *The History of Contract in Early English Equity*, in 4 OXFORD STUDIES IN SOCIAL AND LEGAL HISTORY 89 (P. Vinogradoff ed. 1914).

6. By the sixteenth and early seventeenth centuries, formalized rules were developed for the exercise of equity jurisdiction in order to give relief against accident and sharp practice. 5 W. S. HOLDSWORTH, *supra* note 1, at 330. The relief against penalties was by this time almost a separate branch of equitable jurisdiction.

7. It is characteristic of men, however, that they are likely to be beguiled by the 'illusions of hope,' and to feel so certain of their ability to carry out their engagements in future, that their confidence leads them to be willing to make extravagant promises and commitments as to what they are willing to suffer if they fail.

C. MCCORMICK, *supra* note 3, § 147, at 601.

8. Thompson, *Penalties and Liquidated Damages*, 46 CENT. L.J. 5 (1898) (emphasis in original).

9. This development did not go unnoticed:

[P]laternalism would be well enough . . . if it limited itself to setting aside, or refusing to enforce, stipulations for unconscionable damages. But, as actually administered in the courts, it does not so limit itself. Proceeding on fictitious rules of interpretation, while pretending to be making an effort to ascertaining [*sic*] the real meaning of the parties, the courts have swamped themselves in all manner of difficulties, floundering in which their judgments result in setting aside fair bargains as to damages in cases where the damages are incapable of assessment, as well as unconscionable or extravagant bargains in cases where the damages are capable of assessment.

Id. at 6.

10. Jaquith v. Hudson, 5 Mich. 123, 133 (1858).

11. Henderson-Boyd Lumber Co. v. Cook, 149 Ala. 226, 42 So. 838 (1906); Muldoon v. Lynch, 66 Cal. 536, 6 P. 417 (1885); Metz v. Clay, 101 Kan. 45, 165 P. 809 (1917); Benfield v. Croson, 90 Kan. 661, 136 P. 262 (1913); H. J. McGrath Co. v. Wisner, 189 Md. 260, 55 A.2d 793 (1947); Quigley v. C. S. Brackett Co., 124 Minn. 366, 145 N.W. 29 (1914); City of Rye v. Public Serv. Mut. Ins. Co., 34 N.Y.2d 470, 315 N.E.2d 458, 358 N.Y.S.2d 391 (1974); Chaude v. Shepard, 122 N.Y. 397, 25 N.E. 358 (1890); Feinsot v. Burstein, 161 App. Div. 651, 146 N.Y.S. 939 (1914).

12. The implications of the penalty doctrine are anomalous in terms of the theoretical underpinnings of modern contract law. The prevailing principles of contract developed coincidentally with the emergence of a market economy in the nineteenth century and reflect the assumption of classical economics that no neutral principle exists whereby the end result of a freely bargained exchange can be measured. The doctrinal concern has, therefore, centered on the necessity of reliable, objective evidence of a voluntary exchange, rather than an examination of the disparity in the utility of the exchange to the contracting parties. See L. FRIED-

The continuing vitality of the penalty doctrine can best be explained by examining the assumptions underlying the application of the "just compensation" principle to liquidated damage provisions.¹³ The central assumption is that enforcing an *in terrorem* clause overcompensates the non-breaching party. The current penalty rule is therefore envisioned as a protection against (1) unfair recovery in excess of justifiable reliance,¹⁴ and (2) the performance of a contract through fear of the penalty, where it would be more efficient economically to breach.¹⁵ The assumption of overcompensation, however, is gratuitous. We argue below that its uncritical application frequently induces a costly reexamination of the initial allocation of risks and may also deny the non-breaching party either adequate compensation for the harm caused by the breach¹⁶ or the opportunity to insure optimally against such harm.

MAN, A HISTORY OF AMERICAN LAW 244-46 (1973); J. W. HURST, LAW AND THE CONDITIONS OF FREEDOM IN THE NINETEENTH CENTURY UNITED STATES 11-12 (1956); Horowitz, *The Historical Foundations of Modern Contract Law*, 87 HARV. L. REV. 917, 918 (1974). Prior to the emergence of a market economy, the law of contract had been based upon socially imposed notions of equity. This was a reflection of the prevailing belief that the justification for contractual obligation could be "derived from the inherent justice or fairness of an exchange." *Id.* at 917. One of these equitable notions was that there was some "intrinsic value" to an exchange which could be recognized objectively by the courts through an application of the principles of justice and fairness. As the growth of commercial markets continued, equitable notions such as "intrinsic value" were substantially eroded. By way of illustration, Professor Horowitz states:

Markets for future delivery of goods were difficult to explain within a theory of exchange based on giving and receiving equivalents in value. Future contracts for fungible commodities could only be understood in terms of a fluctuating conception of expected value radically different from the static notion that lay behind contracts for specific goods; a regime of markets and speculation was simply incompatible with a socially imposed standard of value.

Id. at 946-47.

As a consequence of the bargain theory, traditional fairness constraints have been limited to two broad categories: In the first category are process controls designed to identify mistake, fraud, undue influence, duress or other bargaining abnormalities which challenge the assumption of fairly bargained exchange. This category also includes status controls—infancy and insanity—which challenge the assumption that the contracting parties possessed the capacity to evaluate the exchange. These obvious forms of bargaining unfairness render a contract voidable at the option of the promisor. *See generally* 1 A. CORBIN, *supra* note 3, at 13-15, 635-36. The second category is characterized by limitations on the subject matter of permissible bargains. "Illegal" bargains are proscribed because the subject matter of the exchange is deemed socially detrimental. The categories of illegality are as amorphous and impervious to theoretical analysis as is the underlying concept of public policy. Traditional examples of illegal bargains include those that restrain trade, interfere with the administration of justice, promote fraud or other injury to third parties, or are harmful to other social institutions—the family, marriage or perceived morality. *See generally* 6A *Id.* at §§ 1373-1517.

13. The penalty doctrine continues to be strictly followed. *Garrett v. Coast & S. Fed. Sav. & Loan Ass'n*, 9 Cal. 3d 731, 511 P.2d 1197, 108 Cal. Rptr. 845 (1973) (late payment charge measured against unpaid balance of loan constitutes cause of action for invalidation on ground of penalty); *Plymouth Sec. Co. v. Johnson*, 335 S.W.2d 142 (Mo. 1960) (forfeiture of payments on a cemetery purchase after six months; default held a penalty); *City of Rye v. Public Serv. Mut. Ins. Co.*, 34 N.Y.2d 470, 315 N.E.2d 458, 358 N.Y.S.2d 391 (1974) (bond provision specifying liquidated damages for delay in contract with city constitutes an unauthorized penalty); *American Financial Leasing & Servs. Co. v. Miller*, 41 Ohio App.2d 69, 322 N.E.2d 149 (1974) (liquidated damage provision in service station equipment lease held to be a penalty); U.C.C. § 2-718(1), Comment 1; RESTATEMENT OF CONTRACTS § 339, Comment h to subsection (2) (1932).

14. *See Macneil, Power of Contract and Agreed Remedies*, 47 CORNELL L. Q. 495, 499-501 (1962); note 30 *infra*.

15. *See R. POSNER, ECONOMIC ANALYSIS OF LAW*, 55-60 (1973); Hartzler, *The Business and Economic Functions of the Law of Contract Damages*, 6 AM. BUS. L. J. 387, 392-93 (1968).

16. Over the years, occasional critics have challenged the validity of the assumption of

Part I of this Article examines the role of liquidated damage provisions and penalties in the context of a general theory of efficient breach of contract.¹⁷ The proof problems inherent in fully recovering idiosyncratic values within the context of operationally practical damage sanctions may prevent the non-breaching party from recovering his subjective expectations if recovery is limited to legally determined remedies. The expected cost of establishing true losses under conventional damage measures will thus induce parties who face uncertain or nonprovable anticipated losses to negotiate stipulated damage agreements. The current penalty rule subjects these agreements to costly review, based not on the fairness of the process, but on whether the initial estimate sufficiently mirrors the anticipated provable loss.

Part II examines the hypothesis that, absent evidence of process unfairness in bargaining, efficiency will be enhanced by the enforcement of an agreed allocation of risks embodied in a liquidated damage clause.¹⁸ We argue that agreed damage measures and *in terrorem* provisions represent, under many circumstances, the most efficient means by which parties can

over-compensation. Barton, *The Economic Basis of Damages for Breach of Contract*, 1 J. LEGAL STUD. 277, 286-87 (1972) (penalty principle infringes upon freedom to negotiate distribution of risks resulting from breach); Thompson, *supra* note 8, at 6-7 (fictitious rule of interpretation is a judicial abridgment of freedom of contract); C. McCORMICK, *supra* note 3, § 147; Brightman, *supra* note 3, at 279-81. Professor Brightman observed in 1925 that where the courts have invalidated agreed remedies because of an *in terrorem* clause they have often worked unjust results. His inference of unfairness is based on the premise that the court has deprived one of the parties of his bargained-for exchange on the unsupported ground that his motive for the bargain violates public policy. *Id.* at 283.

17. The jurisprudential bases of contract have undergone severe challenges in the last generation. See generally P.S. ATIYAH, CONSIDERATION IN CONTRACTS: A FUNDAMENTAL RESTATEMENT 27-45 (1971); L. FRIEDMAN, CONTRACT LAW IN AMERICA 15-18, 20-24 (1965); Macneil, *The Many Futures of Contracts*, 47 S. CAL. L. REV. 691 (1974). Although the obituary of classical contract theory has only recently been published in Professor Gilmore's provocative essay, THE DEATH OF CONTRACT (1974), its death was presaged by the widespread acceptance of the triumph of legal realism, the Uniform Commercial Code. The non-static conception of promissory liability has destroyed the theoretical efficacy of a formal doctrine of consideration as a necessary and sufficient theory of obligation. Recently, however, several scholars have pointed perceptively to sparks of life remaining in the corpse. For particularly thoughtful essays on the continued vitality of contract jurisprudence, see Speidel, *An Essay on the Reported Death and Continued Vitality of Contract*, 27 STAN. L. REV. 1161 (1975); Henderson, Book Review, 124 U. PA. L. REV. 1466 (1976).

In response to the theoretical limitations of a formal doctrine of consideration as well as the open-ended realism of commercial reasonableness, some have suggested the application of the efficiency criterion. See, e.g., Birmingham, *Damage Measures and Economic Rationality: The Geometry of Contract Law*, 1969 DUKE L.J. 50; Hartzler, *supra* note 15, at 393; Birmingham, *Breach of Contract, Damage Measures and Economic Efficiency*, 24 RUTGERS L. REV. 273 (1970); R. POSNER, *supra* note 15; Barton, *supra* note 16.

18. The hypothesis which forms the basis for the enforcement model presented here was first proposed without significant elaboration by Professor Barton in his article exploring the economic bases of damages measures in the absence of clear market alternatives. See Barton, *supra* note 16. In particular, Professor Barton recognizes that there may be certain subjective expectations that a party to a contract might wish to protect. *Id.* at 281. Allowing the parties to arrange a set of damage rules in the form of liquidated damage provisions may provide such protection. Barton attributes the invalidation of many liquidated damage provisions to a general reluctance on the part of courts to accept damage rules set by the parties. *Id.* at 282-87. See also Brightman, *supra* note 3, at 279-83.

Professor Brightman also reached the conclusion that liquidated damages clauses, unless shown to be unconscionable, should be enforced. It is not clear, however, whether he advocated such a rule to protect any "idiosyncratic" value of the non-breaching party, or only to protect the value of the agreement that the usual laws of compensation identify. See *id.* at 283.

insure against the otherwise non-compensable consequences of breach. Our hypothesis is then also tested against a series of conditions in order to identify alternative legal principles that may provide less costly means of avoiding the harmful effects whose perception apparently prompted the current penalty rule.

I. THE JUST COMPENSATION PRINCIPLE: A THEORY OF EFFICIENT BREACH

A. *The Role of Damage Agreements Under the Just Compensation Standard*

The modern law of contract damages is based on the premise that a contractual obligation is not necessarily an obligation to perform,¹⁹ but rather an obligation to choose between performance and compensatory damages.²⁰ Once a contemplated exchange has been negotiated, the breaching party is merely required to provide "just compensation" equal to the value of performance. The implications of this rule can be usefully illustrated through the principles of economic analysis. Generally, breach will occur where the breaching party anticipates that paying compensation and allocating his resources to alternative uses will make him "better off" than performing his obligation. As long as the compensation adequately mirrors the value of performance, this damage rule is "efficient." It induces a result superior to performance, since one party receives the same benefits as performance while the other is able to do even better. Under the current damage rule, all of these net gains from breaching are retained by

19. One purpose of contract damages which has been suggested is that they tend to prevent breach. The fact that damages must be paid upon breach "makes, therefore, for the security of business transactions and helps to make possible the vast structure of credit, upon which so large a part of our modern prosperity depends." 5 A. CORBIN, *supra* note 3, at 34. This analysis is, however, belied by the absence of punitive measures designed to reduce the incentives to breach. On the contrary, it is clear that the central purpose of contract damages is compensatory and not punitive. See, e.g., Farnsworth, *Legal Remedies for Breach of Contract*, 70 COLUM. L. REV. 1145, 1147 (1970); Speidel & Clay, *Seller's Recovery of Overhead Under UCC Section 2-708(2): Economic Cost Theory and Contract Remedial Policy*, 57 CORNELL L. Q. 681, 683-85 (1972). Indeed, a strong argument can be made that the theory of damages is designed to err toward undercompensation. For example, costs, attorneys' fees and certain categories of pre-judgment interest are not recoverable by the non-breacher. Although the economic bases of the compensation principle have not been fully articulated, it is clear that the principle is, at least intuitively, consistent with a theory of efficient breach—providing incentives for the promisor to breach a contract which has become inefficient *ex post*, compensate the promisee for his expected gain, and reallocate his resources to more highly valued uses. The incentives are provided by permitting the promisor to retain all the gains from breach. See generally Birmingham, *supra* note 17, at 273; Hartzler, *supra* note 15, at 388-97; Barton, *supra* note 16; R. POSNER, *supra* note 15, at 55-59.

20. Nowhere is the confusion between legal and moral ideas more manifest than in the law of contract. Among other things, here again the so called primary rights and duties are invested with a mystic significance beyond what can be assigned and explained. The duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it—and nothing else. If you commit a tort, you are liable to pay a compensatory sum. If you commit a contract, you are liable to pay a compensatory sum unless the promised event comes to pass, and that is all the difference.

Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 462 (1897).

the breacher—the non-breacher is in the same position as if there had been performance. In order to maintain the efficiency value of the rule, however, it is only necessary that some minimal amount of benefits are retained by the breacher in order to induce him not to perform. The allocation of the gains from breach is, therefore, largely a question of wealth transfer between the contracting parties.

Facing this conventional damage measure, contracting parties have incentives to negotiate liquidated damages clauses whenever the costs of negotiating are less than the expected costs resulting from reliance on the standard damage rule for breach. There are two primary factors which might induce the decision to negotiate:

- (1) The expected damages are readily calculable, but the parties determine that advance stipulation will save litigation or settlement costs;
- (2) The expected damages are uncertain or difficult to establish and the parties wish to allocate anticipated risks.

Of course, these factors may be present singly or in combination.

Pre-breach agreements will not be legally enforceable, however, unless two requirements coincide. First, the agreement must be a reasonable forecast of just compensation for the anticipated harm that would be caused by the breach.²¹ Second, the possible damages which might result from the breach must be uncertain and difficult to estimate.²² However, liquidated damages provisions have seldom been voided solely because the damages were easy to estimate.²³ Instead, courts have considered the degree of uncertainty an influential factor in determining the reasonable-

21. RESTATEMENT OF CONTRACTS § 339(1)(a), Comment on subsection (1) (1932); U.C.C. § 2-718(1), Comment 1 (1961 version); 5 A. CORBIN, *supra* note 3, § 1059; C. MCCORMICK, *supra* note 3, § 149. Most courts correctly assess the reasonableness of the stipulated damage provision *ex ante*, *i.e.*, as of the time at which it was made. *McCarthy v. Tally*, 46 Cal. 2d 577, 297 P.2d 981 (1956); *Better Food Mkts. v. American Dist. Tel. Co.*, 40 Cal. 2d 179, 185, 253 P.2d 10, 14 (1953). For an application of the requirement, see *Gorco Constr. Co. v. Stein*, 256 Minn. 476, 99 N.W.2d 69 (1959).

22. RESTATEMENT OF CONTRACTS § 339(1)(b), Comments on subsection (1) (1932); 5 A. CORBIN, *supra* note 3, § 1060; C. MCCORMICK, *supra* note 3, § 148; J. CALAMARI & J. PERILLO, *THE LAW OF CONTRACTS* § 232 (1970). As with the reasonableness requirement, this difficulty in estimating damages must be present at the time of contracting. *Parker-Washington Co. v. City of Chicago*, 267 Ill. 136, 140-41, 107 N.E. 872, 874-75 (1915). This uncertainty requirement has been criticized, especially by Professor McCormick:

If the agreement reasonably approximates the probable damages which a court would give, why forbid the parties thus to agree, even if the damage could be readily and exactly foreseen? If they can by agreement thus relieve the judge or jury of the task of fixing damages, even though that task should be an easy one and its result could readily be forecast, there seems to be no sound objection to their doing so.

C. MCCORMICK, *supra* note 3, at 605. McCormick contends that the ease or difficulty of predicting damages in advance of breach neither increases nor decreases the chances of an oppressive bargain. Therefore, there is little reason to depart from the general rule that courts will enforce the agreement of the parties according to its terms. *Id.* *But see* Macneil, *supra* note 14, at 501-03.

23. *See generally* J. CALAMARI & J. PERILLO, *supra* note 22, § 232; C. MCCORMICK, *supra* note 3, at 605-06 (discussion of cases). *But see* *Garrett v. Coast & S. Fed. Sav. & Loan Ass'n*, 9 Cal. 3d 731, 511 P.2d 1197, 108 Cal. Rptr. 845 (1973).

ness of the estimate.²⁴ If the conditions inducing damage agreements are viewed on a continuum, the application of the penalty rule becomes clearer: as the uncertainty facing the contracting parties increases, so does their latitude in stipulating post-breach damages.²⁵

The threat of subsequent review clearly increases the costs of negotiating a damages clause relative to relying on the standard damages rule. Are these costs accompanied by counterbalancing advantages? The traditional justification for post-breach inquiry is prevention of "unjust" punishment to the breacher, *i.e.* compensation exceeding the harm actually caused.²⁶ This justification has been expressed in two distinct forms. One basis for invalidation is the presumption of unfairness: liquidated damage provisions are unreasonable—a penalty—whenever the stipulated sum is so disproportionate to provable damages as to require the inference that the agreement must have been effected by fraud, oppression, or mistake.²⁷ The other

24. [I]n proportion as the difficulty of ascertaining the actual damage by proof is greater or less, where this difficulty grows out of the nature of such damages, in the like proportion is the presumption more or less strong that the parties intended to fix the amount [reasonably].

Jaquith v. Hudson, 5 Mich. 123, 138 (1858). See also W. HALE, HANDBOOK ON THE LAW OF DAMAGES 214 (2d ed. 1912).

25. It appears that the drafters of the Uniform Commercial Code have tacitly adopted this approach. Section 2-718(1) of the U.C.C. allows parties to liquidate damages for breach as long as the amount stipulated is "reasonable." The reasonableness of a particular amount is determined, in part, by the "difficulties of proof of loss" from the breach. While it might be argued that the U.C.C. rule approximates the common law uncertainty requirement, as does the 1 NEW YORK LAW REVISION COMMISSION, REPORT OF THE LAW REVISION COMMISSION FOR 1955, STATE OF NEW YORK, STUDY OF THE UNIFORM COMMERCIAL CODE 581-82, it appears that a change has been made. The language of U.C.C. § 2-718 itself treats "uncertainty" as merely one factor, and not even a required one, of many to be considered in determining reasonableness.

In addition to uncertainty, courts have also been influenced by the relationship between the stipulated amount and the provable harm *actually* caused by the breach. Although a number of courts have refused to enforce agreements because of the absence of provable losses upon breach, many cases have held that actual loss is irrelevant except as it permits inferences concerning the reasonableness of the agreements viewed *ex ante*. Frick Co. v. Rubel Corp., 62 F.2d 765, 767-68 (2d Cir. 1933); *In re Lion Overall Co.*, 55 F. Supp. 789 (S.D.N.Y. 1943), *aff'd sub nom.* United States v. Walkof, 144 F.2d 75 (2d Cir. 1944); Bryon Jackson Co. v. United States, 35 F. Supp. 665 (S.D. Cal. 1940); McCarthy v. Tally, 46 Cal.2d 577, 297 P.2d 981 (1956). *But see* Rowe v. Shehyn, 192 F. Supp. 428 (D.D.C. 1961); Marshall v. Patzman, 81 Ariz. 367, 370-71, 306 P.2d 287, 290-91 (1957); Gorco Constr. Co. v. Stein, 256 Minn. 476, 481-84, 99 N.W.2d 69, 74-76 (1959). See generally Macneil, *supra* note 14, at 504-509; Sweet, *Liquidated Damages in California*, 60 CALIF. L. REV. 84, 131-33 (1972).

26. RESTATEMENT OF CONTRACTS § 339 (1)(a), Comment on subsection (1), Comment a (1932).

27. This rule was originally outlined in *Sun Printing & Publishing Ass'n v. Moore*, 183 U.S. 642 (1901). In applying the *Sun Printing* rule to a later case, the Supreme Court stated: [E]ffect will be given to the [liquidated damage] provision, as freely as to any other, where the damages are certain in nature or amount or are difficult of ascertainment or where the amount stipulated for is not so extravagant or disproportionate to the amount of property loss, as to show that compensation was not the object aimed at or as to imply fraud, mistake, circumvention or oppression. There is no sound reason why persons competent and free to contract may not agree upon this subject as fully as upon any other, or why their agreement, when fairly and understandingly entered into with a view to just compensation for the anticipated loss, should not be enforced.

Wise v. United States, 249 U.S. 361, 365 (1918).

The unfairness rule has also been applied in *United States v. Bethlehem Steel Co.*, 205 U.S. 105, 119 (1907); and *Gruschus v. C. R. Davis Contracting Co.*, 75 N.M. 649, 655, 409 P.2d 500, 504 (1965). See also *Thompson v. St. Charles County*, 227 Mo. 220, 235-41, 126 S.W. 1044, 1049-50 (1910).

major basis for invalidating agreed remedies is that, since the courts set damages based upon the principle of just compensation, parties should not be allowed to recover more than just compensation from the courts through a privately concocted alternative arrangement, even one fairly negotiated.²⁸

The common theme of these decisions is that a disproportion between the stipulated and the anticipated damage justifies an inference of overcompensation. In turn, overcompensation implies either bargaining unfairness or an objectionable *in terrorem* agreement to secure performance.²⁹ This line of reasoning suggests two benefits which may be expected from the current rule invalidating penalties.³⁰ First, the cost of identifying unfairness may be reduced by a standard rule-of-thumb based on dispropor-

28. The relationship between just compensation and liquidated damages was first articulated clearly in *Jaquith v. Hudson*, 5 Mich. 123, 133 (1858).

29. In some cases, the two invalidating rationales have been entwined. For example, in *Wise v. United States*, 249 U.S. 361 (1918), the Supreme Court found agreed remedies to be reasonable unless they are so disproportionate to provable damages "as to show that [just] compensation was not the object aimed at or as to imply fraud, mistake, circumvention or oppression." *Id.* at 365 (emphasis added). This bifurcated approach has produced conceptual confusion. Some cases identify the compensation limitation as the primary basis of invalidation. *Better Food Mkts., Inc. v. American Dist. Tel. Co.*, 40 Cal. 2d 179, 186-87, 253 P.2d 10, 15 (1953); *Benfield v. Croson*, 90 Kan. 661, 136 P. 262 (1913); *H. J. McGrath Co. v. Wisner*, 189 Md. 260, 55 A2d 793 (1947). Other decisions isolate unfairness as the appropriate decision rule, and implicitly view *in terrorem* provisions as a type of unconscionable conduct. *Equitable Lumber Corp. v. IPA Land Dev. Corp.*, 38 N.Y.2d 516, 522-23, 344 N.E.2d 391, 395-97, 381 N.Y.S.2d 459, 462-64 (1976); *Waggoner v. Johnston*, 408 P.2d 761, 769-70 (Okla. 1965). But the prevailing tendency, as represented by the Uniform Commercial Code, is to specify a reasonableness requirement without indicating an invalidating rationale. U.C.C. § 2-718(1) provides that

Damages for breach by either party may be liquidated in the agreement but only at an amount which is reasonable in the light of the anticipated or actual harm caused by the breach, the difficulties of proof of loss, and the inconvenience or nonfeasibility of otherwise obtaining an adequate remedy. A term fixing unreasonably large liquidated damages is void as a penalty.

Comment 1 to § 2-718 indicates that unconscionability is entwined with the notion of reasonableness and that §§ 2-302 and 2-718 might overlap in certain cases. However, the enumeration of the elements to be considered in determining the reasonableness of a liquidated damage clause seems designed to implement as well the compensation limitation. At least one court has apparently interpreted § 2-718(1) as embracing a dual standard. *See Equitable Lumber Corp. v. I.P.A. Land Corp.*, 38 N.Y.2d 516, 344 N.E.2d 391, 381 N.Y.S.2d 459 (1976).

30. The traditional justification has been that penal clauses violate public policy because contracting parties, overly optimistic about their capacity to perform obligations, will be subject to severe hardship. *C. McCORMICK, supra* note 3, at 601.

Recently, additional justifications have been proposed which require more careful analysis. Professor Macneil has developed the thesis that the criterion for enforcing any agreed remedy should be whether it fills an inadequacy of the law and protects the reliance interests of the promisee. Macneil, *supra* note 14, at 498. Macneil, however, accepts the premise that *in terrorem* provisions violate public policy. He concludes that since the parties to a contract realize that *in terrorem* clauses will not be enforceable, they do not place any reliance upon them. Therefore, by refusing enforcement the courts do not deprive the parties of any reliance interest. Relying upon logic which is not wholly free from circularity, Macneil finds no impairment of freedom of contract, because there is no reliance interest to be protected by the agreed remedy. *Id.* at 499-500. Macneil does not reject the possibility that some *in terrorem* provisions may be enforceable. He states that

[I]f there should be substantial provable or hidden reliance on the 'penalty' clause, *i.e.*, the promisee thinks at the time that the deal is made, that the base promise will not be performed, then we have alternative base promises, rather than simply an additional sanction for performance of one base promise. If there is such reliance on the alternative promise, we are no longer dealing with a penalty but with alternative contracts, and failure to enforce the alternative promise will result in an infringement of the power of contract.

Id. at 500. For a discussion and criticism of the "alternative contracts" concept, see text accompanying notes 60-67 *infra*.

tion.³¹ Second, an enforceable *in terrorem* clause might discourage promisors from breaching and reallocating resources where changed circumstances would ordinarily create efficiency gains from this behavior.³² Inducing performance under these conditions is a misallocation which prevents the net social gain that would result from nonperformance.

As the efficient damages model formalized below will demonstrate, however, this analysis incorrectly assumes that, rather than negotiating out of the penalty, the promisor who is subject to an *in terrorem* clause will inevitably undertake an inefficient performance.³³ In addition, there is no basis for the apparent assumption that the premium placed by the promisee on performance is valueless. Indeed, the market paradigm on which the compensation standard is based requires a contrary presumption; a promisee has a recognizable utility in certain *in terrorem* provisions and this utility is frequently reflected in willingness to pay a price for such clauses.

Clearly, the effects of the current penalty rule cannot adequately be assessed until its underlying assumptions are articulated. In order to develop the efficiency value of agreed remedy provisions within the basic compensation standard, it is first convenient to formalize an analytic model in which "just compensation" for nonperformance can be given precise meaning.

B. *The Efficient Damages Model*

Critiquing the overcompensation assumption requires an analysis of the effects of a penalty rule on the costs of enforcing the compensation standard. Potential enforcement costs include the direct costs of litigation as well as those error costs generated by legal rules which fail to mirror accurately the behavior to be regulated. Formalizing an analytic model in a "friction free" or costless environment provides a useful first step in elevating these effects.

31. See the discussion of the benefits of precision in legal rulemaking in note 55 *infra*.

32. Hartzler notes that enforcing penalty clauses might also have adverse effects on the bargaining process and contract formation. He assumes that contract law provides incentives to contract formation by protecting expectation interests through the award of damages. From this, he observes that damage awards that exceed the "just compensation" amount go beyond the realm of protecting the nonbreaching party's expectations. One result of this excessive protection of expectation damages is that it "may introduce a deterrent to the very contract making behavior to be encouraged." Hartzler, *supra* note 15, at 392. According to Hartzler, this deterrent would have three inefficient manifestations:

- (1) instances in which insistence on an *in terrorem* provision results in a refusal to contract, because the promisor is unwilling to assume the risk of breach;
- (2) instances in which the penalty itself becomes an independent contract objective, resulting in an artificial allocation of risks; and,
- (3) instances in which the promisor may contract below his capacity in order to avoid breach and the subsequent imposition of the penalty, because he perceives the penalty as a real threat.

Id.

The difficulty with this analysis is that it assumes that the expectancy created by the inclusion of an *in terrorem* clause in the contract is not a "valid" interest upon which to predicate an award of damages. This assumption is inconsistent with the premises underlying the bargain theory of contract.

33. See text accompanying notes 38-42 *infra*.

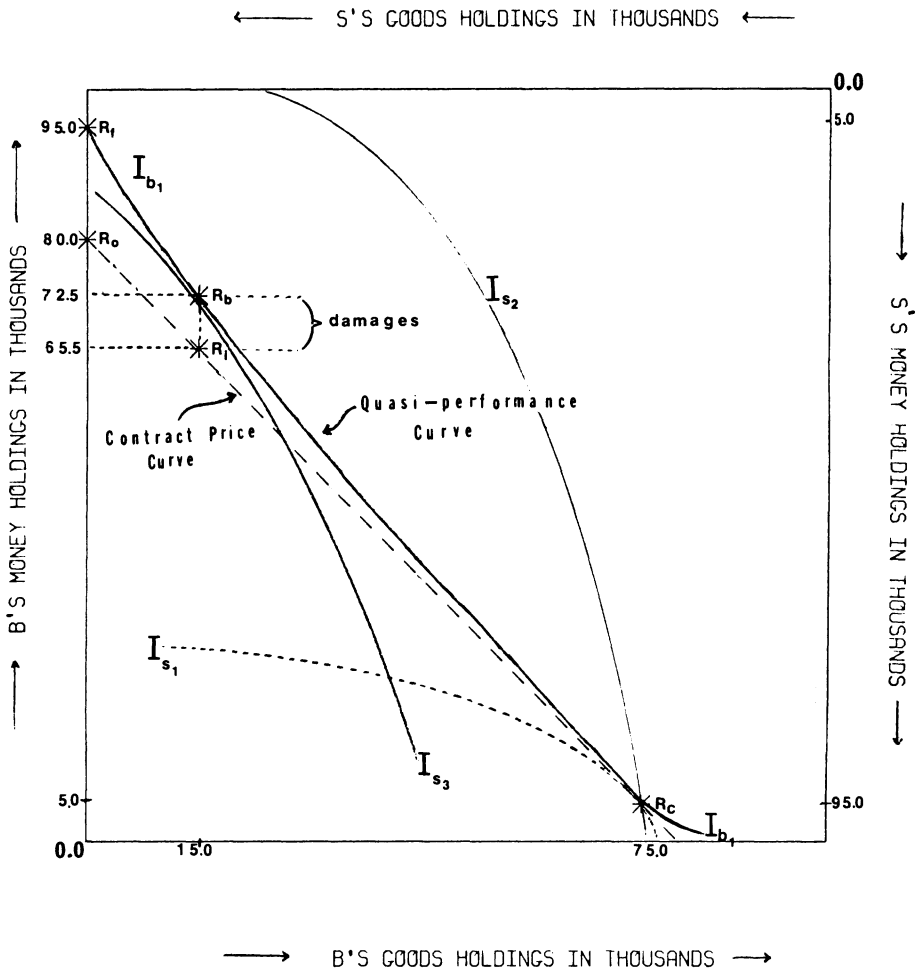


FIGURE 1A

1. *Efficient Contract Reallocation Under the Compensation Standard.* The rationale of modern contract theory can usefully be articulated in terms of the "indifference curve" analysis commonly used in economic theory. Figure 1A is an adaptation of the economic trading model familiarly known as the "Edgeworth Box."³⁴ The potential trade partners, B and S,

34. For a useful explanation of the theory of contract damages in terms of indifference curves and the Edgeworth Box, see Birmingham, *Damage Measures and Economic Rationality: The Geometry of Contract Law*, 1969 DUKE L.J. 50.

Basically, indifference curve maps may be thought of as mapping a "preferredness mountain" wherein higher "elevations" represent the greater preferredness of outcomes. The points in any single indifference curve constitute outcomes having equal preferredness, *i.e.* results among which the individual in question would profess indifference or a state of equal satisfaction. When indifference curves are used in mapping a preference "mountain," their significance is exactly analogous to the equal-evaluation "contour lines" on a standard topographical map. The indifference curve labelled I_{b_1} is a plausibly drawn contour line for B because his more preferred results (closer to his mountain's "peak") are presum-

are assumed to have started with money holdings of \$80,000 and \$20,000 respectively, for an aggregate of \$100,000. Also, B is assumed to have a zero stock of "goods," the product being traded, while S has 100,000 units of goods. By measuring B's holdings from an origin at the southwest corner of the box and S's holdings from the northeast corner, we can represent, as a *point* in the box, any possible division of the aggregates of money and goods between B and S. For instance, the initial situation is represented by R_0 , where B has 0 goods (bottom scale) and \$80,000 (left-hand scale).³⁵

A contract between B and S would be representable as an agreement to move from the original or status quo ante result R_0 to a new result such as R_c . The initial contractual equilibrium of this result is represented in Figure 1A by the tangency of seller's indifference curve (I_{s_1}) and buyer's (I_{b_1}) at R_c ; only at R_c can each party reach the level of preferredness represented by these two curves. The expected post-contract result R_c implies that B will give up \$75,000 in money holdings, with an exactly equivalent receipt to S, in exchange for S's delivery of 75,000 units of goods to B.

Presumably, S and B agreed to this contract initially in the expectation that result R_c was, for both of them, preferred to the original result R_0 . However, the motivation for breach will ordinarily be that the perceived advantages of an agreement at the time of contract have been modified by changed conditions. Hence, if a potential breach is to be analyzed, we should expect to find that one of the parties now "regrets" the promise to move from R_0 to R_c .

As Figure 1A is drawn, changed conditions have shifted the seller's indifference curve (I_{s_2}) back toward the northeast origin.³⁶ The seller now perceives the agreed upon post-performance result (R_c) as an inferior outcome when compared to the status quo ante (R_0) and thus an incentive to breach exists.³⁷ In terms of Figure 1A, the just compensation principle of modern contract theory is representable by B's indifference curve I_{b_1} . This curve not only indicates solutions equivalent (equally preferred by B)

ably in the northeast direction where B's stocks of money and goods are increasing. Exactly the opposite reasoning applies for S, whose "peak" of preferredness is approached by moving toward the southwest, where S's money and goods stocks are maximized. Conceptually, an indifference curve can be drawn through any point on the map of possible results in the Edgeworth Box. However, in order to avoid excessive clutter of the diagrams, it is customary to provide only the indifference curve segments actually necessary to understand the general contours of the individual's preferences about outcome. In this application, it will be necessary to remember only that (1) for B, any point southwest of his indifference curve I_{b_1} is inferior to the promised performance of R_c and (2) for S, any point southwest of his indifference curve I_{s_1} through R_c is preferred to performance while any point northwest is inferior.

35. Note that if we look at the situation in terms of S, the point R_0 is unchanged. It represents S's holding of \$20,000 and 100,000 units of goods.

36. It should be emphasized that conditions inducing breach need not always be represented by a "shiftback" in the seller's indifference curve. It could equally occur where the buyer's curve shifted. The essential condition to breach is that the changed circumstances indicate that the parties' indifference curves are no longer tangent at the contract point (R_c).

37. The movement from R_0 to R_c is a movement down the preferredness surface for the seller. It places S on an indifference curve (I_{s_2}) which is northeast of R_0 and thus inferior to R_0 in S's eyes. See note 34 *supra*.

to full performance at R_c , but also identifies end results southwest of I_{b_1} which would fail to provide the buyer with an outcome meeting his original "expectations" of performance at R_c . Since I_{b_1} does identify results which are, from B's standpoint, interchangeable with full performance (R_c), this important indifference contour will be termed the "quasi-performance" curve. The concept of quasi-performance expresses the fact that any point on the curve places the buyer in the same preferredness situation *as if* performance had taken place.

As Figure 1A is drawn, full breach by the seller would require that S pay B \$15,000 in damages in order to move the buyer from the non-performance solution R_0 to the appropriate point R_f on the quasi-performance curve. Point R_f is the appropriate solution precisely because it supplies B with the equivalent of performance when no goods at all are delivered. Figure 1A is also drawn to reflect the fact that S can do even better by tendering partial performance of 15,000 units, thus creating the post-compensation result of R_b . Result R_b , permitting the seller to move to a new higher indifference curve I_{s_3} , is preferable to both R_c and R_f . It is the most preferred readjustment available to S since the law guarantees to B at least some point on the quasi-performance curve. Not only is solution R_b unambiguously superior to performance at R_c , but it also represents an "efficient" end result in the sense that there is no movement from R_b possible without making at least one party worse off.³⁸ In reaching R_b , the seller would collect the original contract price of \$1 per unit for the 15,000 unit partial performance, but would then be required to remit to the buyer \$7,500 in damages in order to provide B with his quasi-performance expectation at R_b rather than the no compensation result R_f .

This efficient damage model indicates that upon total or partial non-performance the seller's obligation is to provide "just compensation" sufficient to place the buyer on his quasi-performance curve. Since the points along this curve represent the buyer's *subjective* preferences between performance and compensation, "just compensation" may require consideration of nonobjectifiable elements in determining appropriate compensation to the non-breaching buyer. In the costless environment represented by the model, this personal and potentially subjective compensation is readily identifiable. Consequently, legal damages would always be able to provide subjective compensation to the non-breacher, and the parties would never be induced to negotiate a liquidated damages agreement. As the discussion below demonstrates, the incentive for these agreements arises only when costs are reintroduced into the analysis.³⁹

38. The point R_b is superior as Figure 1A is drawn because it makes the seller better off by putting him on indifference curve I_{s_3} while retaining the buyer on his indifference plateau I_{b_1} . The result is also "optimal" in the sense that any movement away from R_b would necessarily produce an inferior result for one of the parties.

39. See text accompanying notes 43-58 *infra*.

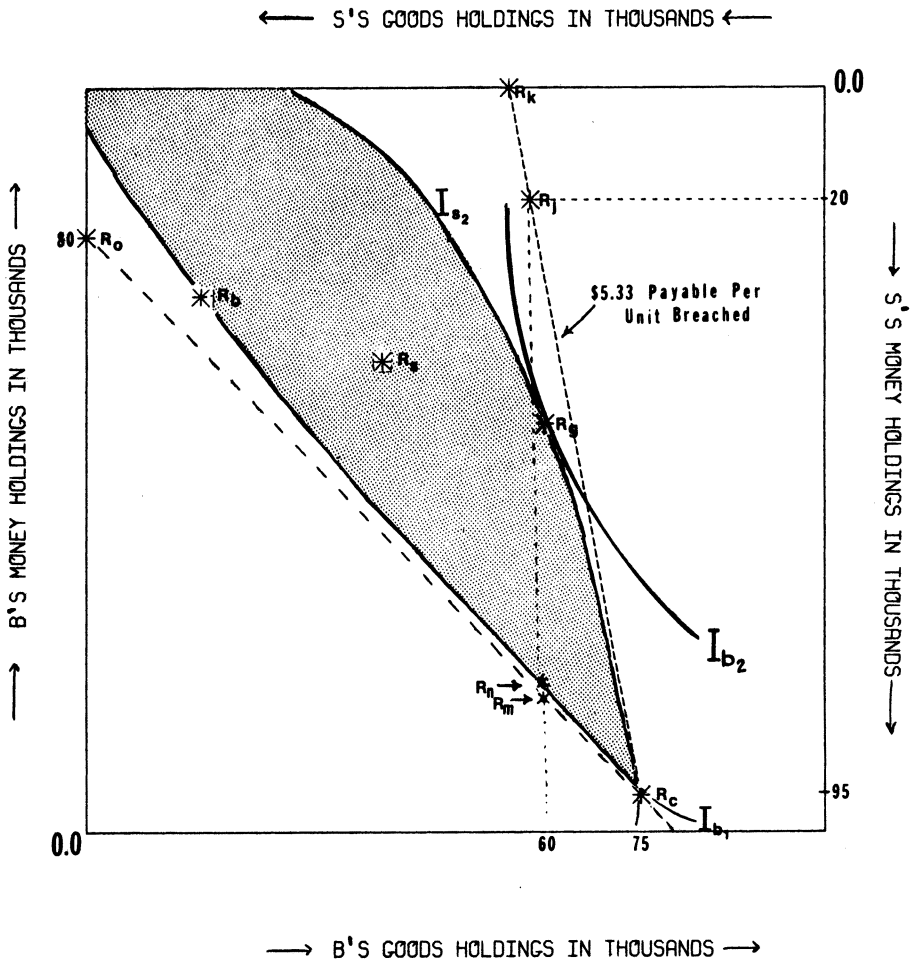


FIGURE 1B

2. *The Operation of a Penalty Provision.* We now define a “penalty” as damages imposed over and above the “just compensation” represented by the non-breaching party’s quasi-performance curve. If paid, penalty damages will place the non-breaching party in a position actually preferable to performance, i.e., “higher” on the preference map than curve I_{b1} of Figure 1A. For instance, assume that S had agreed to pay B non-performance damages at the rate of \$5.33 per unit. In effect, this agreement would replace the “ordinary” damages constraint I_{b1} with the negotiated “penalty” constraint represented on Figure 1B by dashed line R_kR_c . If S chooses any level of non-performance, the penalty component of the damages is the vertical difference between the quasi-performance curve, which denotes the compensatory standard, and the negotiated remedy rate embodied in line R_kR_c . For instance, if S delivers only 60 thousand units, he

must pay \$80,000 ($\$5.33 \times 15,000$) damages, represented by the vertical distance R_m to R_j . Of this amount, compensatory damages correspond to the vertical distance R_m to R_n while penalty damages are exactly as indicated by the vertical distance R_n to R_j .

Figure 1B has deliberately been constructed to show why penalty damages may be considered objectionable. If B insisted on the enforcement of the penalty provision, S would choose to perform fully. Yet, consider any point in the shaded area bounded by the quasi-performance curve (I_{b_1}) and the seller's own indifference curve (I_{s_2}) running through the full performance point R_c . Movement from R_c to any point in the shaded area will represent an unambiguous improvement over R_c in the sense that at least one party will *prefer* the new solution to R_c and the other party will find it at least as good as R_c . Under these circumstances, if the penalty provision really did cause the solution to "stick" at R_c , enforcement of the penalty would lead to results clearly inferior to allowance of the orthodox recovery standard which would motivate movement to a superior point such as R_b .⁴⁰

The situation summarized in Figures 1A and 1B also indicates, however, that there will be an incentive for B to renegotiate his penalty rights whenever the enforcement of these rights leads to an inefficient end result. For instance, B would likely not insist on the penalty provision of \$5.33 per unit but might offer a revised sliding scale of damage rates designed to induce S to partially breach and to slide along S's own performance-equivalent curve I_{s_2} toward solution R_g . This would enable B to move to his superior indifference plateau I_{b_2} at no disadvantage to S. Realistically, R_g is only a limiting solution, since it gives *all* the efficiency gains from breach to B, thus depriving S of any real incentive to breach. However, any infinitesimally lower scale of damages giving some of the benefit of the breach to S might induce S to move to a solution which approaches R_g in the limit.⁴¹

In sum, the principle embodied in present law of giving the non-breaching party "just compensation" or quasi-performance leads to one limiting result, in which all the efficiency gains go to the breacher. But the payment of substantial overcompensation, *i.e.*, penalties, is not necessarily incompatible with an equally efficient result in the neighborhood of the other limiting solution R_g where (infinitesimally less than) all of the efficiency gains from breach go to the non-breaching party. Ranging between these extremes, there are an infinite number of other efficient breach solutions where both parties divide the gains from breach, such as R_s , all of which embody varying combinations of compensatory and "penalty" damages. Obstinate insistence on the enforcement of certain penalties may result in a failure to exploit potential efficiency gains by inducing the

40. See text accompanying note 38 *supra*.

41. In turn R_g is an efficient allocation, "unimprovable" in the sense that there is no possibility of increasing one party's interest except at the cost of a loss to the other because it is a point at which the indifference curves of the two parties are tangent.

penalized party not to breach. However, the very existence of such unexploited gains acts as an incentive to the holder of the penalty rights to renegotiate the penalty provision in question. Hence, the existence of an overcompensation provision is never per se evidence of an efficiency impediment. Absent significant negotiation costs, the stipulation of a penalty still permits alternative post-breach efficient solutions in which the efficiency gains are divided between the breacher and non-breacher in a bargained-for manner.⁴²

Is there, then, no fairness content to the analysis of alternative remedies? Modern contract theory seems primarily concerned with providing a "standard" compensatory rule which supplies an incentive for efficient breaches to be made and resulting gains to be fully exploited. It does not draw upon any obvious principle of fairness in evaluating the post-breach end result. As noted above, the "just compensation" formula gives all of the gains to the breacher. Why should this end result be regarded as any "fairer" than one which splits the gains fifty-fifty or gives them all to the non-breacher? In terms of Figure 1B, this question requires the observer to evaluate alternative efficient solutions such as R_b , R_s , and R_g in fairness terms. Allocatively, for society, these solutions are indistinguishable; they differ only in terms of wealth transfers, the manner in which the gains from non-performance are distributed between the parties. It seems then that the only appropriate fairness inquiry is one concerned solely with process fairness, the bargaining conditions, and not an examination of end results. However, if "penalties" are to be attacked on this basis, the affirmative case must be made that penalties are in some way symptomatic of those market conditions which characterize process unfairness.

C. The Objective Compensation Limitation: Problems of Idiosyncratic Value and Uncertainty

The efficient damages model demonstrates that in an ideal environment the existing damage rule has no allocative or fairness advantages. The just compensation standard can nonetheless be supported as the preferable rule if one views it as the simplest of the alternative means of allocating the gains from breach and setting the conditions for further negotiations. A separate problem, however, is raised by the penalty rule which prevents the parties from altering this allocation by agreement. Evaluating the effects of this rule requires specifying the conditions where the costs of enforcing the just compensation principle might induce parties to negotiate an alternative allocation.

42. In the absence of a penal bond, the settlement would be at R_b and the seller would retain all of his utility gains from breaching at R_c and moving to a higher indifference plateau I_{s_3} . The only effect of the penal provision, absent transaction costs, is to induce a transfer of some of those gains by the seller to the buyer in exchange for a release from the penal

1. *The Costs of Implementing a Compensation Standard.* Under the just compensation rule, the breacher must provide a compensation/performance mix which will place the non-breaching party on the quasi-performance curve which equals his contractual expectations. Where the values represented by the points forming the quasi-performance curve are reflected in an existing market, the legal sanctions necessary to implement the compensation principle are easily formulated.⁴³ The breacher provides a monetary compensation either reflecting the damage based on contract-market differential⁴⁴ or profits from the anticipated exchange.⁴⁵ In either case, the recovery is that sum necessary to provide the equivalency of performance based on an objective market valuation. Alternatively, where the subjective values represented by the non-breacher's indifference curve are not reflected in any established market, the compensation principle

constraint. See text accompanying notes 82-86 *infra* where the *ex ante* estimation of renegotiation costs is explicitly addressed.

43. A basic assumption of contract damages is that the "value" of a performance . . . means the amount of money that can be obtained in exchange for it in some market." 5 A. CORBIN, *supra* note 3, at 137. See text accompanying notes 51-54 *infra*. Even in cases where value is assumed to be so limited, implementation of the compensation principle requires difficult judgments concerning the appropriate "market" from which measurement can be taken. As Professor Corbin has noted:

The term 'value' and the term 'market price' are often used in law books, in judicial opinions, and in statutes as if they represented a definite and easily determinable amount of money. It is very clear that such is not the case, and that, in a determination of values and market prices, a great deal of pure guess work is frequently indulged.

5 A. CORBIN, *supra* note 3, at 137.

44. Where the parties are able to act on the market, for example, by reselling goods upon buyer's breach or purchasing substitutes where seller defaults—the compensation principle is implemented by granting the non-breacher the contract-market differential. This is appropriate because it represents the recovery of the contract risk of a shift in the market price which was allocated to the breaching party. The assumption is that existing market opportunities together with the market-contract differential will be equivalent to the performance expectations of the non-breaching party. This measure, which has traditionally been identified as "lost bargain damages," is implemented in U.C.C. §§ 2-708(1) and 2-713 for the non-breaching seller and buyer respectively.

45. There are circumstances where even though market opportunities are present the compensation principle cannot be implemented by the contract-market differential. This is explicitly recognized with respect to seller's damages in U.C.C. § 2-708(2). Identifying the compensation principle as the relevant criterion, the drafters recognized that circumstances may arise where recovery of market fluctuations would not prove an accurate measure. Where that occurs, the seller is entitled to recover his lost profit. Although recovery of anticipated profit is in reality what the seller receives under the market-contract measure in U.C.C. § 2-708(1), market opportunities plus the market fluctuation is ordinarily a more precise method of measurement than the cost figures necessary to identify anticipated profits. Consequently, only where that method is inconsistent with the compensation principle will the seller be entitled to use the direct profit measurement in § 2-708(2). There are at least two circumstances where this latter measurement becomes appropriate: (1) The "lost volume" case—traditionally applied where the seller is operating at less than optimal capacity so that any action in terms of existing market opportunities would have been available without buyer's breach. *But cf.* Comment, *A Theoretical Postscript: Microeconomics and the Lost-Volume Seller*, 24 CASE W. RES. L. REV. 712 (1973) (economic analysis suggests that few sellers will ever "lose" a sale within the meaning of 2-708(2)); and (2) The "component parts" case—where seller has not yet manufactured the goods so that market opportunities cannot be used. Here, market fluctuations will have no rational relation to seller's contemplated investment, and the seller should be required to establish his anticipated profits under U.C.C. § 2-708(2). See Peters, *Remedies for Breach of Contracts Relating to the Sale of Goods Under the U.C.C.: A Roadmap for Article Two*, 73 YALE L.J. 199, 273 (1963); Harris, *A Radical Restatement of the Law of Seller's Damages: Sales Act and Commercial Code Results Compared*, 18 STAN. L. REV. 66, 70-72, 86-89 (1965).

may be held to be inapplicable in practice; specific remedies mandating full performance are commonly required where no monetary equivalency is ascertainable.⁴⁶

Implementing the compensation principle is much more difficult, however, in those cases where there is a recognized market value for the promised performance, but the promisee attaches an additional idiosyncratic value to the performance. Idiosyncratic value would include any subjective or "fanciful" valuation which varied significantly from the established market value.⁴⁷ Where the subjective utility is greater than the established market evaluation, the compensation principle may require a subjective basis for the compensation payment.

In Figure 2, I_2 is the promisee's purely subjective indifference curve which runs through the contract point R_c . In the absence of any external market, this indifference curve would represent the relevant quasi-performance curve. However, if the promisee can buy or sell goods with parties other than the promisor, the presence of such objective market opportunities will affect the definition of the set of equally preferred quasi-performance solutions. For instance, suppose that the market terms of exchange between goods and money are one-to-one, *i.e.*, goods have a price of \$1. This opportunity to *transform* the promised result R_c is represented by the market transformation curve YX in Figure 2. Although R_c is on the same subjective indifference curve as full performance R_e , R_e may not be regarded by the promisee as equivalent to performance at R_c . The reason is that, if R_c had been achieved according to contract, the promisee could possibly resell goods equal to the distance between X_c and X_a , thus arriving at R_a on a higher indifference curve I_1 . The implication for ordinary remedy measures is that, when the promisor performs less than fully by delivering a quantity such as X_a , the recognition of a foregone market opportunity to resell requires compensation equal to the vertical distance

46. Specific remedies where market opportunities are not available have long been intuitively used by the courts to assure implementation of the compensation principle. *See, e.g.*, U.C.C. § 2-716(1), Comment 2 (buyer's right to specific performance); U.C.C. § 2-716(3) (buyer's right to replevin where substitute goods are not available); U.C.C. § 2-709(1)(b) (seller's right to recover the full purchase price where he is unable to resell the goods after a reasonable effort). For a sophisticated discussion of the economic justification of specific remedies in non-market transactions, see Barton, *supra* note 16, at 293-300. *See also* Birmingham, *supra* note 34, at 69-70 for an analysis advocating the use of specific remedies as a means of providing incentives for a renegotiation in which the efficiency gains derived from breach would be shared between the parties.

In certain categories of contract, such as ones for personal performance in music or athletics, the courts have traditionally been reluctant to impose specific remedies because of the supervisory burdens and attendant costs imposed by such a decree. In this circumstance the deprivation, which has no market, is measured only through the unsatisfactory method of leaving monetary value to the intuitive discretion of the factfinder. 5 A. CORBIN, *supra* note 3, § 1003, at 38.

47. This "sentimental" or "fanciful" value has traditionally been described as "*pretium affectionis*." In *Thomason v. Hackney & Moale Co.*, 159 N.C. 299, 305, 74 S.E. 1022, 1024 (1912), a tort case, the court defined it as "an imaginary value placed upon a thing by the fancy of its owner, growing out of his or her attachment for the specific article, its associations and so forth, which, perhaps, may not inaptly be called its sentimental value."

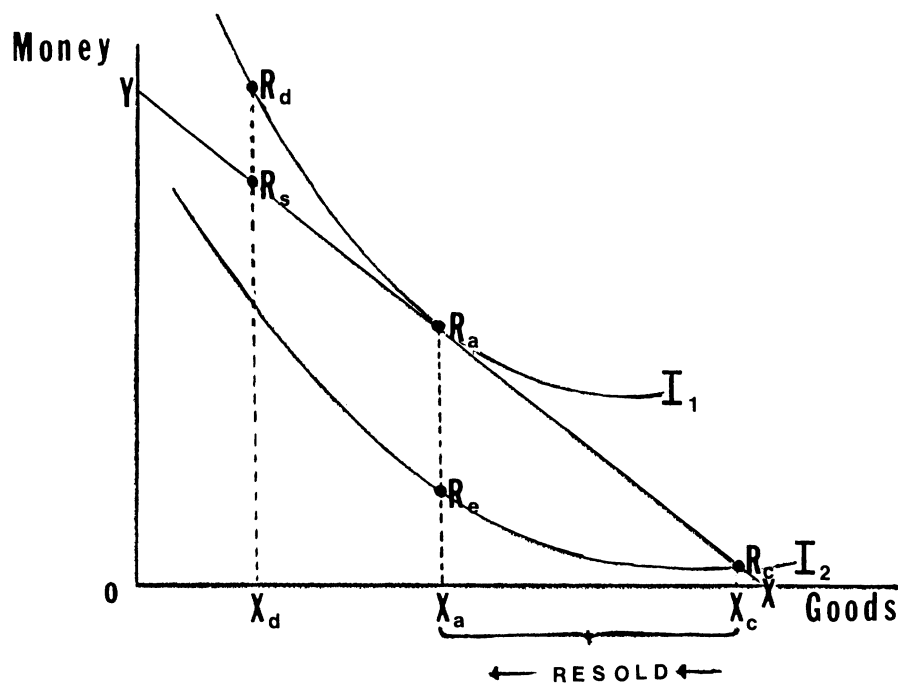


FIGURE 2

R_a R_e in addition to the merely subjective compensation implicit in reaching result R_e .

Thus far, there is no inconsistency with orthodox principles of contract breach compensation because the market measure of loss is exactly equivalent to the true losses. True losses are the *higher* of either (1) objective valuation of lost market opportunities or (2) subjective valuation of lost consumption opportunities. As long as the former exceeds the latter—*i.e.*, the slope of the market transformation curve exceeds that of the indifference curve—objective market measures of damage are perfectly appropriate. Note, however, that when fewer units than X_a are held, the slope of the indifference curve exceeds the slope of the market transformation curve, indicating that the promisee's idiosyncratic valuation of goods becomes greater than the market's valuation.⁴⁸

If, then, the contract point were R_a and the promisor delivered only X_d units, should the non-breaching party be placed, via compensation, at the objectively valued point R_s or at the subjectively valued point R_d ? The answer is that it depends on whether the non-breaching party has a real

48. In order to reduce his holdings of goods from X_a to X_d , the promisee would require more money from a buyer than would the market. The market would require only enough money to move to R_s whereas the promisee would prefer to retain the goods unless he was given enough money to move along curve I_1 to R_d . Thus, the promisee places greater than market value on the goods.

ability to "cover" by purchasing the undelivered goods at the market rate. If this is a genuine possibility, R_s is indeed an equivalent to full performance at R_a , because it can be transformed into R_a by sliding along the market transformation curve, *i.e.*, buying the equivalent goods on the market. At times, however, such ideal ability to cover is not a genuine possibility. For instance, the breach may be unforeseen, so that adequate cover cannot be effectuated in the market during the relevant time period. In this case, subjectively based compensation equivalent to the vertical distance $R_d R_s$ is necessitated in addition to the market's objective damages in order to fully compensate the non-breacher. Existing standard damage measures, however, will frequently not permit recovery at any point other than R_s , the theoretical measure of lost market values.

Why have *ex post* legal sanctions been adopted in which subjective damage elements may not be recognized? Optimal systematic rules establishing post-breach compensation may require some limiting assumptions concerning the extent of harm caused by contract breach. These assumptions reduce the direct or administrative costs of implementing the compensation principle through the legal mechanism, but purchase these advantages at the price of breaking the equivalency between the damage rule and the actual subjective harm which is difficult to assess. In assessing damages, two limiting assumptions—valuation and foreseeability—may operate to prevent the recovery of idiosyncratic value. The rule requiring the loss to be foreseeable, designed to control for causation and remoteness,⁴⁹ can be overcome in some circumstances by pre-contract disclosure of the causal relation between the loss and the breach and of the parameters of reasonable foresight.⁵⁰ The requirement of valuation, given the existence of a market, is a more rigid barrier to the recovery of subjective losses. First, the "value" of a promised performance is generally limited to "the amount of money that can be obtained in exchange for it in some market."⁵¹ Second, where the exchange value is conceded to be inadequate, and "value to the owner" is substituted,⁵² any "fanciful or sentimental" value

49. 5 A. CORBIN, *supra* note 3, §§ 1006-1007, 1020-28.

50. The foreseeability assumption that was first articulated in *Hadley v. Baxendale*, 9 Exch. Ch. 341 (1854). It limits the damage recovery of a non-breaching party to those damages that the parties could reasonably have contemplated to result from a breach. With pre-contract disclosure of the causal relation between the loss and the breach, the breaching party knows which losses he will be liable for and has an opportunity to insure against anticipated loss which the non-breacher might incur as a result of the breach. This is particularly important when the loss is due to some peculiar circumstances of the non-breaching party. *Patterson v. Illinois Cent. Ry. Co.*, 123 Ky. 783, 786, 97 S.W. 426, 427 (1906). See also RESTATEMENT OF CONTRACTS § 330 (1932); U.C.C. § 2-715(2)(a), Comment 3; 5 A. CORBIN, *supra* note 3, § 1008.

51. 5 A. CORBIN, *supra* note 3, § 1022, at 137. "In the process of determining values, market prices will always be used if such prices are available." *Id.* at 50. See also *National Sav. & Trust Co. v. Kahn*, 300 F.2d 910, 913 (D.C. Cir. 1962); *Burke Hollow Coal Co. v. Lawson*, 151 Ky. 305, 306, 151 S.W. 657 (1912); *White v. Schrafft*, 94 N.H. 467, 471, 56 A.2d 62, 65 (1948).

52. Where market value is not a feasible measure of damages, a more elastic "value to the owner" test will be used. The test is frequently used in the case of loss or destruction to

will be excluded on the grounds that such losses are too speculative and uncertain.⁵³ Together, these valuation and certainty limitations may well preclude the non-breacher's recovery of idiosyncratic value.⁵⁴

second hand personal property, where any market value would be plainly inadequate compensation. Value to the owner is generally reached by either replacement cost or original cost less depreciation. *See* *Feldman v. Capital Piece Dye Works, Inc.*, 185 F. Supp. 426, 435 (S.D.N.Y. 1960), *rev'd on other grounds*, 293 F.2d 889 (2d Cir. 1961) (replacement value); *Duka v. Hotel Assocs., Inc.*, 23 Conn. Supp. 500, 502, 185 A.2d 86, 87 (Cir. Ct. 1962) (when article has no market value, its cost may be only evidence of value available); *Cutler-Hammer, Inc. v. Troy*, 283 App. Div. 123, 126 N.Y.S.2d 452 (1953) (replacement value is proper measure of damages for conversion of dies and molds for plastic handles and knives); *McCurdy v. Union Pac. R.R. Co.*, 68 Wash. 2d 457, 467, 413 P.2d 617, 623 (1966). *Cf.* RESTATEMENT OF TORTS § 911 (1939) ("... as used in this Chapter 'value' means exchange value or the value to the owner where this is greater than the exchange value.")

53. In applying the rule that where market value is inadequate, the measure of damage for loss or damage to property is the value to the owner, the overwhelming majority of courts have held that this value does not include any merely sentimental or fanciful value attached to the property by the owner. *St. Louis-S.F.R. Co. v. Kittrell*, 208 Okla. 147, 149, 253 P.2d 1076, 1078 (1953); *Barber v. Moter Inv. Co.*, 136 Or. 361, 366, 298 P. 216, 218 (1931); *De Spirito v. Bristol County Water Co.*, 102 R.I. 50, 54, 227 A.2d 782, 784-85 (1967). *See also* *Nelson v. Leo's Auto Sales, Inc.*, 158 Me. 368, 374, 185 A.2d 121, 124 (1962) (error to permit plaintiff to testify as to the "value to me at this time" of an automobile).

The nature of the dilemma faced by the courts in these cases is illustrated best in *Furlan v. Rayan Photo Works, Inc.*, 171 Misc. 839, 12 N.Y.S.2d 921 (N.Y. Mun. Ct. 1939), in which the plaintiff sought substantial recovery for sentimental value and mental anguish for destruction of his deceased mother's photograph. The court, in limiting recovery to nominal damages, held that while one could easily empathize with the plaintiff's feelings, this reaction was sentiment, not the law, which was bound by substantive and procedural limitations. *See also* *Missouri, K. & T.R. Co. v. Dement*, 115 S.W. 635 (Tex. Civ. App. 1909) (damages for loss of family portraits is actual loss in money sustained by the owner).

A few courts have granted recovery of sentimental value for loss or conversion of heirlooms where special value can be presumed. *See, e.g.*, *Brown v. Frontier Treaties, Inc.*, 369 S.W.2d 299, 304 (Tex. 1963) (rule denying recovery for sentimental value not applicable where the items have their primary value in sentiment). *See also* 1 T. SEDGWICK, DAMAGES § 251 (9th ed. 1912).

In contracts damage cases, this issue is most frequently raised as an aspect of the requirement that the damages for non-performance be proved with reasonable certainty. The certainty aspect of the valuation assumption is a standard which requires a minimum level of proof in the fact and amount of damage. *See, e.g.*, *Central Coal & Coke Co. v. Hartrom*, 111 F. 96, 98 (8th Cir. 1909) ("The actual damages which will sustain a judgment must be established, not by conjecture . . . but by facts from which their existence is logically and legally inferable."); *Griffin v. Colver*, 16 N.Y. 489 (1858). Loss of profits remains the largest category for the application of the standards of certainty. *See, e.g.*, *Fontainebleau Hotel Corp. v. Crossman*, 323 F.2d 937 (5th Cir. 1963); *U.S. v. Griffith, Gornall & Carman*, 210 F.2d 11 (10th Cir. 1954); *Chicago Coliseum Club v. Dempsey*, 265 Ill. App. 542 (1932); *Broadway Photoplay Co. v. World Film Corp.*, 225 N.Y. 104, 121 N.E. 756 (1919); *Keystone Diesel Engine Co. v. Irwin*, 411 Pa. 222, 191 A.2d 376 (1963).

54. Sentimental value is something that cannot be considered in the law of contracts. The amount of compensation for a loss of this kind, however, may not be determined by the market value of the article in question. . . . Sometimes the cost of replacement may be allowed, that is, the market value of some similar article even though it will not service some of the dearest of the purposes for which the original article was treasured; but in the law of contracts no price will be put upon mere feelings of pleasure or affection or feelings of sorrow and distress.

5 A. CORBIN, *supra* note 3, § 1004, at 49. *See* *Sinclair Ref. Co. v. Jenkins*, 289 U.S. 689 (1933); *McGregor v. Watts*, 254 App. Div. 904, 5 N.Y.S.2d 525 (1938).

A classic example of the denial of idiosyncratic value as compensation for breach is provided in *Carpel v. Saget Studios, Inc.*, 326 F. Supp. 1331 (E.D.Pa. 1971). In *Carpel*, the court granted defendant's motion for summary judgment in a diversity action for breach of a contract to take wedding photographs on the ground that plaintiff's allegations of sentimental value were, as a matter of law, insufficient to achieve the monetary jurisdictional limits of the federal court. The court held that

the alleged lost sentimental value of the pictures is so highly speculative that it is not a proper element of damages for consideration by the jury. There are no guidelines available to aid the jury in determining a dollar value for this loss. . . . Since in this case the photographs never came into existence, there is the additional question of whether they would have had any sentimental value at all.

A compensation limitation is justified where the reduction in the litigation costs of proving idiosyncratic harm exceeds the error costs of inaccurate damage measures.⁵⁵ The assumption signals the promisee who places a higher subjective value on performance to provide for that idiosyncratic valuation at the time of contracting unless satisfactory post-breach cover is really achievable at market terms. One method of covering in advance for this idiosyncratic value would be for the promisee to secure the performance to the extent necessary by proposing a liquidated damages clause to the prospective promisor. In addition, the compensation limitation affects the entire continuum of cases where the losses upon breach are uncertain or difficult to establish. Parties to contracts within this range will be particularly induced to negotiate liquidated damage agreements because the uncertainty in damage recovery increases the probability of error in enforcing the compensation standard. Clearly, these agreements will not necessarily approximate provable damages. There are at least two circumstances where the stipulated amount may be significantly disproportionate to anticipated provable loss:

- (1) The expected provable damages are less than the subjective value of performance and the agreement reflects performance security;
- (2) The parties allocate a number of different risks of uncertain damages by specifying a single damage provision for all contingencies.

Both of these agreements fail under the overcompensation assumption of the penalty rule. The first will be characterized as an impermissible effort to induce performance by penalty even where the facts strongly support an inference of subjective non-provable value.⁵⁶ In the second

Id. at 1333.

This view is supported by the RESTATEMENT OF CONTRACTS § 341 (1932). *See, e.g.*, Illustration 2:

A contracts with B, a cremation society, for the cremation of the body of A's child, B promising to preserve the ashes and deliver them to A on request. The body is cremated by B, but the ashes are lost, because the label on the urn in which they were preserved falls off and identification is impossible. B's breach of contract was not willful. A cannot get judgment for damages for his mental suffering.

55. It is plausible to hypothesize that the legal damage rule promotes an efficiency-enhancing reduction in the total costs of enforcement. The compensation principle is an open-ended standard which operates at a very broad level of generality. By eliminating the consideration of idiosyncratic value in defining the legal damage rule, an increase in ascertainability is achieved by reducing the number and complexity of the relevant facts. The benefits of this more certain legal damage rule will be reduced litigation costs and increased foreseeability. The primary cost of this ease in measurement is the divergence between the scope of the rule and the behavior being regulated (compensation for contract breach). In this case the rule tends to permit conduct (inadequately compensated breach) that would be prohibited by an unqualified application of the compensation principle. Provided that the transactions costs of optional, more inclusive alternatives to the standard rule are relatively modest (*e.g.*, by private negotiation of liquidated damages), then a crude, underinclusive standard rule produces cost savings in most cases while not unduly disadvantaging parties involved in situations where a specially negotiated recovery rule is indeed cost-effective. *See generally* Ehrlich & Posner, *An Economic Analysis of Legal Rulemaking*, 3 J. LEG. STUD. 257, 262-68, 285 (1974).

56. One of the clearest illustrations of an agreement where subjective value induced a

case, where a contract specifies a single sum payable for the breach of stipulations of varying importance, the courts have universally refused enforcement.⁵⁷

If these conditions represented the universe of agreements affected by the penalty rule, it might be argued that the costly effects were exceeded by the savings on the more numerous cases where process unfairness infected the agreement and it was unnecessary to actually prove the unfairness of the bargaining, a costly procedure. However, a more significant aggregate cost is produced over the entire range of cases where true losses from breach are uncertain to some degree. Here the efficiency incentives identified earlier induce negotiated damage agreements. But the current penalty rule imposes additional transaction costs on all of these cases, by inducing the party who regrets the initial allocation to litigate on whether the *ex ante* agreement sufficiently mirrors the anticipated losses.⁵⁸ The actuality—or

liquidated damages clause is *Muldoon v. Lynch*, 66 Cal. 536, 6 P. 417 (1885). The plaintiff agreed to erect a marble monument over the grave of the defendant's husband for \$18,788, the contract providing for liquidated damages of \$10 per day for delay in completion. In declaring the agreement void as a penalty, the court concluded:

There is nothing in this case to indicate that the defendant has suffered any *actual damage which can be measured or compensated by money*. . . it has been generally held that the party in whose favor the penalty or forfeiture exists must prove his damage. In the case before us there is no claim of special damage; it might have been quite difficult for the defendant to show any damage of a pecuniary nature for the non-completion of the monument at the time specified, *though its completion might have been of great comfort and consolation to her affectionate remembrance*.

Id. at 539-40 (emphasis added).

In *City of Rye v. Public Serv. Mut. Ins. Co.*, 34 N.Y.2d 470, 315 N.E.2d 458, 358 N.Y.S.2d 391 (1974) an agreement for \$100 per day and \$100,000 maximum for delay in completion of a building complex was held void as a penalty. The court found that the harm which the city contended it would suffer by the delay was speculative, or simply non-cognizable: "The most serious disappointments in expectation suffered by the city *are not pecuniary in nature and therefore not measurable in monetary damages*." *Id.* at 473 (emphasis added). See also *Security Safety Corp. v. Kuznicki*, 350 Mass. 157, 213 N.E.2d 866 (1966); *Gorco Constr. Co. v. Stein*, 256 Minn. 476, 99 N.W.2d 69 (1959); *Norman v. Durham*, 380 S.W.2d 296 (Mo. 1964).

At the beginning of the nineteenth century in *Astley v. Weldon*, 126 Eng. Rep. 1318 (C.P. 1801), Lord Eldon unsuccessfully challenged the developing penalty rule. In arguing for strict adherence to a fair exchange presumption he stated:

But nothing can be more obvious than that a person may set an extraordinary value upon a particular piece of land, or wood on account of the amusement which it may afford him. In this country a man has the right to secure to himself a property in his amusements: and if he choose to [do so], I see nothing irrational in such a contract; and it appears to me extremely difficult to apply with propriety the word 'excessive' to the terms in which parties choose to contract with each other.

Id. at 1321.

57. See generally *Wallis v. Smith*, 21 Ch. D. 243, 250 (1882); *Kemble v. Farren*, 130 Eng. Rep. 1234 (C.P. 1829); *Astley v. Weldon*, 126 Eng. Rep. 1318, 1322 (C.P. 1801) (Heath, J.). RESTATEMENT OF CONTRACTS, § 339 Comment (b) (1932), provides that where a contract "promises the same reparation for the breach of a trivial or comparatively unimportant stipulation as for the breach of the most important one or of the whole contract, it is obvious that the parties have not adhered to the rule of just compensation."

This rule has caused the invalidation of various allocations of risk by parties dealing in arm's length transactions. See *Benfield v. Croson*, 90 Kan. 661, 136 P. 262 (1913) (single payment of \$500 for breach by either party of agreement to trade a stock of merchandise for a tract of land held void as a penalty); *H.J. McGrath Co. v. Wisner*, 189 Md. 260, 55 A.2d 793 (1947) (liquidated sum fixed the same for a total as for a partial breach declared a penalty).

58. See, e.g., *Hutchison v. Tompkins*, 240 So.2d 180 (Fla. Dist. Ct. App. 1970), *rev'd on other grounds*, 259 So.2d 129 (Fla. 1972).

It is to be remembered that in the instant case the parties contracted at arm's length in

even the mere threat—of such post-breach attacks increases the transaction costs of contracting even in those cases where no disproportion between agreed and anticipated damages is ultimately discovered by the court.

2. *The Alternative Contract: Oasis or Mirage?* Under the penalty rule, an agreement that fails the "reasonable estimate" formulation will not be enforced despite the expressed intention of the parties.⁵⁹ Some effort to deal intuitively with the inconsistencies of this rule can be found in the theory of alternative contracts—the agreement does not stipulate damages for breach, but rather an alternative contract in the event that the primary contract cannot be performed.⁶⁰ Although a few courts have used the device of alternative contracts to sustain agreements which appeared invalid under the penalty rule,⁶¹ there is little support in the cases for the assertion that proof of reliance upon the stipulated sum will permit its enforcement as an alternative promise. Traditional analysis has distin-

no uncertain terms and there was no intimation of fraud or over-reaching on the part of the vendors. . . . The vendees are in unexcused, unexplained and flagrant breach of their contract. Thus, it is a strange and unpalatable circumstance to note the Court's preference over the blameless vendors.

Id. at 187 (Walden, J., dissenting).

59. In *Condan v. Kemper*, 47 Kan. 126, 27 P. 829 (1891), the court stated [T]he words of the parties with respect to damages, losses, penalties, forfeitures, or any sum to be paid . . . must be given due consideration; and, in the absence of anything to the contrary, must be held to have controlling force; but when it may be seen from the entire contract, and the circumstances under which the contract was made, the parties did not have in contemplation actual damages or actual compensation, . . . then the amount stipulated . . . cannot be considered as liquidated damages.

Id. at 130, 27 P. at 830 (1891). See also C. McCORMICK, *supra* note 3, § 149.

The common law doctrine that liquidated damages provisions were unenforceable only if the parties *intended* them as a penalty has generated much doctrinal confusion. W. HALE, *supra* note 23, §§ 48-50. It appears that this confusion, and the inability of the courts to deal consistently with the intent rule, has resulted in its abandonment. As one commentator stated, [w]hether the parties made a real effort to estimate probable damages or intended the stipulation to act *in terrorem* to encourage performance, is often stated to be important in determining if the provision is a penalty. But intent alone is seldom the basis of decision, and in a recent Michigan case is expressly rejected as immaterial. (citing *Wilkinson v. Lanterman*, 314 Mich. 568, 22 N.W.2d 827 (1946)).

It seems unlikely that the expressed intent of the parties ever had much influence, and [it has been argued that] since . . . the only real indication of intent is the relation between the amount stipulated and probable actual damages, intent as such is of no importance.

Developments in the Law—Damages, 61 HARV. L. REV. 113, 129 (1947). See also Dunbar, *Drafting the Liquidated Damage Clause—When and How*, 20 OHIO ST. L. J. 221, 225 (1959); Brightman, *supra* note 3, at 279-81; RESTATEMENT OF CONTRACTS § 339, Comment on Subsection (1), Comment (a) (1932); 5 A. CORBIN, *supra* note 3, § 1058; *Jaquith v. Hudson*, 5 Mich. 123, 138 (1858); *Goeco Constr. Co. v. Stein*, 256 Minn. 476, 481-82, 99 N.W.2d 69, 74 (1959).

60. Macneil, *supra* note 14, at 500.

61. In *Blank v. Borden*, 11 Cal. 3d 963, 524 P.2d 127, 115 Cal. Rptr. 31 (1974), the California Supreme Court avoided the rigid statutory restrictions on liquidated damages by finding a provision to pay a real estate broker the full 6% commission if the property were withdrawn from sale to be a valid alternative contract. The Court's effort to find an analytical basis for its decision was largely unpersuasive. As noted in dissent:

The majority never reached the issue whether the . . . clause . . . was an invalid penalty clause. . . . Instead, the majority neatly sidestep [*sic*] this issue by labeling the brokerage contract as one contemplating an 'alternative performance' by the owner. . . . [T]he issue in this case cannot be avoided by the facile use of labels—otherwise any illegal penalty could be disguised as a "true option" by the promisor to pay a substantial sum for the privilege of breaking his contract.

11 Cal. 3d at 975, 524 P.2d at 133, 115 Cal. Rptr. at 37 (Burke, J., dissenting).

guished the alternative provision designed to secure performance of the primary promise from two promised alternatives between which the promisor can choose, each an agreed exchange for the consideration given by the promisee.⁶² The former would be an invalid penal sanction, while the latter an enforceable alternative contract. Since the relative *market* value of the alternatives will frequently be decisive in making the distinction, the promisee who attempts to secure idiosyncratic value or allocate uncertain risks may still be found to have exacted an invalid penal sanction.⁶³

The Restatement of Contracts, in classic understatement, recognizes that enforceable alternative contracts may easily be confused with invalid liquidated damage provisions.⁶⁴ But a careful analysis of the legal status of alternative contracts reveals a distinction which considerably diminishes prospects of using this principle to uphold disproportionate agreements. If the provision is determined to be a valid alternative contract, damages for breach without an election of alternatives will be based on the *least valuable* alternative.⁶⁵ Consequently, where one alternative is the payment of a specific sum of money, that sum will not determine the amount of damages for breach where the market value of the alternative performance is less than the stipulated sum. The effect of this provision is to preclude recovery of a stipulated sum where it includes unprovable idiosyncratic value, or represents an agreed allocation of risks based on uncertain damage recoveries.

There is in the current legal position on liquidated damages a certain admirable tenacity. Having explicitly refused to permit the identification of idiosyncratic value in post-breach damage measures, the legal mechanism has resolutely precluded parties from protecting that value, either directly or indirectly, in the contracting process. Consequently, the presumption of overcompensation continues to be employed by many courts in avoiding bargained-for damage agreements. Those courts unwilling to presume overcompensation from the fact of disproportion must undertake their analysis within the confines of the "reasonable estimate" formulation which does not allow for idiosyncratic value.

62. *Garrett v. Coast & S. Fed. Sav. & Loan Ass'n*, 9 Cal. 3d 731, 738, 511 P.2d 1197, 1201, 108 Cal. Rptr. 845, 850 (1973); *Paolilli v. Piscitelli*, 45 R.I. 354, 121 A. 531 (1923). See also RESTATEMENT OF CONTRACTS § 339 Comment f (1932); C. McCORMICK, *supra* note 3, at 617-18. Professor McCormick stated:

[W]hile an alternative promise to pay money when it presents a conceivable choice is valid, yet, if a contract is made by which a party engages himself either to do a certain act or to pay some amount which at the time of the contract no one would have considered an eligible alternative, the alternative promise to pay is unenforceable as a penalty.

Id. (footnotes omitted).

63. C. McCORMICK, *supra* note 3, at 617-18.

64. RESTATEMENT OF CONTRACTS § 334, Comment c (1932).

65. *Id.* § 344 & Illustration 1 at 566-67: "The damages for breach of an alternative contract are determined . . . in case of breach without an election, in accordance with the alternative that will result in the smallest recovery."

II. AN ENFORCEMENT HYPOTHESIS: A MODEL OF THE MOST EFFICIENT INSURER

Applying an efficiency analysis to contract damage rules suggests the following enforcement hypothesis:

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.

This hypothesis is based on the assumption that liquidated damage provisions will (1) reduce transaction costs where the parties determine that the costs of negotiation are less than the expected costs of litigation upon breach; and (2) reduce the error costs produced upon breach when the promisee is denied recovery for his non-provable idiosyncratic value. It follows, unless enforcement produces other inefficiencies, that enforcing agreements negotiated *ex ante* will enhance efficiency by permitting the parties to minimize the costs of transacting. The current penalty rule seems to produce significant inefficient effects by limiting the possibilities of mutually beneficial exchange. In addition, negotiated damage agreements are now subject to post-breach attack as penal sanctions. This increases the direct costs of litigation in all cases—even where the agreement is upheld.

The situational model which will be used to test the hypothesis can be illustrated by the hypothetical *Case of the Anxious Alumnus*. Assume the following facts: Dean Smith, a 1957 graduate of the University of Virginia, is a loyal, some would say fanatical, fan of the University of Virginia Cavalier college basketball team. For the 1976 season, after years of second division performances in the highly competitive Atlantic Coast Conference, the Cavaliers finally produce a team that advances to the finals of the conference championship tournament at the end of the season. Through hard work and financial sacrifice, Smith acquires twenty-five tickets to the conference championship game in Landover, Maryland. Smith enters into contract negotiations with the Reliable Charter Service, Inc. to arrange for a bus to transport himself and twenty-four other Virginia fans to Landover on the day of the game. The standard price for this service is \$500.

Smith considers his attendance at the game to be of supreme importance and does not relish the thought of anxious and sleepless hours worrying whether the bus will arrive and successfully accomplish the desired purpose. He is eager to quiet his fears by securing adequate protection in case Reliable fails to perform. However, under the current legal rule Smith cannot protect his unprovable reliance either by securing fully compensating post-breach damages or a bargained-for stipulation of the value of performance.⁶⁶ Consequently, he is forced to consider other protective alternatives.

66. See note 56 *supra*.

One option is to attempt to insure against the subjective consequences of breach with a third party (Lloyd's of London, for example). Assuming that a policy could be secured and enforced up to the assessed valuation of performance,⁶⁷ adding the proceeds of the policy to the award of provable expectation damage which Smith could recover under existing law would provide him with full recovery for his idiosyncratic value upon breach by Reliable.

Alternatively, Smith could negotiate for direct insurance from Reliable or any of its competitors offering the same service. Dean might propose to pay Reliable \$1,000 for the charter service if, in return for the additional premium, Reliable would agree to a penal sanction of \$10,000 upon failure of performance. The stipulated sum of \$10,000 would represent that amount at which Dean would be indifferent between performance and breach. Unfortunately, insurance purchased directly from the promisor, Reliable, is not a real alternative; the mere labelling of the idiosyncratic damages as pursuant to an "insurance" contract is unlikely to prevent a perceptive court from recognizing that such payments are de facto equivalent to a penalty. Hence, legally enforceable insurance for damages not recoverable as breach damages is in practice obtainable only from third parties.

This insurance model and the enforcement hypothesis pose the following issues for resolution:

- (1) As between the third-party insurance company and the promisor, which is the more efficient provider of insurance?
- (2) To what extent do the rationales supporting the "indemnity principle" suggest significant additional social costs due to enforcement of agreements at stipulated values?
- (3) Assuming changed conditions after an insurance agreement has been negotiated, such that the value of performance to the promisor is reduced, does enforcement of a penal sanction produce a high probability of inefficient effects?
- (4) What presumptions of unfairness can be developed to cope with those special classes of cases where bargaining abnormalities would produce inefficient effects if stipulated damage provisions were enforced?

A. *The Efficient Insurer Model*

Identifying the efficient insurer requires an analysis of the costs of providing insurance. At what price would a profit-making commercial enterprise be willing to offer Smith \$10,000 worth of protection against the contingency of bus failure en route to the fabled final game of the championship?

Perhaps the overwhelming element in the cost of this insurance would

67. Insurance recovery may be limited to actual economic loss. See text accompanying notes 74-77 *infra*.

be the expected value of the underwriting loss to the insurer. This expected value is defined as the product pR where p is the probability of non-performance and R is the recovery payable to the insured. In addition, an insurer will also have other transactions costs, such as the costs of ascertaining the true probability p and the costs of negotiation and communication with the insured. These transaction costs will be subsumed in the portmanteau variable T , so that the total cost C of a policy paying R on the occurrence of an event with probability p can be summarized as

$$C = pR + T.$$

We assume that since the services in question are marketed competitively in the presence of alternative sellers, the cost of breach insurance to Smith will be $(1 + \alpha)C$ where α is the competitive rate of return or profit for the insurer. The question, then, is whether C would differ between the bus company and the third-party insurer.

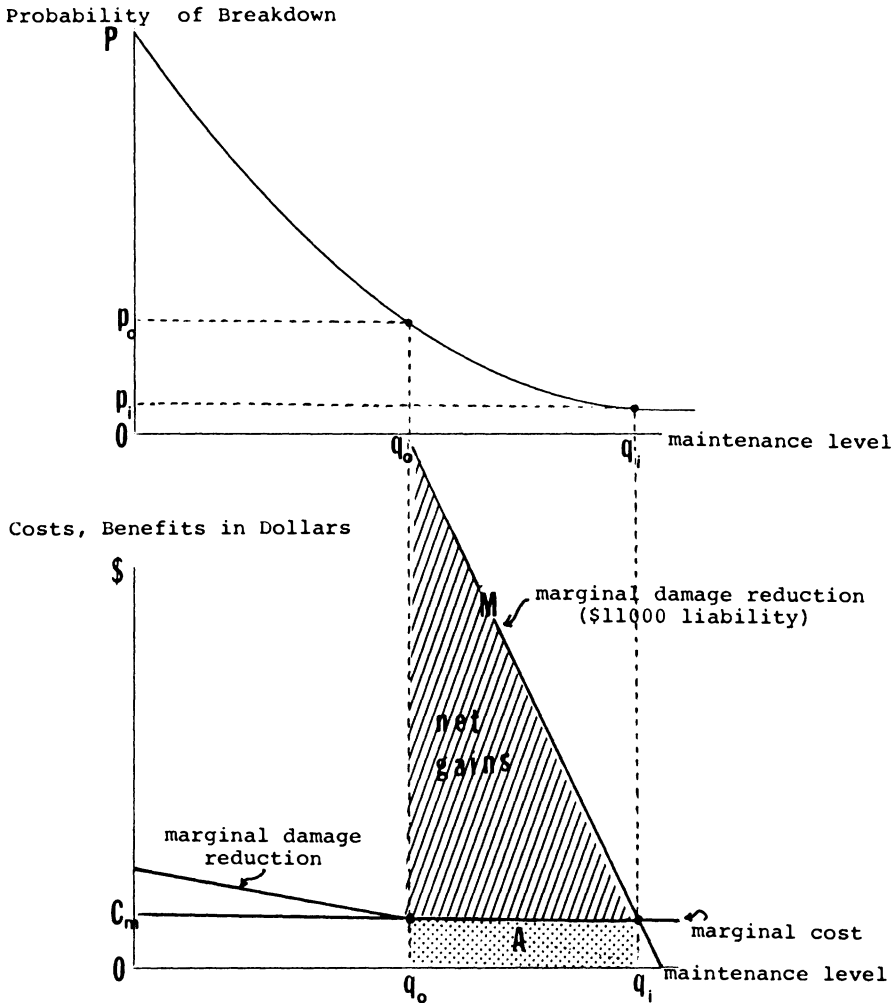
An obvious focal point of interest is the transaction cost element T . Here, it is tempting to argue that the advantage lies with the bus company. In the first place, the bus company is in a superior position to know the breakdown probability p . Secondly, many of the other transactions costs normally incident to customer communication may be negligible when communication is already being undertaken relative to the carriage service itself. Hence, T may be lower for the bus company and thus so would C and the offering price of the insurance to Smith.

Actually, however, the transactions cost element is not the strongest argument in favor of the bus company as the most efficient insurer. The bus company's main advantage derives from its power to exercise some control over the breakdown probability p . This can be illustrated by examining how the rational enterprise will make the maintenance and repair decisions which affect the breakdown probability.

Figure 3 reflects the fact that there is an inverse relationship between the probability of breakdown and the level of maintenance.⁶⁸ Each additional "dose" of maintenance reduces p , presumably by smaller and smaller amounts as the maintenance level increases.

In Figure 4, the technological facts of Figure 3 are converted into their investment implications. If D is the damage liability upon non-performance (here \$10,000), the *expected* damages decline as increasing maintenance decreases the probability of breakdown. Since these expected damages are pD , the saving in expected damage costs attributable to a marginal unit of maintenance is rD where r is the marginal reduction in p brought about by the maintenance increase. Hence, the "marginal damage reduction" curve in Figure 4 is simply a schedule of the value rD for alternative maintenance

68. "Maintenance" is used as a summary word for all bus company activities which potentially affect the breakdown probability.



FIGURES 3 & 4

levels.⁶⁹ The “marginal cost” curve is, in turn, a reflection of the incremental costs incurred as the maintenance level is varied.⁷⁰ The firm will maintain the buses up to the point where the last or marginal unit of maintenance is “just worth it,” *i.e.* where the marginal damage reduction curve intersects the marginal cost curve at quantity q_0 .⁷¹

69. This is an inverse relationship as it is closely related to the curve in Figure 3. As the maintenance increases (from q_0 to q_i), the marginal damage reduction decreases, eventually tending to zero, a point where increased maintenance does not reduce the probability of breakdown at all.

70. In Figure 4, this marginal cost is a constant. Each additional unit of maintenance has the same cost.

71. If the company would continue to maintain the buses, spending another unit of maintenance beyond q_0 , it would cost C_m but reduce the expected damage liability by an amount less than C_m . This would clearly be uneconomical for the bus company.

Absent the \$10,000 liquidated damage agreement, the bus company anticipates that D will embody only the standard objective damage recovery which, let us assume, amounts to \$1,000 for the Smith bus trip. This anticipation underlies the marginal damage reduction curve in Figure 4 and is the ultimate basis for the p_0 breakdown probability. In computing the expected underwriting loss, the third party insurance company will therefore arrive at a value ($p_0 \cdot \$10,000$).

Suppose, however, that the bus company can offer the same insurance. The expected value of damages is now based, not on a D of \$1,000, but on a D of \$1,000 actual provable damages plus \$10,000 insurance recovery. Therefore, the new marginal damage reduction curve (M) for maintenance will be exactly eleven times higher than the original one. The company will expend additional maintenance costs A, equivalent to the shaded area in Figure 4 or $C_m \cdot (q_1 - q_0)$, and the breakdown probability will consequently decline to p_1 .

What are the implications of these adjustments on cost? For the bus company, the insurance cost must now be modified to reflect the net benefits of possible risk-avoidance efforts. Hence, the appropriate cost function for the provision by the bus company of \$10,000 coverage is $C = (p_0 \cdot \$10,000) - [\$11,000 (p_0 - p_1) - A] + T$ where the terms in the square bracket are net gains from adjusting maintenance levels: $(p_0 - p_1)$ is the change in breakdown probability, its product with \$11,000 is the expected damage reduction, and A is the added cost of maintenance.⁷² We know that these net gains are positive from the nature of their computation⁷³ and that they would not be achieved when the third-party insurance is purchased. Hence, even where the transaction cost component T is identical for the alternative insurers, the bus company has an efficiency advantage equal to the square-bracketed term in the equation above. (Graphically, this quantity is represented in Figure 4 by the crosshatched triangular area of net gains above the shaded area of incremental costs A and below the new marginal benefits curve M.)

The preceding argument has been that non-enforcement of liquidated damages provisions has the result of inducing individuals to protect against

Note that this decision of q_0 maintenance also determines the probability p_0 in Figure 3, the probability of breakdown when the bus company stops its maintenance and the bus actually sets out.

72. Note that the relevant cost function for the bus company is the expected loss due to the provision of the insurance. The present expected loss to the bus company is $C_1 = (\$11,000 \cdot p_1) + A + T$. In computing the insurance premium, the pre-insurance expected loss of $C_0 = \$1,000 \cdot p_0$ (provable damages upon breach multiplied by probability of breach) must be deducted. The cost function is thus $C_1 - C_0 = (\$11,000 \cdot p_1) - (\$1,000 \cdot p_0) + A + T$.

The equation in the text is equivalent to this one. It is presented in a more complicated form to demonstrate the relationship between the bus company's insurance and the third party's. The first term is equal to the cost function of the insurance company ($p_0 \cdot \$10,000$). The bracketed term, which is positive, is then subtracted which indicates that the bus company's insurance is less expensive by this amount than the third-party insurance.

73. See text accompanying notes 69-71 *supra*.

otherwise non-recoverable losses through special third-party insurance. This is likely to be an extremely inefficient alternative since there are strong economic arguments that suggest that the vendor is the lowest-cost insurer against non-performance. Although our argument has been framed in terms of the bus company example, a similar conclusion may be generalized to all cases in which the vendor has some control over the probability of externally caused non-performance.

In sum, many people may not want to make deals unless they can shift to others the risk that they will suffer idiosyncratic harm or otherwise uncompensated damages. To the extent that the law altogether prevents such shifts from being made or reduces their number by unnecessarily high costs, it creates efficiency losses; that is, it prevents some welfare-increasing deals from being achieved.

B. *The Indemnity Principle: The Problems of Wagering and Moral Hazard*

The preceding model assumed that legal principles governing the formation and enforcement of aleatory contracts would permit the promisee to secure third-party insurance to cover unprovable idiosyncratic value. We concluded that the third-party option is likely to be relatively inefficient. The availability of even the inefficient option must be examined more carefully, however, in light of the principle of indemnity.⁷⁴ Since a contract insuring property rights is a contract of indemnity only, payment upon loss is generally based not upon the face value of the policy, but upon the provable economic loss.⁷⁵ Does this general limitation on recovery suggest reasons against enforcing insurance of idiosyncratic value? The policies enjoining wagering and moral risk are traditionally identified as the reasons supporting the indemnity rule. Examining these policies, however, reveals

74. See generally R. KEETON, BASIC INSURANCE LAW 69-162 (1959) [hereinafter cited as BASIC INSURANCE]; R. KEETON, INSURANCE LAW—BASIC TEXT 88-173 (1971) [hereinafter cited as INSURANCE LAW]; E. PATTERSON, CASES AND MATERIALS ON THE LAW OF INSURANCE 247-307 (3d ed. 1955).

75. In its broadest definition the indemnity principle merely implies that the value of the insurance benefit should not exceed the loss. At that level of generality the principle is a valid recognition that the prevention of net gain from an insurance contract promotes two perceived policy objectives: (1) the discouragement of wagering contracts and (2) the reduction of moral hazard. The problem lies not with the principle as an articulated "standard," but in the legal "rule" adopted to implement that standard which defines the measure of recovery with reference to the exchange value of the insured's loss. The measure of recovery generally used to determine the "actual value" of the loss is the difference between the market value of the property before and after the loss. See, e.g., *Cassil v. Newark Ins. Co.*, 274 Wis. 25, 79 N.W.2d 101 (1956); *Engl v. Calvert Fire Ins. Co.*, 266 Wis. 419, 63 N.W.2d 831 (1954). In cases where the market value is not a fair measure of indemnity of the insured's loss, an alternative measure—replacement cost less depreciation—is frequently used. See, e.g., *Agoos Leather Cos. v. American & Foreign Ins. Co.*, 342 Mass. 603, 174 N.E.2d 652 (1961); *Pinet v. New Hampshire Fire Ins. Co.*, 100 N.H. 346, 126 A.2d 262 (1956). At least one commentator has recognized that these exchange value measures might produce less than full indemnity for the insured, particularly where the insured claims a "special relationship . . . to the property . . ." The result is justified on the basis of administrative costs, however, since "getting into the area of special value to the insured would substantially increase problems . . ." See INSURANCE LAW, *supra* note 74, at 146-47.

adequate substitutes for controlling the increased social costs of the enforcement hypothesis.

Traditional analysis distinguishes a gambling contract which creates risk, from a valid insurance contract which merely allocates existing risk to a third party for a price.⁷⁶ The indemnity principle prevents parties who shift existing risk by contract from creating additional risk through overvaluation of the insured object. Collection of insurance coverage for an over-valued "loss" yields a net gain to the insured. Such a contract is therefore viewed as a gamble by the insured on the occurrence of such a "loss."

Clearly, the policy against wagering is potentially inconsistent with the enforcement hypothesis. Assuming that the bargaining process is fair, the hypothesis requires enforcing all stipulated agreements without inquiry into the nature of the underlying risks. An additional post-breach inquiry would increase transaction costs by requiring proof of valid risk allocation by the promisor. This issue, however, appears collateral to the central question of enforcement. If it is determined that wagering contracts are violative of public policy, a stipulated damages provision used to disguise a gambling transaction would be unenforceable. Without empirical data indicating that wagering contracts are prevalent in liquidated damages provisions, a presumption of legitimate risk allocation would appear to reflect more accurately the use of stipulated damage provisions in commercial contracts. The danger of wagering seems sufficiently remote to leave invalidation to explicit proof of a gambling transaction, rather than using it as a rationale for invalidating all agreements which allocate risks based on unprovable reliance.

The second rationale supporting the indemnity principle is the perceived moral hazard that accompanies the allocation of idiosyncratic risk to a third-party insurer. The very nature of unprovable value heightens the risk that the assessed valuation will overcompensate the insured, providing incentives for fraudulent claims of loss.⁷⁷ Limiting the recovery to provable loss, irrespective of the stipulated value, will reduce the moral hazard by some amount. Again, the choice that must be resolved is whether the savings produced by reducing the frequency of fraudulent claims of loss exceeds the increase in enforcement costs due to undercompensation and reduced contracting flexibility. Absent empirical evidence, there is no persuasive reason to invalidate an entire range of legitimate bargains in order to produce an indeterminate reduction in moral risk. Faced with uncer-

76. See INSURANCE LAW, *supra* note 74, at 89-90; Patterson, *Insurable Interest in Life*, 18 COLUM. L. REV. 381, 386 (1918). Professor Keeton asserts that the policy against wagering has been a major influence in formulating the indemnity principle. The underlying evils to be discouraged are felt to include vice, idleness, impoverishment, misery, crime, and the discouragement of useful work. BASIC INSURANCE, *supra* note 74, at 89; Patterson, *supra*, at 386.

77. INSURANCE LAW, *supra* note 74, at 88.

tainty, efficiency may well be maximized by a less inclusive rule which denies recovery of loss induced by fraud once evidence of fraudulent conduct is established.

Although the probability of fraud induced by moral risk will not be eliminated in the direct insurer model, it should be a smaller risk than in the case of third-party insurance against idiosyncratic loss. Assume that Lloyd's of London sells a policy to the promisee with an excessive recovery amount. There comes a point under this contract where collusive breach between the promisor and promisee becomes profitable. Our hypothetical alumnus, for example, would bribe the bus company to breach and make Lloyd's pay most of the cost. The risk of this collusion against the insurer is clearly nonexistent when the provider of the product is the insurer.

The rationales supporting the indemnity principle indicate that limiting recovery for nonprovable loss may produce a decrease in the probability of fraud and the enforcement of socially detrimental gambling contracts. Unlike the absolute sanction of the penalty rule, however, the indemnity principle in insurance contracts is not applied when competing values with greater benefits are identifiable.⁷⁸ The valued policy doctrine permits parties to an insurance contract to avoid the limitations of the indemnity principle by stipulating a valuation which will be payable upon proof of loss.⁷⁹ Stipulated value policies originated in marine insurance because valuation based on indemnity was too imprecise to insure full recovery. They are commonly used to insure items of tangible personalty where market value is not a reliable indicator of subjective worth.⁸⁰

This valued policy alternative indicates that the indemnity principle can be efficiently applied.⁸¹ Where the expected costs of undervaluation

78. "In some instances, however, the evils are more than balanced by the costs of effective safeguards—the two most significant costs being a reduction in flexibility of insurance arrangements and an increase in expense of administration. The result is that rules of insurance law fall short of compelling strict adherence to the principle of indemnity." *See BASIC INSURANCE*, *supra* note 74, at 69.

79. *See generally* *INSURANCE LAW*, *supra* note 74, at 140-42; W. YOUNG, *INSURANCE, CASES & MATERIALS* 558-59 (1971). In selecting a valued policy, the parties "mean to substitute their present assessment for the result of later controversy. *In the absence of fraud this is as conclusive as in the case of any other damages; and indeed valued insurance is only an instance of stipulated damages.*" *St. Paul Fire & Marine Ins. Co. v. Line Oil Co.*, 63 F.2d 771, 772 (2d Cir. 1933) (emphasis added).

80. Professor Young identifies jewelry, works of art, vessels, and cargo as those subjects that are commonly insured with a valued policy. W. YOUNG, *supra* note 79, at 558.

81. In a related development, many jurisdictions have enacted valued policy statutes which require the insurer to satisfy claims for specific types of real property insurance at full policy valuation. This legislation is based primarily on the notion that the insurer should be required to make the risk and valuation determinations at the time of contracting. Assuming competitive bargaining, the insurer should not be permitted to exact a premium for a stipulated amount of insurance and, upon loss, assert the indemnity principle to limit his liability to provable exchange valuation. "Valued policy" legislation commonly applies to fire insurance on buildings and other fixtures. In some jurisdictions the coverage is extended to other types of property insurance. *See, e.g.,* KAN. STAT. ANN. § 40-905 (1964) (coverage for loss caused by fire, tornado and lightning). Some statutes have been extended to personal property in spite

exceed the benefits in reducing the probability of fraud or wagering, the legal system will enforce insurance contracts covering nonprovable idiosyncratic value. We conclude, therefore, that existing legal restraints on fraud and gambling are adequate substitutes for controlling the increased social cost resulting from the enforcement hypothesis.

C. *Changed Conditions: Inefficient Performance and Induced Breach*

The enforcement hypothesis proves a persuasive basis for a new legal doctrine as long as the assumption of fairly-bargained exchange justifies confidence in end-result efficiency. A challenge to that assumption is posed, however, by the problem of changed economic conditions. The nature of the problem can be illustrated by examining the case of *Rockingham County v. Luten Bridge Co.*⁸² In *Rockingham* the county contracted with the bridge company to construct a bridge to provide a connecting link with a hard surface road to be constructed simultaneously. Subsequently, the construction of the road was terminated during a political dispute, and the county moved to repudiate the bridge contract. At the time of repudiation, the company had incurred \$2,000 in reliance costs. The company continued work on the bridge until six months after repudiation and then instituted action against the county for \$18,500 representing work performed under the contract.

Denying recovery for the work performed after repudiation, the court applied classic mitigation principles in holding that, "it is inflicting damage on the defendant without benefit to the plaintiff to allow the latter to insist in proceeding with the contract."⁸³ The court in *Rockingham* correctly identified and applied the just compensation principle to the changed conditions problem. Any further recovery would have produced a disincentive to efficient breach, and prevented a preferred reallocation of resources by the promisee.

Assume, however, that in the initial contract the county attached a substantial, nonprovable value to full performance, and secured a promise by the company that upon nonperformance it would pay a stipulated damage of \$20,000. The cancellation of the road building project reduced the value of the performance to the county to \$1000. Assume further that a

of the increased moral hazard. *See, e.g.,* *Duckworth v. United States Fidelity & Guar. Co.*, 452 S.W.2d 280 (Mo. App. 1970).

Although subject to extensive criticism by commentators, *see, e.g.,* BASIC INSURANCE, *supra* note 74, at 142; W. YOUNG, *supra* note 79, at 559, valued policy statutes intuitively reflect the jurisprudential paradigm of the mutually beneficial market exchange.

82. 35 F.2d 301 (4th Cir. 1929).

83. *Id.* at 307. The court correctly grounded its decision on the principle of compensation:

The legal right of either party to violate, abandon, or renounce his contract, on the usual terms of compensation to the other for the damages which the law recognizes and allows, subject to the jurisdiction of equity to decree specific performance in proper cases, is universally recognized and acted upon.

Id.

change in raw material prices will produce an estimated \$15,000 loss to the bridge company if full performance is required. Under the current legal rule, the *in terrorem* \$20,000 damage clause would be unenforceable and the parties would be left to conventional damage measures. This would induce a breach by the bridge company, and the payment of \$1000 expectation damages—an efficiency-maximizing result. However, under the enforcement hypothesis the county is provided with an incentive to demand full performance by the company and simultaneously attempt to induce breach in order to collect the \$20,000.

Inefficient performance is the first problem posed by the changed conditions dilemma. The efficiency criterion indicates that the bridge should not be built where the return performance is no longer valued by the promisee. Given ordinary market assumptions, however, the bridge never will be built. Since performance has become less valuable to the promisee he is induced to negotiate with the promisor to permit him to buy out from performance. The county would be willing to accept any payment over \$1,000, and the bridge company would be willing to pay any amount less than \$15,000. These sums represent the respective gain and loss which will accrue to each party if the bridge is completed. At some point, they will negotiate an agreement to divide the gains from breach that would be captured entirely by the breaching bridge company under the current legal rule.

Clearly, renegotiation produces transaction costs that would not arise under the penalty rule. It is tempting to measure these costs against the expected savings in litigation costs and propose whichever rule promotes the least costly alternative. However, since this measurement can only be made *ex post* by the legal mechanism, taking this approach would produce additional litigation costs in all cases. The enforcement hypothesis is the more efficient rule, because it permits an *ex ante* measurement of the expected costs.⁸⁴ The parties must be assumed to have taken all the expected costs into consideration at the time they elected to negotiate an agreed damages provision. In sum, the very existence of a freely negotiated agreed damages provision is compelling presumptive evidence that it constitutes the cost-minimizing alternative. This analysis is consistent with contract rules which assign responsibility for allocated risks, and excuse performance only where circumstances indicate risk was unassigned.⁸⁵

84. It could be hypothesized that if the legal rule could be constructed precisely to categorize and enforce only those agreements where renegotiation costs are less than the litigation costs, *ex post* measurement would be superior. This hypothesis is rejected because such a rule would require numerous factual considerations, and would necessarily be relatively imprecise, *see* Ehrlich & Posner, *supra* note 55, at 261-62, and since the total costs of litigation are not imposed on the losing litigant, there is an incentive to attempt to secure non-enforcement even where it is probable that the agreement will be found enforceable. This would impose transaction costs also on those cases where enforcement would be required.

85. In general, in order for inability to perform fully to be considered a legal excuse, the

The second problem posed by the changed conditions dilemma is the increased incentive that the promisee has to induce breach in order to capture the \$19,000 premium contained in the *in terrorem* clause. The prospective payoff would encourage the county to urge full performance and then attempt to induce a breach rather than negotiate an efficient release from the contract. The enforcement hypothesis would produce an increase of some amount in the social costs of policing against fraud or non-cooperation by the promisee. However, absent empirical data indicating the extent of these costly effects, the least restrictive assumption—invalidation for induced breach when those facts are identified—is suggested again. In any event, the moral risk produced by the changed conditions dilemma is reduced where the promisor provides the insurance. Third-party insurance in this situation would produce the heightened risk of collusive breach between the contracting parties.⁸⁶

D. *Conditions for Non-Enforcement: Unfairness and Bargaining Abnormalities*

1. *Presumption of Fair Exchange.* The underlying premise of the enforcement hypothesis is that, in the absence of bargaining unfairness, a stipulated damage clause reflects equivalent value. The possibility that a given provision does not reflect subjective compensation, but is penal in nature, is irrelevant to the question of enforcement unless this fact is caused by bargaining abnormalities. This premise is a derivative of what can be described as the flexibility principle of private exchange.⁸⁷ Assuming no violation of process constraints, the subjective value of exchange is not amenable to judicial scrutiny.⁸⁸ Except by controlling the subject

failure of performance must be caused by factors which were not allocated risks of the bargain. A failure of presupposed conditions (*e.g.*, destruction of specific goods) or unforeseeable, and therefore unallocated, circumstances will avoid the contract to the extent of the failure. See U.C.C. §§ 2-613, 2-614, & 2-615. Allocation of casualty risks by agreement has long been a permissible exercise of contracting flexibility. For example, sales contracts frequently expand the excuses for nonperformance well beyond the parameters of legal excuse outlined above. See 1 A. CORBIN, *supra* note 3, § 148; 3 *Id.* § 642; 6 *Id.* § 1331. Similarly, agreements frequently allocate risks to the promisor which would otherwise form a basis of legal excuse. See, *e.g.*, RESTATEMENT OF CONTRACTS §§ 451 (Comment c), 456 (Comment c), 458 (Comment d), 465 (Comment g) (1932).

86. See text accompanying notes 77-78 *supra*.

87. This principle has been variously described as the freedom or power to contract. It most readily identifies the efficiency criterion—permitting resources to flow to their highest valued uses.

The following model is suggested for the application of the flexibility principle: In a costless environment, the sole principle of market exchange would be flexibility. The parties would negotiate deals without constraints, and renegotiate upon non-performance with the only role of legal institutions being to enforce the private agreement, including sanctions, according to its terms. Facing positive transactions costs, however, the legal system provides ready-made rules based on common assumptions about typical contracting behavior. These "off the rack" contract rules reduce the costs of exchange by specifying the legal consequences of typical bargains where the expected cost of explicit negotiation exceeds the utility derived from individualized exchange. It is only where idiosyncratic value exceeds negotiating costs, therefore, that contractual flexibility induces privately concocted alternative arrangements.

88. The doctrine of adequacy of consideration, for example, represents an explicit at-

matter,⁸⁹ no neutral principle has been devised to evaluate the relative worth of a voluntary, freely-bargained exchange.⁹⁰ Instead, contracts doctrine has developed fairness constraints which focus on the maintenance of process values—full access to information and competitive market opportunities.⁹¹ The enforcement hypothesis relies on this jurisprudential tradi-

tempt to reject the moral imperatives of the early common law principle of objective value. The basic premise of the adequacy doctrine is that if there is a legal (*i.e.* "sufficient") consideration for a promise the courts will not question its adequacy. *Buckner v. McIlroy*, 31 Ark. 631, 634 (1877); RESTATEMENT OF CONTRACTS § 81 (1932). As expressed by the Restatement, "the relative values of a promise and the consideration for it do not affect the sufficiency of consideration." *Id.* The rationale for the doctrine is that the parties to a bargain, and not the courts, are best suited to determine the adequacy of an exchange. Therefore, "inadequacy of consideration, exorbitance of price or improvidence in a contract will not, in the absence of fraud, constitute a defense. . . ." *Hotze v. Schlarsen*, 410 Ill. 265, 102 N.E. 2d 131, 133 (1951). The doctrine does not, however, prevent a court from questioning the integrity of the bargaining process in cases in which the consideration given for a promise is so insufficient as to indicate that fraud, unfairness or unconscionability entered into the bargaining process.

Courts, though they have long arms, cannot relieve one of the consequences of a contract merely because it was unwise. They are not guardians in general to the people at large, but where inadequacy of price is such as to shock their conscience equity is alert to seize upon the slightest circumstance indicative of fraud, either actual or constructive.

Planters Nat'l Bank v. Heflin Co., 166 Va. 166, 173, 184 S.E. 216, 219 (1936). *See also* Horowitz, *supra* note 12, at 918, 923-27; 35 COLUM. L. REV. 1090-91 (1935).

In spite of the perceived movement in American law toward increased scrutiny of private exchange, the general principle of party autonomy remains. *See, e.g.*, Braucher, *Freedom of Contract and the Second Restatement*, 78 YALE L. J. 598 (1969).

89. *See generally* note 12 *supra*; G. GRISMORE, *supra* note 3, at 499. Traditionally, control over the subject matter of a bargain has been founded on the doctrine of illegality. This doctrine prohibits the making and, in some instances, the enforcement of certain agreements. It does so for much the same reason that the criminal law prohibits certain conduct, because such conduct is "contrary to the best interests of society. . . ." *Id.* In determining which agreements are, in fact, illegal, courts rely alternatively upon statutes and the broad rubric of "public policy." For example, bargains involving a wager have long been held to be illegal and unenforceable, based in part on the rationale that they act so as to create risk rather than to allocate an already existing risk. For a full discussion of illegality, see 6A A. CORBIN, *supra* note 3, §§ 1373-1541.

90. Instead of attempting to evaluate the end result of exchange, the doctrinal concern of bargain contract has traditionally centered on the necessity of reliable, objective evidence of a voluntary exchange. *See generally* L. FRIEDMAN, *supra* note 12, at 29-32, 69-77; J. W. HURST, *supra* note 12, at 11-12; Horowitz, *supra* note 12. The objective test of the existence of bargain, the formal doctrine of consideration focusing on the quid pro quo, and the benefit-detriment expression of an exchange arrived at by bargaining—all represented efforts to provide legal support to the operation of a free entry market.

Professor Speidel views the bargain theory as serving crucial needs in a market economy by providing formal controls to channel the market exchange, protecting the exchange transaction and "shielding the creative or idiosyncratic bargainer from later claims that the agreed exchange was disproportionate." Speidel, *An Essay on the Reported Death and Continued Vitality of Contract*, 27 STAN. L. REV. 1161, 1170 (1975).

91. The most ambitious effort to bring the traditional categories of unfairness—fraud, duress, incapacity—under a single analytical umbrella is represented in the emerging doctrine of unconscionability. As reflected in U.C.C. § 2-302 the unconscionability principle is largely devoid of substantive context. Some commentary has concluded that efforts to provide any analytical structure are essentially futile—that unconscionability represents the subjective reactions of the particular decision-maker to the unfairness of a particular transaction. *See* Leff, *Unconscionability and the Code—The Emperor's New Clause*, 115 U. PA. L. REV. 485 (1967).

The major thrust of much of the unconscionability scholarship has been to seize upon Comment 1 to U.C.C. § 2-302 proscribing "unfair surprise" and "oppression." These two factors have been described respectively as "procedural" and "substantive" unconscionability. *See id.* at 503, and have been used by many in an effort to merge the traditional constraints identified earlier. *See generally*, Braucher, *The Unconscionable Contract or Term*, 31 U. PITT. L. REV. 337 (1970); Murray, *Unconscionability: Unconscionability*, 31 U. PITT. L. REV. 1 (1969); Spanogle, *Analyzing Unconscionability Problems*, 117 U. PA. L. REV. 931 (1969);

tion by incorporating a presumption of fair exchange.⁹² Under this presumption, evidence that equivalent value was not exchanged for the liquidated damages provision would be relevant only to the extent that it permitted an inference as to the relative unfairness of the bargaining process. This analysis has been consistently applied by courts and legislatures to agreements for underliquidated or "limited" damages. These partial allocations of the risk of breach to the nonbreacher have been enforced absent specific evidence of unfairness.⁹³ The enforcement hypothesis, by validating liquidated damage clauses which allocate similar risks to the breacher, does not raise any unique dangers of fraud or duress. Rather, the two situations present perfectly symmetrical fairness issues.⁹⁴

Speidel, *Unconscionable Assent and Consumer Protection*, 31 U. PITT. L. REV. 359 (1970). All of these efforts at providing analytical clarity to the unconscionability doctrine seem to founder on the problem of structuring a neutral principle by which a contract can be determined to be substantively unfair.

The problem of substantive evaluation of the utility of the bargain can be avoided by limiting the unconscionability principle to a process control—falling solely within procedural unconscionability. The doctrine can thus be viewed as a method of defining bargaining abnormalities more precisely than through the traditional doctrines of fraud and duress as well as a more precise status control where assumptions concerning competitive market conditions are invalid. See text accompanying notes 95-103 *infra*. See also Epstein, *Unconscionability: A Critical Reappraisal*, 18 J. L. & ECON. 293 (1976) for a recent discussion which argues for the limitation of the unconscionability principle to a process control.

92. The primary significance of the presumption of fair exchange is the reallocation of the burden of proof of unfairness from the promisee to the promisor. The impact of the burden of proof under the penalty rule is illustrated in *Waggoner v. Johnston*, 408 P.2d 761 (Okla. 1965), where the court allocated the burden of establishing that the damages were difficult to ascertain on the non-breacher seeking to enforce the agreement. The fact that the parties have expressly stated that the damages are difficult to estimate does not shift the burden of proof to the breacher under present law. *Id.* at 768.

93. The disparity between the controls on underliquidated provisions and the limitations on penal agreements has long been recognized. See, e.g., 5 A. CORBIN, *supra* note 3, § 1068: Public policy may forbid the enforcement of penalties against a defendant, but it does not forbid the enforcement of a limitation in his favor. Parties sometimes make agreements and expressly provide that they shall not be enforceable at all, by any remedy legal or equitable. . . . Where a contract provides that damages for breach shall not be recoverable beyond a specified sum, it is obvious that the risk of loss beyond that sum is being assumed by the promisee. If the law allows him to assume the whole risk, with no remedy whatever, it is obvious that it will allow him to assume a part less than the whole.

See also 3 S. WILLISTON, *supra* note 3, at § 781A; RESTATEMENT OF CONTRACTS § 339, Comment g (1932); Sweet, *Underliquidated Damages As Limitations of Flexibility*, 33 TEX. L. REV. 196, 203-06, 212-19 (1954).

In *Cellulose Acetate Silk Co. v. Widnes Foundry*, [1933] A.C. 20 (H.L.), Lord Atkin in sustaining an agreement to limit delay damages to £20 per week stated:

It appears to me that such sum is provided as compensation in place of the no compensation at all which would otherwise have been the result. . . . I agree that it is not a preestimate of actual damage. I think it must have been obvious to both the parties that the actual damages would be much more than £20 a week; but it was intended to go toward the damage, and it was all that the sellers were prepared to pay.

Id. at 25 (emphasis added).

A similar result was reached in *American Dist. Tel. Co. v. Roberts & Son, Inc.*, 219 Ala. 595, 122 So. 837 (1929), where the court sustained an underliquidated damage agreement fixing recovery for defects in a fire alarm system to \$50.00. The Court stated:

As we view this clause, it is but a limitation of the amount recoverable in case of the breach of the contract. . . . Conceding, however, that, by limiting the amount of recoverable damages, it operates as a burden or hardship on the plaintiff, still we are not at liberty, by analogous authorities, to make a new contract for the parties or to strike therefrom a clause well understood and evidently within the intention of the parties.

Id. at 598 (emphasis added).

94. Although the analysis is identical in both cases, the legal rules diverge. Underliqui-

2. *Unfairness and the Efficiency Criterion.* The enforcement hypothesis identifies unfairness or other bargaining aberrations as the only limitations on the use of liquidated damages provisions. The normative notions of fairness implicit in the common law tradition are consistent with the analytical model of economic efficiency. Bargaining unfairness precludes the assumption of fair exchange and increases the risk of allocative inefficiencies. The inefficient effects of unfairness include an increase in the incidence of erroneously valued exchange as well as the increased social costs of fraud, misrepresentation, and duress. Asserting the inefficiencies of unfairness is not helpful analytically unless neutral principles can be identified within the fairness rubric. In the bargain context, two neutral principles may justify constraints on contracting flexibility.

Access to information at minimum cost is the first principle of bargain fairness. Where the bargain reflects processes which inhibit information exchange, the risk of allocative inefficiencies is enhanced. This constraint, identified in the unconscionability doctrine as "unfair surprise,"⁹⁵ would incorporate contracting behavior ranging from fraudulent exchange of false or misleading information⁹⁶ to failures to reasonably disclose essential contract terms.⁹⁷ This incentive to information exchange will maximize

dated damages have been traditionally subject only to a process constraint of bargaining fairness and not to the additional controls that have been imposed on liquidated damage agreements. See *Personal Fin. Co. v. Meredith*, 39 Ill. App.3d 695, 350 N.E.2d 781 (1976); *Shaer Shoe Corp. v. Granite State Alarm, Inc.*, 110 N.H. 132, 262 A.2d 285 (1970). The clearest portrait of this anomaly is captured by U.C.C. § 2-718 and its comments.

Comment 1 provides: "A term fixing unreasonably large liquidated damages is expressly made void as a penalty. An unreasonably small amount would be subject to similar criticism and might be stricken under the section on unconscionable contracts or clauses." (emphasis added).

The clear import of this comment is that underliquidated damages are subject only to the fairness limitations of § 2-302, while "overliquidated" provisions are subject to the additional, absolute constraint of the penalty rule in § 2-718(1).

95. U.C.C. § 2-302, Comment 1. Section 2-302 is the most significant statutory expression of the unconscionability doctrine. Other than directing the focus of the court to the bargaining process, the statutory language contains no substantive guidance as to the nature of the bargaining unfairness to be found void under the doctrine. Comment 1 to § 2-302 contains an expression of principle: "The principle is one of the prevention of oppression and unfair surprise . . . and not of disturbance of allocation of risks because of superior bargaining power." Illustrations of the "unfair surprise" component of bargaining unfairness would presumptively include traditional instances of fraudulent misrepresentation, as well as attempts to obscure the nature of the exchange by disguising relevant information in extra language or contract terms. See, e.g., *Murray*, supra note 91, at 16-18; *Slawson*, *Mass Contracts: Lawful Fraud in California*, 48 S. CAL. L. REV. 1, 11-14 (1974). The doctrine is typically applied to the paradigm "fine print" provision in a standard form contract creating a presumption that there was no consent to the bargain in fact.

96. Traditional controls on false or misleading information incorporate fraudulent misrepresentation or deceit as well as negligent and innocent misrepresentation. See, e.g., *King v. O'Reilly Motor Co.*, 16 Ariz. App. 518, 494 P.2d 718 (1972) (car sold "as new" was a presumably wrecked demonstrator); *Town & Country Chrysler Plymouth v. Porter*, 11 Ariz. App. 369, 464 P.2d 815 (1970) (car odometer turned back); *Eno Brick Corp. v. Barber-Greene Co.*, 109 N.H. 156, 245 A.2d 545 (1968); *Lanners v. Whitney*, 247 Or. 223, 428 P.2d 398 (1967). See also Epstein, supra note 91, at 298-300.

97. Simple non-disclosure of material facts has rarely been considered the kind of information barrier which will preclude enforcement of the agreement. On the other hand, evidence of concealment or partial disclosure will frequently bar enforcement of the resulting bargain.

The concealment issue is manifested in standardized contracts in the requirement that provisions allocating significant risks be conspicuous. See *Lacks v. Bottled Gas Corp.*, 215 Va. 94, 205 S.E.2d 671 (1974); U.C.C. §§ 1-201(10), 2-316(3).

efficiency by reducing the transaction costs of acquiring information.

The second fairness principle supports the maximizing of competitive market opportunities. Bargaining aberrations which inhibit competitive exchange will tend to produce inefficient resource allocation. The identifiable bargaining abnormalities would encompass duress⁹⁸ as well as the more traditional cases of monopoly.⁹⁹ The fairness value of enhanced market opportunities has also traditionally been reflected in the unconscionability doctrine. Scrutinizing a bargain produced by "oppression"¹⁰⁰ or "absence of meaningful choice"¹⁰¹ is a response to the perceived inefficiencies of reduced markets. The benefits of this response by the private law doctrine of unconscionability, however, remain indeterminate.¹⁰²

If this elaborated definition of unfairness is incorporated into the enforcement hypothesis, the following decision rule would be proposed:

Liquidated damage provisions should be enforced in all cases unless evidence of information barriers or reduced competitive opportunities rebuts the presumption of fair exchange.¹⁰³

3. *Party Sophistication and Presumptions of Unfairness.* The jurisprudential anomaly of the penalty rule is the imposition of a second level fairness constraint. There is no reason to presume that liquidated damages provisions are more susceptible to duress or other bargaining aberrations than other contractual allocations of risk. Consequently, the extraordinary limitation seems to produce many more costly effects than are warranted

98. Duress focuses on impermissible means used to obtain consent to the exchange. Duress requires an individual to abandon one legal right (property or physical integrity) in order to secure another. See Dawson, *Economic Duress—An Essay in Perspective*, 45 MICH. L. REV. 253 (1947); Epstein, *supra* note 91, at 295-698.

99. The traditional objection to the monopolist is that he will only offer additional units of goods until his marginal revenue equals his marginal cost; an excess of marginal cost over marginal revenues would imply losses on additional units produced. Since marginal revenue is less than price, society evaluates additional output at a magnitude higher than its additional cost, but it is no longer profitable for the monopolist to produce such socially desirable output increments. In that sense a monopoly is inefficient. Asserting the inefficiency of monopoly does not, however, necessarily require the conclusion that a sale by a monopolist should be considered unfair. The buyer from the monopolist is clearly different from the buyer facing information barriers, where it can be argued that the individual bargainer has misperceived his utility. Professor Alan Schwartz has argued that the monopolist will ascertain and supply buyer preferences as well as competitive sellers. Schwartz, *Seller Unequal Bargaining Power and the Judicial Process*, 49 IND. L.J. 367 (1974). However, if the monopolist's utility function is perceived as a mix of price, quantity, and other qualitative factors (*i.e.*, plus or minus risks), it may be that although consumers would pay for a different risk allocation, the monopolist will not offer it.

100. U.C.C. § 2-302, Comment 1.

101. *Williams v. Walker-Thomas Furniture Co.*, 350 F.2d 445 (D.C. Cir. 1965).

102. Even conceding the inefficiencies of monopoly, the difficult question remains. How should private contract law respond? One possible response is to refuse to enforce certain risk allocations where the seller has monopoly power. This would require the monopoly power to be exercised through price alone, which might reduce the search costs of disclosing monopolies. However, any benefits in aiding antitrust enforcement may well be exceeded by the litigation costs of challenging contracts of monopolists as unconscionable. After a thorough analysis of the problem Professor Schwartz argues that monopoly should not be a relevant factor in determining whether or not to declare a contract unconscionable. See generally Schwartz, *Unconscionable Factors*, 63 VA. L. REV. — (1977).

103. See text accompanying notes 87-94 *supra*.

by the perceived risk of unfairness. Nonetheless, it is clear that party sophistication will often be a relevant issue in determining the fairness of a stipulated damages provision.¹⁰⁴ Many contracting parties may not be capable of calculating the risks necessary to bargain for the *in terrorem* clause at an equivalent price. It is clear that some parties are incompetent to act as direct insurers of idiosyncratic value.¹⁰⁵

The problem of status does not justify the current rule under which these agreements are conclusively unenforceable in all cases. Nonetheless, a presumption of unfairness (and unenforceability) might well be appropriate in those factual contexts where the expected unfairness costs exceed the expected gains from unlimited contracting flexibility. For instance, if there exists an identifiable class of cases where application of the enforcement hypothesis predictably produces a high incidence of unfairness, the social costs can be reduced by attaching the unfairness presumption to those cases alone. This less restrictive limitation on contracting flexibility could be rebutted by the promisee's demonstrating that the clause was a product of a fairly-bargained exchange. As part of his burden of proof, the promisee would be required to demonstrate that the parties had sufficient commercial sophistication and access to information to allocate fairly the identified risks.

CONCLUSION

The historical background of the penalty rule discloses initial judicial interference to protect against fraud and duress in a legal context where alternative, less costly, protections were not available. Subsequently, invalidation was grounded on a presumption of unfairness based on indications that information barriers prevented rational assessment of the nature and extent of the risk allocations produced by the agreement. The costs of identifying unfairness in individual cases generated pressure for a rule invalidating these clauses on more precise criteria. Applying the compensation limitation to liquidated damages appeared to satisfy these wants and consequently received wide acceptance. Since the roots of the penalty rule were nourished on fairness concerns, it is not surprising that generations of lawyers have clung to the view that penalties are "bad." This notion does

104. The following statement is an example:

[C]ourts are beginning to look with favor upon stipulated damage provisions between parties who have equality of opportunity for understanding and insisting upon their rights.

Waggoner v. Johnston, 408 P.2d 761, 770 (Okla. 1965).

105. The problem of inability of certain contracting parties to maximize their own welfare through liquidated damages clauses should be regarded as part of the general question of incompetence, which, like infancy, insanity and drunkenness, rebuts the presumption of fair exchange. See Epstein, *supra* note 91, at 300-01. A similar analysis has been proposed as a limitation on state enforcement of consensual arrangements for provisional creditor remedies upon default. See Scott, *Constitutional Regulation of Provisional Creditor Remedies: The Cost of Procedural Due Process*, 61 VA. L. REV. 807, 862-64 (1975).

not withstand rigorous, dispassionate analysis. The current penalty rule does not promote end results which are any "fairer" than an enforcement rule. The behavior which requires regulation is unfairness in bargaining. The penalty rule, however, fails to mirror accurately the proscribed behavior.¹⁰⁶ Consequently, the penalty rule is used as a second level control together with standard restrictions on process unfairness.

Challenging the penalty rule demonstrates that it has numerous costly effects. First, the rule denies true compensation to the promisee with non-provable idiosyncratic wants, inducing him either to protect those wants with inefficient third party insurance¹⁰⁷ or to suffer exposure to inefficient breaches.¹⁰⁸ Secondly, assuming that cases of non-compensable idiosyncratic value are rare, the rule produces a more significant cost by inducing review of the entire continuum of cases where liquidated damages provisions are intended to reimburse true losses which are to any extent uncertain.

The modern development of unconscionability as a unifying fairness principle presents a less costly alternative to the sweeping invalidation powers exercised under the penalty rule.¹⁰⁹ With our present enhanced access to information and consequential greater accuracy in individual risk evaluation, an enforcement model which facilitates the recovery of difficult-to-prove values would appear to maximize the allocative efficiency of the contracting process.

In sum, contemporary cost-benefit analysis suggests that the traditional penalty rule is anachronistic for several reasons: (1) the efficiency costs of the rule are now apparent in the light of modern analysis; (2) the market imperfections once addressed by the rule have become empirically less important; and (3) more selective legal doctrines, such as unconscionability, have developed as remedies for those market imperfections which retain practical importance.

106. See Ehrlich & Posner, *supra* note 55, at 262-68. The penalty rule is both underinclusive and overinclusive. That is, some bargains which are infected by unfairness are not proscribed; while others—as in the case of idiosyncratic reliance—are invalidated in spite of a fairly-bargained exchange.

107. See text accompanying notes 66-71 *supra*.

108. See note 55 and accompanying text *supra*.

109. A number of courts have indicated a willingness to limit the penalty rule to a fairness test. See, e.g., *Gruschus v. C. R. Davis Contracting Co.*, 75 N.M. 649, 409 P.2d 500 (1965) ("As a general rule, enforcement of such a clause will only be denied when the stipulated amount is so extravagant or disproportionate as to show fraud, mistake or oppression.').