University of Virginia Law School
The John M. Olin Program in Law and Economics Working Paper
Series

Year 2004

Paper 4

Embedded Options and the Case Against
Compensation in Contract Law

Robert E. Scott*

George G. Triantis[†]

*University of Virginia School of Law, res8f@virginia.edu

[†]University of Virginia School of Law, gtriantis@law.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://law.bepress.com/uvalwps/olin/art4>

Copyright ©2004 by the authors.

Embedded Options and the Case Against Compensation in Contract Law

Robert E. Scott and George G. Triantis

Abstract

Despite the fact that compensation is the governing principle in contract law remedies, it has tenuous historical, economic and empirical support. A promisor's right to breach and pay damages (which is subject to the compensation principle) is only a subset of a larger family of termination rights that do not purport to compensate the promisee for losses suffered when the promisor walks away from the contemplated exchange. These termination rights can be characterized as embedded options that serve important risk management functions. We show that sellers often sell insurance to their buyers in the form of these embedded options. We explain why compensation is of little relevance to the option price agreed to by the parties, which is a function of the value of the option to the buyer, its cost to the seller and the market in which they transact. We thus propose a novel justification for why penalty liquidated damages may be higher than seller's costs: they are option prices that reflect the value of the options to the buyer. The regulation of liquidated damages is thus tantamount to price regulation, which is outside the realm of contract law. Moreover, in light of the heterogeneity among optimal option prices, we also make the case against having an expectation damages default rule to begin with. In thick markets, we argue for enforcing the parties ex ante risk allocation with market damages. In thin markets, we propose that parties be induced to agree explicitly with respect to all termination rights, including breach damages, by the threat of specific performance of their contemplated exchange or, in the case of consumers, by a default rule that provides them a termination option at no cost.

EMBEDDED OPTIONS AND THE CASE AGAINST COMPENSATION IN CONTRACT LAW

*Robert E. Scott**
*George G. Triantis***

Revised Second Edit Draft: June 10, 2004

Despite the fact that compensation is the governing principle in contract law remedies, it has tenuous historical, economic and empirical support. A promisor's right to breach and pay damages (which is subject to the compensation principle) is only a subset of a larger family of termination rights that do not purport to compensate the promisee for losses suffered when the promisor walks away from the contemplated exchange. These termination rights can be characterized as embedded options that serve important risk management functions. We show that sellers often sell insurance to their buyers in the form of these embedded options. We explain why compensation is of little relevance to the option price agreed to by the parties, which is a function of the value of the option to the buyer, its cost to the seller and the market in which they transact. We thus propose a novel justification for why penalty liquidated damages may be higher than seller's costs: they are option prices that reflect the value of the options to the buyer. The regulation of liquidated damages is thus tantamount to price regulation, which is outside the realm of contract law. Moreover, in light of the heterogeneity among optimal option prices, we also make the case against having an expectation damages default rule to begin with. In thick markets, we argue for enforcing the parties ex ante risk allocation with market damages. In thin markets, we propose that parties be induced to agree explicitly with respect to all termination rights, including breach damages, by the threat of specific performance of their contemplated exchange or, in the case of consumers, by a default rule that provides them a termination option at no cost.

CONTENTS

INTRODUCTION

I. RECONSIDERING THE COMPENSATION PRINCIPLE IN CONTRACT LAW

- A. Doctrinal Dominance of the Compensation Principle
- B. Historical Roots of Compensation in Breach Remedies
- C. Economic Contract Theory and Compensation
- D. Commercial and Consumer Practices of Termination

*David and Mary Harrison Distinguished Professor of Law, University of Virginia; Justin W. D'Atri Visiting Professor of Law, Business and Society, Columbia Law School.

**Perre Bowen Professor of Law, University of Virginia.

We are grateful to workshop participants at Columbia, George Mason, Southern California, Virginia and Yale Law Schools, and particularly Barry Adler, Ian Ayres, Victor Goldberg, Edward Iacobucci, Bruce Johnson, Eric Posner and Eric Talley for helpful comments on earlier drafts. We also thank Jason Bregge and Maruti Racherla for valuable research assistance.

II. EMBEDDED OPTIONS

- A. Termination Rights and Call Options
- B. Real Options and Ex Post Efficiency
- C. Written Embedded Options and Insurance
- D. Risk Management and Insurance in Thick- and Thin-Market Contracts
- E. Pricing Embedded Options in Thin Market Contracts

III. EMBEDDED OPTIONS AND OPTIMAL DAMAGE DEFAULTS

- A. Conditions for Efficient Damages Defaults
- B. Optimal Damages in Thick-Market Contracts
- C. The Case Against Compensation in Thin-Market Contracts
- D. Bargain-Forcing Defaults

CONCLUSION

INTRODUCTION

Compensation is the governing principle in contract law remedies. This principle shapes the key doctrines that specify the consequences of breach. Expectation damages, the default measure of damages, aim to put the promisee in the position she would have occupied had the promisor performed, while specific performance is available at the option of the promisee only when the court believes that money damages are inadequate to compensate for her loss. Although parties may agree to liquidated damages, contract doctrine instructs them to abide by the compensation norm. But despite its profound influence on contemporary contract law doctrine, the compensation principle has tenuous historical, economic and empirical support. Its evolution in the common law resulted primarily from ill-conceived path-dependence; compensation is virtually ignored in the theoretical analysis of efficient contract design; and compensation plays little role in the contracts expressly negotiated by commercial parties and agreed to by consumers. We argue that the right to breach should be treated as a valuable option sold by the promisee to the promisor. As a result of an unfortunate turn in history, however, contract damages are instead analogized to compensation for wrongs. This has impeded the efficient evolution of both default remedies and the regulation of liquidated damages.

The right to breach and pay damages is a subset of a broader category of termination

rights that give one party an option to walk away from the contemplated exchange.¹ A firm offer or unilateral promise, for example, grants the promisee such an option. Broad warranties, such as satisfaction-or-your-money-back provisions, give buyers similar options. Requirements, output or installment contracts grant one party substantial discretion to determine the contract quantity. And a contract may provide that one party has the right to terminate, to cancel, to renew, or to return or redeem goods.

Either or both parties to a contract, therefore, commonly enjoy the right to terminate at some cost. For the purposes of our analysis and argument, we focus on the option held by a buyer of goods or services. The buyer holding an option has the right to avoid the exchange by paying either a termination fee or damages.² The price of an embedded option is determined just as the price of any other product: It is a function of the option's value to the option holder, the

¹It is well known that contract damages effectively give the promisor an option between performing the promise or breaching and paying damages. The classic statement is by Oliver Wendell Holmes: 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 mean a prediction that you must pay damages if you do not keep it—and nothing else. Oliver Wendell Holmes, *The Path of the Law*, 10 Harv. L. Rev. 457, 462 (1897).

Over the past decade, legal scholars have begun to analyze contract remedies explicitly in terms of options. In this article, we extend prior analyses to consider the full range of contractual options to terminate. Our normative claims follow from this broader perspective. For a review of the prior work on contract options, see Robert E. Scott, *The Case for Market Damages: Revisiting the Lost Profits Puzzle*, 57 U. Chi. L. Rev. 1155 (1990); George G. Triantis, *The Effects of Insolvency and Bankruptcy on Contract Performance and Adjustment*, 43 Univ. Tor. L. J. 679 (1993); Alexander J. Triantis and George G. Triantis, *Timing Problems in Contract Breach Decisions*, 41 J. Law & Econ. 163 (1998); Paul G. Mahoney, *Contract Remedies and Options Pricing*, 24 J. Legal Stud. 139 (1995); Victor P. Goldberg, *The Net Profits Puzzle*, 97 Colum. L. Rev. 524 (1997); Victor P. Goldberg, *Discretion in Long-Term Open Quantity Contracts: Reining in Good Faith*, 35 U.C. Davis L. Rev. 319 (2002); Avery Wiener Katz, *The Efficient Design of Option Contracts: Principles and Applications*, (forthcoming Va. L. Rev. 2004).

² Several previous articles characterize the right to breach as if the buyer held a put option on the agreed upon exchange that she could exercise by paying damages to the seller. By exercising the put, the buyer in effect sold the contract good or service back to the seller for the assessed damages liability. Triantis, *Effects of Insolvency and Bankruptcy*, supra note – at 680-4; Mahoney, supra note –; Triantis & Triantis, supra note – at 168-9, 169 n15. In this article, however, we analyze the buyer as effectively incurring an obligation to pay damages in exchange for a call option on the subject matter of the contract. The buyer's prospective liability for damages is effectively the price of the call option, as if the buyer makes a nonrefundable deposit or payment for the call option and pays an additional price to exercise it. The exercise price of the option is the difference between the contract price and the damages liability. For discussion, see text accompanying notes 100-102 infra.

cost to the option writer and the competitiveness of the market in which they transact. Options are essentially insurance contracts that divide risks according to their exercise prices.³ For any given exercise price, the option price divides between the parties the surplus created by the option; namely, the difference between the option's value to the buyer and its cost to the seller. The parties should choose a pairing of option price and exercise price that maximizes this surplus. Sometimes the option price is fixed by the contract (e.g. in the form of a nonrefundable deposit or termination fee) and at other times it is left to be judicially determined as damages for breach of contract. In light of the great variety in the conditions under which parties contract for this option, it should not be surprising that commercial and consumer contracts contain a wide range of option prices. We explain this heterogeneity in option prices and argue that they are equivalent only by coincidence to the measure of the seller's expectation in a completed sale.

Consider the electronics store that sells television sets for \$400 and offers full refunds for any returns made within 30 days. This contract gives the buyer a free 30-day call option to purchase the television set for \$400. The option is valuable to the buyer because she is uncertain as to the value of the television set to her family. The retailer in this case bears significant costs in accepting returns, including the cost of receiving, inspecting and reselling the returned goods, often through a discount outlet or internet sale. Yet, the retailer often does not charge for the call option.⁴ Such "free" options are remarkably common in both commercial and consumer

³In his Nobel lecture, Robert Merton stated: "When [an option is] purchased in conjunction with ownership of the underlying asset, it is functionally equivalent to an insurance policy that protects its owner against economic loss from a decline in the asset's value below the exercise price for any reason." Robert C. Merton, *Applications of Option-Pricing Theory: Twenty Five Years Later*, 88 Am Econ. Rev. 323, 336-7 (1998).

⁴ It is inaccurate to conclude that the cost of the option is included in the overall contract price because the buyer does not pay this price if she walks away from the option. Where the seller does not charge the buyer for an unexercised option, the seller recovers the cost from buyers who exercise the option. This cross-subsidization leads to adverse selection and moral hazard issues discussed later in the paper. In many of these cases, the option is in fact not free because the seller holds the buyer's payment of the contract price during the option term and, if the buyer has possession of the good during that period, she often must also bear the cost of bringing the product back to the seller.

contracts.⁵ They are particularly interesting for our purposes because these contracts make no attempt to compensate the seller for losses it suffers when the buyer walks away from the contemplated exchange.

Some buyers enjoy no option, however, because they pay (or owe) the entire purchase price and cannot escape this obligation by cancelling the transaction.⁶ Retailers, for example, may offer refunds for ordinary course sales but do not accept returns of sale merchandise. In the range between no-option and free-option, lie termination provisions under which buyers pay positive option prices in the form of fees or damages. These fees sometimes exceed the compensatory amount – for example, the seller's foregone profit. Consumers are familiar with instances of such overcompensation in a variety of transactions. For example, economy fares on airlines are typically conditioned on a penalty of \$100 if the passenger chooses to cancel and apply the ticket price against the fare for another flight.⁷ The penalty applies even on flights that are overbooked and almost certain to be full. Thus, the penalty bears no necessary relationship to compensation for the airlines losses.

Indeed, casual observation reveals that the termination provisions affecting consumers regularly depart from the predictions of the compensation principle. Examples of such embedded options abound in commercial contracting as well. Retailers often have the right to return unsold merchandise to the wholesaler or distributor. Rights of return are common, for example, in the retailing of books, journals, newspapers, musical compact discs, jewelry, and cigarettes.⁸ Aircraft manufacturers permit purchasers to cancel orders or to change the type of

⁵ Another example of a free option is the right of consumer borrowers to prepay their loans or mortgages without compensating their lenders for consequential losses that might be caused, for example, by declines in the market rate of interest. Similarly, a borrower who defaults is liable to pay the accelerated balance owing, but not compensation for the lender's foregone opportunities when market rates have dropped since the loan.

⁶Of course, fresh start in bankruptcy (and limited liability of some organizations) creates the well known option held by debtors, which we set aside in this article.

⁷In the event of a change in plans, the passenger also may have to pay any increase in the fare from the time she bought the ticket.

⁸Eugene Kandel, *The Right to Return*, 39 J. L. & Econ. 329, 330 (1996).

aircraft even after the manufacturer has made significant investment in production.⁹

In this article, we offer an explanation for these contracting patterns based on the insurance role played by embedded options. It is well known that compensatory remedies insure the buyer against the seller's breach and thereby against the risk of fluctuations in the seller's *cost* of performance that lead it to breach. The buyer, however, still bears the risk of fluctuations in the *value* of the seller's performance. Options created by termination rights insure buyers against this risk and thereby may promote the risk management objectives of business contractors.

When the buyer's option comes in the form of the right to breach and pay expectation damages, the buyer's liability on termination is a function of the profit that the seller would gain from the exchange. Under an expectation damages rule, therefore, the buyer shrugs off some of the risk of fluctuations in performance value, but assumes some of the risk in fluctuations in the seller's profit. If the buyer contracts for the right to terminate and pay a fixed amount rather than expectation damages, the buyer can also avoid the risk associated with the seller's cost. The buyer may be prepared to pay a premium in order to be able to shift both the risk in the value of performance and the risk in the seller's costs. In these cases, the parties may agree to pay fixed liquidated damages that are greater than the (ex ante) expected amount of the seller's loss from breach. Under the current doctrine, however, courts would refuse to enforce this term. Our analysis suggests a benign explanation for this supercompensatory damages term: It is the product of a negotiated sale of insurance from the seller to the buyer in the form of a call option.

Given the close link between options and insurance, it should not be surprising that the optimal terms of embedded options are a function of considerations that determine insurance contracts; namely, risk bearing capacity, adverse selection and moral hazard. Although we discuss the role of these considerations in the structuring of embedded options, our principal

⁹John Stonier and Alexander J. Triantis, *Natural and Contractual Real Options: The Case of Aircraft Delivery Options*, in A. Micalizzi and L. Trigeorgis, eds., *Real Options Applications* 159-195 (1999); John Stonier, *The Change Process*, in Tom Copeland and Vladimir Antikarov, *Real Options: A Practitioner's Guide* 47 (2001).

contribution is to highlight the heterogeneity of optimal terms and prices, which is consistent with our observation of a wide variety of termination provisions in practice.

This diversity has important normative implications for the default rules of contract damages and for the freedom of parties to contract away from the defaults. The characterization of breach damages as an option price separates damages from the compensation principle and reinforces the criticisms of the penalty rule that have been raised by contract scholars.¹⁰ The recent litigation concerning late fees charged by Blockbuster video stores is illustrative.¹¹ The court recognized that the right to extend the rental period was valuable, that a rational consumer would pay a price for this right that might be higher than the cost to Blockbuster, and that the court should not engage in regulation of this price.¹²

The courts are not well suited to set default prices either. In most contract relationships, an expectancy default rule is no more appropriate with respect to damages than with respect to the price of any other good, service or contract term for which there is no established market price. In thick markets, courts should simply enforce the risk allocation specified in the contract by awarding market damages. Outside thick markets, we argue in favor of a bargain-forcing approach. This could take one of two forms. First, the courts could adopt a general default of specific performance and thus refuse to recognize any option to terminate unless the parties clearly state the price and exercise price of the option. Alternatively, under certain circumstances the courts might grant a promisor the option to terminate at no cost if the parties failed to make an express provision regarding termination rights. Where commercial parties

¹⁰See note 94 *infra*. The compensation principle is also responsible for leading courts as well as scholars to search in the wrong place for solutions to lost-volume seller cases by debating alternative measures of the seller's loss. See TAN 164-167 *infra*.

¹¹*Pickens v. Blockbuster, Inc.*, 2004 WL 339594 (Cal. App. 1 Dist.).

¹²The California Court of Appeal held that the test whether the late fees were liquidated damages or provision for alternative performance is whether the customer would have a "realistic and rational choice" between two alternative performances (returning the video when due or keeping it). The court then ruled that "this is not a case in which no rational person would choose to pay the greater liability." *Id.* at 3. The court cited as authority *Blank v. Borden* 11 Cal.3d 963, 971 (1974).

have failed to provide expressly for termination rights, we propose that the courts specifically enforce the exchange. In the context of consumer transactions, where sellers typically draft the contracts, the default should give the consumer a free option.

The article proceeds as follows. In Part I, we contrast the contemporary dominance of the compensation principle of contract damages against its tenuous historical roots, its marginal relevance in contract economics and its limited use in commercial and consumer contracts. Part II characterizes the termination provisions available to a buyer – including the right to breach and pay damages -- as call options that serve an important risk management function. We demonstrate that options prices, including breach damages, might take a wide range of values. We specify factors that explain the variations in the price paid for these options and show that these may appear either undercompensatory or overcompensatory to a court. Part III draws the implications for optimal default rules for breach of contract summarized above. The Conclusion summarizes.

I. RECONSIDERING THE COMPENSATION PRINCIPLE IN CONTRACT LAW

A. Doctrinal Dominance of the Compensation Principle

The doctrinal view of contract breach is that it is a wrong that is remedied by compensating the victim for her loss.¹³ This view is reflected in the default provisions for damages and the constraints imposed on the freedom of parties to stipulate other measures of damages by contract. The Restatement of Contracts, for example, states that “[t]he traditional goal of the law of contract remedies has not been compulsion of the promisor to perform his

¹³ See, e.g. Uniform Commercial Code § 1-305 (2003) (formerly §1-106): “The remedies provided by this Act shall be liberally administered to the end that the aggrieved party may be put in as good a position as if the other party had fully performed...”; Restatement(Second) of Contracts §347(1979). Although the compensation principle is firmly enshrined in the Code and the Restatement, we suggest in the next section that contract law embraced the compensation principle relatively late in its common law development.

promise but compensation of the promisee for the loss resulting from breach.”¹⁴ Since the classic article by Fuller and Perdue in 1936,¹⁵ courts and scholars have typically compared three alternative bases for evaluating the promisee’s entitlement to compensation: restitution, reliance and expectation. The dominant measure is expectation damages--the payment necessary to put the promisee in as good a position as if the promisor had performed.¹⁶ Over the past 150 years, the courts have expanded the availability of compensation for the promisee’s consequential losses, subject only to relatively modest constraints of mitigation, foreseeability and uncertainty.¹⁷ Specific performance is available only when the court determines that money damages are inadequate to compensate for the promisee’s loss.¹⁸

The most notable characteristic of the compensation principle is its expansive reach: The compensation default applies in all contexts and, in addition, purports to justify the legal regulation of parties’ ability to stipulate damages expressly in their contract. Although most of contract law provides default rules from which parties are free to contract away, remedial defaults carry heavier presumptive weight than other provisions. Contract breach terms may not deviate from the compensation principle. According to doctrinal statements, the only permissible ground for stipulating damages is the anticipated difficulty of measuring the

¹⁴Restatement (Second) of Contracts, Intro. Note to Ch. 16, Remedies. See also, E. Allan Farnsworth, *Legal Remedies for Breach of Contract*, 70 Colum. L. Rev. 1145, 47 (1970) (describing the common view that the central purpose of contract damages is compensation); UCC § 1-305, Comment 1 (same).

¹⁵Lon L. Fuller and William R. Perdue, Jr., *The Reliance Interest in Contract Damages: I*, 46 Yale L.J. 52 (1936).

¹⁶Restatement (Second) of Contracts §§344, 347(1979). “The initial assumption is that the injured party is entitled to full compensation for his actual loss.” Intro. Note to Topic 2, Enforcement by Award of Damages.

¹⁷In a recent article, Richard Craswell observes that courts appear willing to soften or disregard these traditional limitations when they find that the breach was willful. Richard Craswell, *Against Fuller and Perdue*, 67 U. Chi. L. Rev. 99, 138-43 (2000). Craswell cites Arthur Corbin as noting that “a lesser degree of certainty will be required as against one whose breach is described as ‘willful’ or is motivated by malice or avarice than against one whose breach was due to misfortune and whose efforts to perform were honest and in good faith.” *Id.* at 140. Craswell also speculates that courts may choose reliance damages as means of effecting some sharing of losses between contracting partners (for example, in response to an unforeseen contingency).

¹⁸Restatement (Second) of Contracts §359. See ROBERT E. SCOTT & JODY S. KRAUS, *CONTRACT LAW AND THEORY* 116-118, 992-95 (3D. ED. 2003).

promisee's loss. For instance, the Restatement of Contracts and Article 2 of the Uniform Commercial Code each provide that stipulated damages must be an amount that is reasonable in light of the anticipated or actual loss caused by the breach and the difficulties of proving the amount of loss.¹⁹ In practice, the courts strike down supercompensatory liquidated damages far more often than they police undercompensatory limitations on damages.²⁰ In cases in which it is especially difficult to measure the promisee's loss, a court may order specific performance or an injunction even in the face of a contractual provision for liquidated damages.²¹ This additional basis for setting aside liquidated damages provides further evidence that the law regards measurement difficulty as the sole legitimate reason for deviating from the default of expectation damages. Although contracting parties attempt to circumvent the regulation of liquidated damages by providing explicitly for alternative contracts such as take-or-pay or pay-or-play provisions, even these drafting techniques are more likely to succeed when the alternative payment has a compensatory basis, such as reimbursing fixed costs incurred by the seller.²²

¹⁹ UCC § 2-718 (2003); Restatement (Second) of Contracts §356 (1979). Restatement (Second) of Contracts § 347, Comment a states that "The parties to a contract may effectively provide in advance the damages that are to be payable in the event of breach as long as the provision does not disregard the principle of compensation." UCC § 2-718, Comment 1 states: "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." The modern penalty rule found in the Code and the Restatement is derived from a line of common law cases invalidating any stipulated damages where the amount specified exceeded the "just compensation for the loss or injury actually sustained." *Jaquith v. Hudson*, 5 Mich. 123, 133 (1858). For discussion, see Charles J. Goetz & Robert E. Scott, *Liquidated Damages, Penalties and the Just Compensation Principle*, 77 Colum. L. Rev. 554, 555-557 (1977) and TAN *infra*.

²⁰ See e.g., *Martin Rispens & sons v. Hall Farms*, 621 N.E. 2d 1078 (1993) ("Courts have rejected claims that contractual limitations of remedy are substantively unconscionable"); 5 ARTHUR CORBIN, CONTRACTS, § 1068 (1964) ("public policy may forbid the enforcement of penalties against a defendant, but it does not forbid the enforcement of a limitation in his favor."); Justin Sweet, *Underliquidated Damages As Limitations of Flexibility*, 33 Tex. L. Rev. 196, 203-06 (1954). Although we would expect that undercompensatory liquidated damages are in fact at least as common as overcompensatory provisions, our review of reported cases between 1998 and 2003 reveals that the incidence with which courts strike down penalties far exceeds their rejection of undercompensatory damages provisions [We need more detail from Maruti's research here].

²¹ Restatement (Second) of Contracts § 361, Comment a states: Merely by providing for liquidated damages, the parties are not taken to have fixed a price to be paid for the privilege not to perform. *The same uncertainty as to the loss caused that argues for the enforceability of the provision may also argue for the inadequacy of the remedy that it provides.* (emphasis added)"

²² E.g., *Lake River Corp. v. Carborundum Co.*, 769 F.2d 1284 (7th Cir. 1985); *Superfos Investments Ltd. v. Firstmiss Fertilizer, Inc.*, 821 F. Supp. 434 (S.D. Miss.1993).

B. Historical Roots of Compensation in Breach Remedies

Although now firmly entrenched in doctrine, the contemporary understanding of the compensation principle is a recent development in contract law. Well into the 19th century, contract enforcement was heterogeneous. The common law courts provided separate remedies for discrete contracting categories. Thus, for example, the courts a) awarded reimbursement damages for breach of informal contracts in thin-markets; b) awarded market damages in thick-market contexts such as commodities or stock transactions; c) specifically enforced contracts under seal, and d) generally enforced parties' intentions regarding liquidated damages. Beginning in the 19th century, however, the courts began to collapse these categories under an expanding compensation principle that they had adopted from tort law. Thereafter, the compensation notion gradually mutated to include full expectancy damages and this broad conception of compensation spread to each of the previously discrete transactions. This development severely narrowed the freedom of contracting parties to stipulate damages and contract for specific performance.

1. The Roots of the Compensation Principle

At early common law, there was no cause of action for breach of an informal (unsealed) executory promise. The only actions available for breach of contract were the action for debt and the action in covenant (for promises under seal).²³ The notion of compensation was foreign to either action. The action for debt lay only for the recovery of a sum certain. One party was seeking relief for a debt that was due and owing, fixed by the parties prior agreement, and the court in no sense awarded compensation for breach of contract.²⁴ The evolution of commercial exchange during the late middle ages increased the demand for a legal mechanism for enforcing

²³ JAMES BARR AMES, LECTURES ON LEGAL HISTORY 92 (1913); *Id.* at 92, 122-23; A. W. B. SIMPSON, A HISTORY OF THE COMMON LAW OF CONTRACT 47-48 (1986).

²⁴ AMES, LECTURES ON LEGAL HISTORY at 88-89. Where a seller tendered goods to a buyer and the buyer refused to accept delivery, the seller was able to sue in debt for the purchase price and force the buyer to take delivery of the goods (for which title had passed under the contract). Alternatively, if the buyer tendered the purchase price and the seller refused to transfer goods that were then available, the buyer's only recourse was to bring an action in equity for specific performance since the remedy at law was inadequate.

informal promises to supplement the self-enforcing mechanisms of the law merchant and medieval trade fairs.²⁵ The response of the English common law courts was to recognize a promisee's right to recover for breach of an informal promise by bringing an action in assumpsit.

Assumpsit had developed initially to provide a cause of action for the negligence of a bailee or carrier for hire.²⁶ The principle of compensation that supported the action in assumpsit was thus a distinctly tort notion premised on the idea of ex post redress for a harm committed by the defendant.²⁷ Over time, the action in assumpsit was extended to nonperformance of certain promissory undertakings.²⁸ Specifically, the action in assumpsit for breach of promise lay for plaintiffs who had either conferred benefits and/or had taken action in preparation for performance in reliance on the defendant's promise. In either case, a plaintiff who was seeking relief via assumpsit for breach of an informal promise was asking for compensation under a theory of *reimbursement* for the loss of that which had been given (directly or indirectly) to the promisee.²⁹

²⁵Avner Grief, *Informal Contract Enforcement: Lessons from Medieval Trade* in 2 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND LAW 287 (Peter Newman, ed. 1998).

²⁶SIMPSON, supra note — at 210-215. The traditional tort action of trespass on the case would not permit recovery for negligent bailment because the plaintiff might have been equally careless in entrusting a third party with his property. Id. The action in assumpsit for negligent bailment was first recognized in the *Humber Ferryman Case*, 22 Lib. Ass., Edw. III pl.41 (1348). The plaintiff alleged that defendant bailee *undertook* to carry his goods safely. The failure to perform this undertaking was the gravamen of the action, and, as was traditional in tort actions, the resulting injury to the plaintiff's property required *compensation*.

²⁷ AMES, supra note -- at 130. A parallel line of cases permitted recovery in deceit for a false warranty for goods sold and delivered. This action was also, in its origin, a pure action in tort. Id. At 136-7.

²⁸ Initially, a plaintiff was allowed to bring assumpsit only where the defendant performed his promise unskillfully (i.e., a carpenter who undertook to build a house for the plaintiff and performed poorly). Subsequently, the English courts held that a plaintiff could recover in assumpsit for the promisor's failure to act altogether. Id.

²⁹ During this early period of the action in assumpsit, a plaintiff could bring an action for breach of promise independent of the doctrine of consideration and the concept of exchange. The early notion of *special assumpsit* (the contract action) did not require a quid pro quo as was required for an action for debt which was explicitly tied to the notion of exchange. AMES, supra note – at 143-144.

For many years, the orthodox view among legal historians was that the final stage of the development of assumpsit – the enforcement of purely executory promises--matured in England sometime in the 16th or 17th centuries.³⁰ But the weight of evidence suggests that the enforcement of executory contracts occurred much later, at the end of the 18th century (in England) and the beginning of the 19th century (in America).³¹ Throughout the 18th century, contract law was still dominated by the action in debt, exchange was not conceived in terms of future returns and thus expectation damages were not recognized on either side of the Atlantic.³² During this pre-market period, courts almost universally declined to instruct juries on damages measures or to revise damage judgments (whether excessive or inadequate).³³ Thus, the common law courts that granted recovery for an action in debt were, in essence, specifically enforcing the actual bargain that the parties had struck.

Executory contracts thus were not enforced in the United States until the early 19th

³⁰ Morton Horwitz, *The Historical Foundations of Modern Contract Law*, 87 Harv. L. Rev. 917, 921-22 (1974); E. ALLAN FARNSWORTH, *CONTRACTS* 18 (4th ed. 2004).

³¹ See Horwitz, *Historical Foundations*, supra at 921-22 (1974)(arguing that enforcement of executory promises did not occur until the rise of industrialization and the development of commercial markets in the late eighteenth and early nineteenth centuries). Horwitz's basic thesis—that prior to the industrial revolution the common law of contract was dominated by notions of equity and fairness and that it was thereafter adapted to legitimate the inequalities of the nineteenth century market economy—has been vigorously contested. See e.g., A.W.B. Simpson, *The Horwitz Thesis and the History of Contracts*, 46 U. Chi. L. Rev. 533 (1979). Simpson's critique does not, however, challenge the basic point relevant to the emergence of the compensation principle—that courts did not regularly enforce executory contracts until the nineteenth century. Rather, as we discuss below, the penal bond was the preferred device for legal enforcement of commercial exchange transactions in the 17th and 18th centuries. See TAN 40-43 infra.

³² Horwitz cites only two English cases in the 18th century that raise the issue of expectation damages. Horwitz, supra note – at 921. In *Fleureau v. Thornhill*, 96 Eng. Rep. 635 (C.P. 1776), the court limited the plaintiff to restitution damages, holding that “plaintiff could not be entitled to damages for the fancied goodness of the bargain which he supposes he has lost.” In the United States, only a few actions for breach of executory contracts were brought before the Revolution. See, e.g., *Boehm v. Engle*, 1 Dall. 15 (Pa. 1767) where the seller was allowed to sue for the contract price of breach of a contract for the sale of land. *Id.* at 922.

³³ Chancellor Kent articulated this principle of contract damages as compensation being fixed by the jury “with a moderation agreeable to equity and good conscience and where the claims and pretensions of each party can be duly attended to...” *Seymour v. Delancy*, 6 Johns. Ch. 222, 232 (N.Y. Ch. 1822).

century.³⁴ This development coincided with a period of commercial expansion and with the emergence of markets in stock transactions and commodities.³⁵ Courts began awarding market-based damages for failure to deliver stock certificates in a rising market,³⁶ and for the breach of fixed-price forward contracts for the delivery of commodities.³⁷ This link between commodities and stock transactions and contract law was thus the major step in the development of a market damages default rule for breach of thick market contracts. Contract thereafter became an instrument for managing the exogenous price changes in well-developed markets.³⁸

2. *The Penal Bond and the Roots of Judicial Oversight of Liquidated Damages*

If legal enforcement of executory contracts was unavailable prior to the industrial revolution, how did commercial parties conduct exchange transactions? To be sure, many contracting parties relied on self-enforcement of their bargains.³⁹ But in addition, commercial parties developed alternative legal mechanisms to make credible commitments. The most important of these devices was the *penal bond*. The bond was a sealed promise to pay a sum of money subject to an express condition subsequent.⁴⁰ The bond was made conditional on an

³⁴ See *Sands v. Taylor*, 5 Johns. 395 (N.Y. 1810). Under the older common law rule, when a buyer breached a contract to purchase goods, the seller would have been required to tender the contract goods and sue for the contract price. But in *Sands*, the seller covered on the market by reselling the goods to a third party and then sought damages based upon the contract-market differential. The court conceded that this was a case of first impression in America and granted market damages to the plaintiff. *Id.*

³⁵ Horwitz, *supra* note – at 921-93.

³⁶ See e.g., *Groves v. Graves*, 1 Va. (1 Wash.) 1 (1790).

³⁷ See e.g., *Shepherd v. Hampton*, 16 U.S. (3 Wheat.) 200 (1818).

³⁸ Horwitz, *supra* note --at 941. A market damages default rule was generalized in England in 1826 with the publication of the first treatise announcing a general rule of damages for failure to deliver goods:

In an action for *assumpsit*, for not delivering goods upon a given day, the measure of damages is the difference between the contract price and that which goods of similar quality and description bore on or about the day when the goods ought to have been delivered.

Id. at 941. It was at this point that contract fully separated from property and courts, for the first time, granted promisees a property right in the contract itself.

³⁹ Grief, *supra* note – at –.

⁴⁰ Initially, the bond developed as a debt instrument to circumvent restrictions on interest under rigid usury laws. A debtor would promise under seal to pay a creditor £1,000 in six months. This promise was subject to an

executory promise to perform a specific service or to deliver goods or money at a time certain. The advantage of this formal instrument was that the promise to pay was conclusively enforceable at its face value by an action in covenant and was an absolute obligation once the defeasing condition failed to occur.⁴¹ In a real sense, therefore, the bond was a means of specifically enforcing a contractual obligation with the stated monetary obligation serving as security to ensure performance. According to Horwitz, at the beginning of the 19th century virtually all large business transactions took the form of two independent bonds, each of which stipulated damages for failure to perform an executory promise.⁴²

Penal bonds thus effectively served as the 18th century substitute for legally binding executory contracts. Indeed, at the beginning of the 19th century the number of bonds used to effect commercial transactions in America greatly exceeded the number of contracts that sought to enforce mutual promises.⁴³ The dominance of bonds, bills of exchange and other sealed instruments meant that commercial parties had little reason to take their transactional disputes to common law courts. As a result, the modern default rule of damages for executory contracts did not develop until the mid-nineteenth century.

From the beginning of the 18th century, English courts adjudicating disputes over the terms of bonds had sought to distinguish between penalties – where they would grant relief from the bond in appropriate cases⁴⁴ – and liquidated damages, which the parties were free to stipulate

express condition subsequent that it would be null and void if the debtor paid the creditor £900 (the amount of the loan) on the day following the execution of the sealed promise. THEODORE SEDGWICK, *A TREATISE ON THE MEASURE OF DAMAGES* 392-93 (1847); THEOPHILUS PARSONS, *THE LAW OF CONTRACTS* 433-34 (1855).

⁴¹ A.W.B. Simpson, *The Penal Bond with Conditional Defeasance*, 82 L.Q. Rev. 392, 411-12 (1966); SEDGWICK, *supra* note -- at 393.

⁴² Horwitz, *supra* note -- at 928.

⁴³ Horwitz, *supra* note -- at 929.

⁴⁴ W. S. HOLDSWORTH, *A HISTORY OF ENGLISH LAW* 663 (1924). During the 17th and 18th centuries, courts of equity began to grant claims for relief from the bond where there was only a minor breach of the stipulated conditions.

in the bond without interference from the courts.⁴⁵ Recall that the penalty in the bond was designed only to secure performance. It was not a sum that the promisor intended to pay. But the penalty was a poor security device in that the monetary payment was independently valuable to the promisee and this raised the risk of opportunism and induced breach.⁴⁶ In response, courts of equity policed performance under the bond to ensure that minor deviations from performance by the promisor were not used as a pretext for enforcing the face value of the bond.⁴⁷ By the end of the 18th century, common law courts had adopted the equity rule of relief from the bond where the amount owed greatly exceeded the loss to the plaintiff from the breach of the condition.⁴⁸ American courts followed this tradition and “chancered” these bonds.⁴⁹

3. *The Mutation of Market Damages to Expectancy.*

During the early decades of the 19th century, executory contracts began to dominate the bond for most transactions between commercial parties.⁵⁰ The market damages default rule that had developed in the commodities market context was adopted as well for contracts where the parties had made contract-specific investments. But plaintiffs seeking such “expectancy”

⁴⁵By the time of Lord Mansfield, English courts were predisposed to enforce liquidated damages provisions. Mansfield announced as settled the proposition that “where the covenant is to pay a particular liquidated sum, a court of equity cannot make a new covenant for a man...” *Lowe v. Peers*, 98 Eng. Rep. 160, 162 (K.B. 1768). Lord Eldon declared in 1801 that he “could not but lament any supposed principle that even an enormous and excessive damage provision, to which the parties agreed, should be void as a penalty.” *Astley v. Weldon*, 126 Eng. Rep. 1318, 1321 (C.P. 1801). Presumably, American courts at this time followed a similar presumption favoring the enforcement of liquidated damage clauses.

⁴⁶ For a discussion of the hostage function of security and the attendant risk of induced breach, see Robert E. Scott, *A Relational Theory of Secured Financing*, 86 Colum. L. Rev. 901, 927-29 (1986).

⁴⁷ 5 W.S. HOLDSWORTH, *A HISTORY OF ENGLISH LAW* 293 (1924); FARNSWORTH, *supra* note – at 812.

⁴⁸ Thompson, *Penalties and Liquidated Damages*, 46 Cent. L. J. 5,6 (1898); SEDGWICK, *supra* note –at 394.

⁴⁹SEDGWICK, *supra* note – at 394.

⁵⁰ Liquidated damages provisions in formal bonds were not well-suited to contracts premised on the allocation of market risk. See *Graham v. Bickham*, 2 Yeates 32 (Pa. 1795) (sharp market fluctuations justify a recovery for market damages in excess of the stipulation in a bond). Horwitz, *supra* note – at 932.

damages for breach of contract were generally denied recovery for consequential losses.⁵¹ Sedgwick, writing in 1847, stated: “The law does not aim at complete compensation for the injury sustained; it seeks rather to divide than to satisfy the loss. In cases of contract,... the direct pecuniary damages...form the measure of relief.”⁵² Thus, he concluded, “both the English and American courts have generally concurred in denying profits as any part of the damages to be compensated for [breach of contract].”⁵³ Justice Story, in a widely-cited opinion, held that “an allowance of damages upon the basis of a calculation of profits, is inadmissible. This would be a calculation upon conjecture and not upon facts.”⁵⁴

Thus, by the mid 1840's treatise writers could announce a general default rule governing damages for breach of contract: The breaching party was liable for losses that fairly were in contemplation of the parties at the time of contract; that is, the “plaintiff must have turned the mind of the defendant to the consequences likely to ensue from default.”⁵⁵ The early crystallization of this rule suggests that *Hadley v. Baxendale*,⁵⁶ decided in 1854, announced a

⁵¹ See cases cited in SEDGWICK, *supra* note – at 68-69 et seq., and in PARSONS, *supra* note – at 458-59. Prior to the decision in *Hadley v. Baxendale* in 1854, courts generally refused to take into consideration any damages that “remotely” resulted from the breach. Remoteness, in turn, was interpreted to mean that contracting parties could only recover that which *both parties could have contemplated at the time of contract*. SEDGWICK, *supra* note -- at 65-67.

⁵² SEDGWICK, *supra* note – at 57.

⁵³ *Id.* at 69. The remoteness principle, denying recovery for lost profits, was affirmed often by American courts. In *Blanchard v. Ely*, the New York Court of Appeals reviewed an action brought for the price of a steamboat. The buyer showed that the machinery of the vessel was defective and that those and other defects caused considerable delay in putting the boat into operation in his established ferry business. The buyer sought to deduct from the contract price not only the cost of repairs but also the lost profits for the trips that the vessel might have run during the period of the delay, having proved that the net profits lost would be \$100. The Court of Appeals disallowed the recovery of lost profits. *Blanchard v. Ely*, 21 Wendell 342 (18). In another New York case, a seller contracted to deliver a steam engine to a buyer who owned a mill for manufacturing oil. The seller failed to deliver the machine at the time for performance and the buyer sought the profits it could have made from manufacturing oil had the machine been delivered on time. The court disallowed the claim, limiting damages to the loss of full use of the mill and other machinery and interest on the additional stock purchased in anticipation of the installation of the engine. *Freeman v. Clute*, 3 Barb. S.C. R. 424 (18–).

⁵⁴ *The Schooner Lively*, 1 Gallison 314, 325.

⁵⁵ SEDGWICK, *supra* note – at 112.

⁵⁶ 9 Exch. 341, 156 Eng. Rep. 145 (1854).

rule that had already developed in America.⁵⁷ The primary effect of *Hadley* was not to limit recovery of consequential damages that previous courts would have allowed, as is generally believed, but rather to extend damage recovery by granting the plaintiff the right to recover consequential damages in those instances where he had “communicated special circumstances” to the defendant.⁵⁸

4. *The Merger of the Compensation Principle with the Penalty Rule*

During the early decades of the 19th century, common law courts found themselves with a vexing problem of interpretation. On the one hand, they had to police penal bonds. The restrictions on enforcing these bonds that had developed initially in courts of equity were subsequently enacted into statutes in many states.⁵⁹ As a result, the general rule evolved that no other sum could be recovered on a bond than that which compensated the plaintiff for his actual loss.⁶⁰ On the other hand, liquidated damages clauses were increasingly common terms in executory contracts between commercial parties. Breaching promisors argued that these damages agreements were, in fact, disguised “penalties” and should be subject to the same legal

⁵⁷In addition to formulating a damages default, the courts during this period developed more clearly the principle that the “contract itself furnishes the measure of damages.” The mid-19th century contract law thus distinguished and rejected earlier cases giving the jury wide latitude and discretionary authority to determine the measure of damages, either by reducing or enlarging the award. The amount of compensation was now regulated by the direction of the courts, and the sole object was to ascertain the agreement of the parties, which agreement controlled the measure of damages. SEDGWICK, *supra* note – at 200. Among other things, this principle was an explicit rejection of the concept of breach as “fault.” The motives behind the breach were “irrelevant.” PARSONS, *supra* note – at 443.

⁵⁸ See SCOTT & KRAUS, *supra* note – at 120-127 (discussing the conventional understanding of *Hadley*); Richard Danzig, *Hadley v. Baxendale: A Study in the Industrialization of the Law*, 4 J. Legal Stud. 249, 279-83 (1975) (same).

⁵⁹ These statutes had two principal provisions: 1) they gave defendants on bonds for the payment of money the right to pay into court the principal sum of the debt with interest and costs even though the condition was in default; and 2) they required plaintiffs suing to enforce bonds conditioned on a particular performance to assign specific breaches by the defendant and to prove the amount of loss for each breach. SEDGWICK, *supra* note – at 393-96.

⁶⁰Technically, since the bonds were sealed instruments and thus enforceable on their face, the plaintiff was awarded judgment for the face amount of the bond (since the “penalty” was the debt), but the statute only allowed execution to issue for the amount of the loss actually suffered and proved by the defendant. *Id.*

restrictions that had come to narrow the scope of the bonds.⁶¹

The courts initially seized on the “intent of the parties” as the key to distinguishing liquidated damages and penalties. An invalid penalty was a sum intended only as security for the performance of the executory promise and not intended to be paid. Enforceable liquidated damages, on the other hand, were intended to be paid by the promisor if she elected not to perform the agreement (a termination option, as it were).⁶² Complicating the analysis further was the fact that the courts also recognized that a contract to do or refrain from doing an act *or, in the alternative*, to pay a given sum of money was neither liquidated damages nor a penalty, but rather was an enforceable alternative contract (well-known to the civil law at this time).⁶³ The understandable result of these diverse holdings was a universally decried confusion.

A much applauded “solution” to the confusion of the earlier case law was proposed in *Jaquith v. Hudson* in 1858.⁶⁴ In *Jaquith*, the court upheld a liquidated damage clause enforcing a covenant not to compete in a contract for the sale of a partnership interest in a mercantile business. The court explicitly rejected the intent test of enforceability. Instead, the court held that the governing principle reconciling the cases was that damages must be based on “the principle of just compensation for the loss or injury actually sustained; considering it no greater

⁶¹SEDGWICK, *supra* note – at 398.

⁶² See cases cited in SEDGWICK, *supra* note – at 398-420.

⁶³SEDGWICK, *supra* note -- at 398. By the mid-century, the following generalizations could be advanced: First, the primary objective of courts in scrutinizing stipulated damage clauses was to determine the true intent of the parties, although the language they used was not conclusive evidence of that intent. Second, in England there was a presumption in favor of enforcing liquidated damage clauses in executory contracts, while American courts were more reluctant than their English counterparts to admit the agreement of the parties as conclusive. Third, when the agreement was in the alternative, courts would uniformly enforce the obligation to pay money. Fourth, where a court determined that the intention of the parties was to evade the statutory restrictions on penal bonds, the clause would be treated as a penalty and damages limited to actual losses. Fifth, independent of the above, if damages were uncertain and incapable of being measured except by conjecture, then any stipulated damages would be enforceable. *Id.* at 424

⁶⁴5 Mich. 123 (1858). The “intent of the parties” test that the courts had developed to distinguish between valid liquidated damages and invalid penalties simply did not square with the outcome of the cases. Many courts found an “implied intention” to create a penalty even though the parties insisted otherwise by their express language. *Jaquith* purported to reconcile these contradictions.

violation of this principle to confine the injured party to the recovery of less, than to enable him, by the aid of the court, to extort more.”⁶⁵ The compensation principle, the court declared, was a mandatory rule, a “principle of natural justice” and not a default rule, and thus the intention of the parties was irrelevant. Since the compensation principle was the law of the contract, parties were not permitted by express stipulation, however clear the intent, to set it aside.⁶⁶

But if the compensation principle is a mandatory rule, what role do the parties have in setting damages defaults? The court in *Jaquith* had a ready answer to this question. The task for the parties was to specify just compensation ex ante in those instances where they had a comparative advantage over a court seeking to do so ex post. Such comparative advantage would exist where the provable loss from the breach of the contract was uncertain, remote or speculative. Thus, while *Jaquith* purported to restrict party sovereignty over stipulated damages, the *Jaquith* rule actually gave greater latitude to 19th century contracting parties than would be true under modern damage rules. Given the then-prevailing view that lost profits could not be recovered because they were too remote or speculative, the anticipated losses in most commercial contracts would be difficult to prove. Consequently, parties had considerable freedom to stipulate damages under the *Jaquith* rule, but only to improve the accuracy of compensation and not for other purposes.

5. *From Jaquith to the Present: The Expansion of Expectation Damages*

The 150 years following *Jaquith* reveals a consistent pattern of expansion in the recovery available under default damages for breach of contract. In *Globe v. Landa*, Justice Holmes sought to generalize the common law rule of consequential damages as requiring a “tacit

⁶⁵ Id. at —.

⁶⁶ The court cited no authority for the proposition that the compensation principle of expectation damages was an immutable rule. This is particularly surprising given that this “principle” had only been generally recognized for less than thirty years. The court reverted to maxims rather than to authority, holding that courts could set aside the parties intention on the familiar ground: “conventus privatorum non potest publico juri derogare.” Id. at —.

agreement” that the breacher would be liable for losses caused by special circumstances.⁶⁷ But this limitation on consequential damages soon came under substantial pressure. One source of pressure was the increasing sophistication of expert testimony in establishing the quantum of lost profits that might result from breach.⁶⁸ Meanwhile, as commercial enterprise expanded, courts came to regard lost profit claims as a natural and direct result of a breach rather than as a special circumstance that required communication.⁶⁹ Eventually, the constraint on the recovery of lost profits was limited to the so-called “new-business rule” which denied recovery of profits only in the case of a non-established business enterprise.⁷⁰

The universal adoption of Article 2 of the UCC has completed the expansion of the expectation damages default. In addition to explicitly rejecting the tacit agreement test and relaxing the restrictions on recovery of consequential damages,⁷¹ the Code (unlike its predecessor the Uniform Sales Act) specifically invites lost profit claims by volume sellers.⁷² Even more significantly, the Code elevates the compensation principle to an overarching norm and a fundamental principle of interpretation.⁷³ Consequently, most courts under the Code have

⁶⁷ *Globe Ref. Co. v. Landa Cotton Oil Co.*, 190 U.S. 540 (1903).

⁶⁸ Since the denial of lost profits was based on claims of uncertainty, the ability to establish losses with reasonable certainty justified recovery under an expectation default.

⁶⁹ See e.g., UCC §2-715, Comment 6(2003).

⁷⁰ The new-business rule is now treated by many courts merely as a presumption in favor of the defendant rather than as an absolute bar to the recovery of lost profits for start-up businesses. See *Drews Company, Inc. v. Ledwith-Wolfe Assoc., Inc.*, 371 S. E. 2d 532 (S.C. 1988).

⁷¹ UCC §2-715(1) and Comment 2(2003); Restatement (Second) of Contracts §351, Comment a (1979)(same). Compare §67 of the Sales Act of 1906 which retained the traditional rule.

⁷² The Code abandons the common law preference for market damages, which had been enshrined in the Uniform Sales Act, and instead explicitly authorizes the recovery of lost profits against a breaching buyer so as to “eliminate the unfair and economically wasteful results arising under the older law when fixed price articles were involved.” UCC §2-708(2) and Comment 2(2003).

⁷³ UCC §1-305 (2003) (“The remedies provided in this Act shall be liberally administered to the end that the aggrieved party may be put in as good a position as if the other party had fully performed....”). See Scott, *The Case for Market Damages*, *supra* note— at 1169. Section 1-305 (formerly §1-106) seemingly trumps the arguments for measuring recovery other than by ex post expectancy.

concluded that, even in thick market contexts, market damages should not be used where they depart from the economic gain the promisee would have enjoyed had the contract been performed.⁷⁴ And, as this default notion of compensation has expanded, the penalty rule has become a tighter constraint on the ability of parties to stipulate damages.⁷⁵

The dominance of the compensation principle is now unquestioned, but the preceding inquiry into its historical roots suggests that the elevation of compensation to a universally applicable norm results more from path dependence than from a sustained and systematic appreciation of the merits of the rules governing contract damages. In the following section, we ask whether the compensation principle is nevertheless theoretically justifiable on economic grounds.

C. Economic Contract Theory and Compensation

Given that compensation is now entrenched doctrinally as the motivating principle behind default damages rules and the regulation of liquidated damages, it is perhaps not surprising that scholars have advanced economic justifications for the rules in each case. However, the literature in the economic analysis of contracts reflects a growing academic consensus against the penalty rule and raises substantial doubts about the desirability of expectation damages. Indeed, the economics scholarship of the last decade scarcely mentions compensation either as an end or a means to maximizing the joint welfare of contracting parties.

⁷⁴See cases cited in note 160 *infra*.

⁷⁵ As noted above, the alternative contract doctrine has also been harnessed to the compensation principle in recent years. See Part I(A) *infra* and cases cited in note 23 *supra*. This development is reflected in the differing treatment of alternative contracts in the first and second Restatements. The first Restatement affirmed the enforceability of alternative contracts and provided for damages in the case of a breach without an election in accordance with the alternative that resulted in the smallest recovery (presumably without regard to whether that recovery was supercompensatory). Restatement of Contracts §344 (1932). The second Restatement eliminates any specific default rule for damages for breach of an alternative contract. Instead, the comments to the section on liquidated damages and penalties state that “although the parties may in good faith contract for alternative performances and fix discounts or valuations, a court will look to the substance of the agreement to determine whether this is the case or whether the parties have attempted to disguise a provision for an [unenforceable] penalty...” Restatement (Second) of Contracts §356, comment c (1979).

1. Expectation Damages and Efficient Breach

The principal economic justification for expectation damages is that they compel the promisor to internalize the costs that her breach inflicts on the promisee. The promisor consequently has the incentive to make the efficient breach decision.⁷⁶ By internalizing the promisee's loss, the promisor also has the incentive to take the efficient precautions against contingencies that threaten to increase the cost of performance.⁷⁷ In addition, expectation damages provide insurance to the promisee against the risk of increases in the promisor's cost of performance that might lead her to breach. This insurance may be efficient if the promisor is risk neutral and the promisee is risk averse.⁷⁸

Notwithstanding the possible benefits of the current rule, more recent literature has identified a large number of countervailing considerations against expectation damages. Our purpose is not to provide a comprehensive review of these factors, but simply to illustrate that they present a very significant challenge to the case for expectation damages and indeed compensatory damages in general. Expectation damages ensure efficient breach only if a court can accurately and predictably measure the promisee's losses at no cost. Yet, the practical difficulty in verifying the promisee's losses makes judicial assessment of expectation damages uncertain and costly, and this may lead to costly and inaccurate breach. For example, even if the

⁷⁶ See e.g., Goetz & Scott, *supra* note – at 558-59:

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 receive 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 the breacher. 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.

⁷⁷ *Id.* at 579-83.

⁷⁸ E.g., A. Mitchell Polinsky, *Risk Sharing Through Breach of Contract Remedies*, 12 J. Legal Stud. 427 (1983); Samuel A. Rea, Jr., *Efficiency Implications of Penalties and Liquidated Damages*, 13 J. Legal Stud. 147, 151-63 (1984). To many, the insurance objective seems inapt in contracts for goods or services. Several decades before Holmes wrote his contract options statement (*supra* note –), Pollock wrote: “A man who bespeaks a coat of his tailor will scarcely be persuaded that he is only betting with the tailor that such a coat will not be made and delivered to him within a certain time. What he wants and means to have is the coat, not an insurance against not having the coat.” F. POLLOCK, *PRINCIPLES OF CONTRACT* xix (1881).

courts are accurate on average, a risk averse promisor may choose an inefficient performance (i.e. where the cost exceeds the promisee's anticipated benefit) out of a desire to avoid the uncertainty in her prospective damages liability. Moreover, the expected enforcement costs may outweigh the incentive and insurance gains from expectation damages, so that the parties may be better off with undercompensatory damages that reduce the likelihood of litigation.⁷⁹ At least in part, the common law doctrines of foreseeability and uncertainty may be explained as striking just such a compromise between the goals of minimizing enforcement costs and inducing efficient breach..

2. *The Excessive Reliance Problem.*

More prominent in economic contract theory is the observation that expectation damages give rise to well known moral hazard incentives for the promisee to overrely on promises and to fail to take steps to avoid losses from breach.⁸⁰ Several common law doctrines cut back on compensation in order to address this problem. In particular, the mitigation doctrine deprives the promisee of compensation for post-breach losses that she reasonably could have avoided.⁸¹

⁷⁹If damages are undercompensatory, the stakes in litigation are lower and the parties will therefore spend fewer resources in litigation. Within some range, the resulting saving in litigation cost may outweigh the incremental loss in the incentive and insurance gains caused by agreeing to lower-than-compensatory damages. The tradeoff is discussed more generally in A. Mitchell Polinsky and Daniel Rubinfeld, *The Welfare Implications of Costly Litigation for the Level of Liability*, 17 J. Legal Stud. 151 (1988).

⁸⁰This is the well-known overinvestment problem that is created by reliance or expectation damages (as well as specific performance). See Steven Shavell, *Damage Measures for Breach of Contract*, 11 Bell J. Econ. 466 (1980); William P. Rogerson, *Efficient Reliance and Damage Measures for Breach of Contract*, 15 Rand J. Econ. 39 (1984). Assume, for example that the seller contracts to produce a specialized good for the buyer and the buyer makes a specific investment (one that has no value other than with the good) in order to increase its valuation of the good. The buyer's expectation damages if the seller breaches will be the difference between its realized valuation and the price. The buyer will always receive the return on his investment, even if trade does not occur and the investment is effectively wasted. The buyer's marginal return from investment exceeds the social return, leading to inefficiently high investment.

⁸¹ The doctrine of avoidable consequences has both negative and affirmative aspects. The affirmative branch of the doctrine permits recovery of all reasonable expenses that the plaintiff incurs in seeking to avoid damages. See e.g., *Rench v. Hayes Equip. Mfg. Co.*, 134 Kan. 865, 8 P.2d 346 (1932). The negative branch of the doctrine precludes a plaintiff from passively incurring losses which she could reasonably avoid or from actively increasing such losses where prudence would require an adjustment. See, e.g., Restatement (Second) of Contracts §350(1979); *Rockingham County v. Lutten Bridge Co.*, 35 F.2d 301 (1929). For discussion, see Charles J. Goetz & Robert E. Scott, *The Mitigation Principle: Toward a General Theory of Contractual Obligation*, 69 Va. L. Rev. 967, 973-76 (1983).

The foreseeability and uncertainty doctrines remove unusual losses from the measure of damages and thereby dampen the incentive to rely more than the average promisee.⁸² Moreover, the fact that plaintiffs tend not to be compensated for waiting for payment until after trial and that they do not recover their legal costs restores some of the promisees' incentives to protect themselves from breach. Nevertheless, the problem of excessive reliance and insufficient precautions by the promisee remains whenever the damages are compensatory in the sense that they vary at the margin with the losses of the promisee. To correct for overinvestment, a contract might fix liquidated damages at the expected compensatory level, so that damages do not vary with the promisee's actual expectancy loss.⁸³ Although this may result in optimal precautions, it will not produce efficient "trade or breach" decisions ex post because the realized damages will be either over- or undercompensatory at the time set for performance.⁸⁴

3. *Shift in Theoretical Focus from Efficient Breach to Efficient Investment.*

These analyses of contract remedies thus suggest that any potential benefits of expectation damages in promoting efficient ex post decisions to trade or breach may be offset by the failure of the damage rule to motivate efficient ex ante investment and reliance.⁸⁵ One problem is the aforementioned tendency of the promisee to overrely. In addition, when the promisee is insured against breach by an expectation damages award, she lacks the incentive to

⁸² In general, speculative losses cannot be recouped. These include lost good will, reputation or emotional distress. See SCOTT & KRAUS, *CONTRACT LAW & THEORY* at 1050 et seq. The foreseeability limitation has been considerably eroded over time. Under UCC §2-718, for example, the plaintiff can recover "all losses resulting from the general or particular requirements and needs of which the seller at the time of contracting had reason to know...." See also, Restatement (Second) of Contracts §351(1979).

⁸³ Robert Cooter & Melvin Aron Eisenberg, *Damages for Breach of Contract*, 73 Calif. L. Rev. 1434 (1985).

⁸⁴ If the variance in realized damages is a continuous function, any stipulated sum that is set ex ante to reflect expected damages ex post will always turn out to be wrong at the time of breach.

⁸⁵ Aaron Edlin and Stefan Reichelstein resuscitate expectation damages in their model to show that they can lead to efficient investment where only the promisee's investment is relevant and where the parties can renegotiate their agreement. Aaron S. Edlin & Stefan Reichelstein, *Holdups, Standard Breach Remedies, and Optimal Investment*, 86 Am. Econ. Rev. 478 (1996).

incur costs that help or induce the promisor to avoid breach.⁸⁶ In recent years, therefore, contract theorists have focused attention on designing contracts that optimize *both* the ex post decision whether to trade or breach and the ex ante incentives to make specific investments. The challenge comes from the fact that the conditions for efficient trade and efficient investment cannot be specified fully in the contract in light of the transaction costs of foreseeing and describing all possible future states of the world, and of verifying the realized state to a court.

Yet, even when information is not verifiable, contract theorists have shown that ex post efficiency is easy to achieve without expectation damages through a variety of “implementation mechanisms,” as long as the parties themselves are symmetrically informed.⁸⁷ These mechanisms elicit from the parties their information about the state of the world that can then be used to identify efficient payoffs for that state to the court.⁸⁸ The availability of these solutions diminishes the significance of compensatory damages in forcing promisors to internalize the losses caused by their breach.

The contribution of expectation damages in promoting efficient decisions to trade is even less valuable if contracting parties are able to renegotiate their agreement. The premise that parties can often renegotiate to efficient ex post outcomes is sufficiently well accepted that

⁸⁶Yeon-Koo Che & Donald Hausch, *Cooperative Investments and the Value of Contracting*, 89 Am. Econ. Rev. 125 (1999).

⁸⁷Jean Tirole, *Incomplete Contracts: Where Do We Stand?*, 67 Econometrica 741 (1999); Benjamin Hermalin and Michael Katz, *Judicial Modification of Contracts Between Sophisticated Parties: A More Complete View of Incomplete Contracts and Their Breach*, 9 J. Law, Econ. & Org. 98 (1993). Under some conditions, an option contract might reveal even asymmetrically held private information. Georg Noldeke and Klaus M. Schmidt, *Option Contracts and Renegotiation: A Solution to the Hold-up Problem*, 26 Rand J. Econ. 163 (1995)

⁸⁸Mechanism contracts induce parties to reveal the ex post state truthfully to a court or other decisionmaker. As an example, the contract may provide that the buyer must write down how much she values the seller’s performance and the seller must write her cost of performance. If the buyer’s written valuation exceeds the seller’s written cost, then the court will order the parties to trade; otherwise the court will prohibit trade. It is a dominant strategy under such a mechanism for each party to report the ex post state truthfully. The inability of the parties to renegotiate the consequences of a mismatch is crucial to the revelation mechanism. Some authors have noted that option contracts are message-contingent in this way. See e.g., Ilya Segal, *Complexity and Renegotiation: A Foundation for Incomplete Contracts*, 66 Rev. Econ. Stud. 57, 57 n1 (1999); For discussion, see Robert E. Scott, *Rethinking the Default Rule Project*, 6 Va. J. 84 (2004).

contract theorists have largely set aside the concern with efficient breach to focus on the hold up problems caused by renegotiation.⁸⁹ In particular, the division of the renegotiation surplus will not reflect the fact that one or both of the parties have made specific investments contributing to the surplus, and therefore the investing party will be compelled to share the payoffs from her investment. In turn, the anticipation of a renegotiation of contractual commitments undermines attempts to establish efficient investment incentives.⁹⁰ Given that states of the world and performance obligations cannot be fully described in the initial contract, the courts often cannot discern when a party forcing renegotiation is acting strategically by refusing to carry out an efficient trade. One set of solutions to this problem proposes that the parties create legal or practical obstacles to renegotiation, such as bargaining through organizations whose internal regulations require many layers of consent.⁹¹ A more popular theoretical approach is to try to predetermine the allocation of the renegotiation surplus by effectively assigning bargaining power to one party or the other.⁹²

⁸⁹Oliver Hart and John Moore, *Incomplete Contracts and Renegotiation*, 56 *Econometrica* 755 (1988).

⁹⁰The key to the hold-up threat is the presence of specific investment, whose value in the exchange is shared between the two parties in the course of the renegotiation. Oliver E. Williamson, *Transaction Cost Economics: The Governance of Contractual Relations*, 22 *J. Law & Econ.* 233 (1979); Benjamin Klein, Robert G. Crawford and Armen A. Alchian, *Vertical Integration, Appropriate Rents, and the Competitive Contracting Process*, 21 *J. L. & Econ.* 297 (1978).

⁹¹See e.g., Alan Schwartz & Joel Watson, *Economic and Legal Aspects of Costly Contracting*, 20 *J. L. Econ. & Org.* #2 (2004); Tirole, *Incomplete Contracts*, supra note —. A precommitment not to renegotiate the contract generally cannot be made simply by contract. Under current law, an agreement not to modify a contract is not enforceable. Restatement (Second) of Contracts §311 comment. a (1979); *Beatty v. Guggenheim Exploration Co.*, 122 N.E. 2d 378, 387 (N.Y. 1919) (“Those who make a contract can unmake it. The clause that forbids a change can be changed like any other ... Whenever two men contract, no limitation self-imposed can destroy their power to contract again.” Cardozo, J); See also, *Zumwinkel v. Legget*, 345 S.W. 2d 89 (Mo. 1961); Christine Jolls, *Contracts as Bilateral Commitments: A New Perspective on Contract Modification*, 26 *J. Leg. Stud.* 203 (1997).

⁹²E.g. Tai-Yeong Chung, *Incomplete Contracts, Specific Investments, and Risk Sharing*, 58 *Rev. Econ. Stud.* 1031 (1991); Philippe Aghion, Mathias Dewatripont & Patrick Rey, *Renegotiation Design with Unverifiable Information*, 62 *Econometrica* 257 (1994); Georg Noldeke and Klaus M. Schmidt, *Option Contracts and Renegotiation: A Solution to the Hold-up Problem*, 26 *Rand J. Econ.* 163 (1995). The idea in these two articles is to set the correct investment incentives for one party through the setting of the default point in renegotiation and for the other party by assigning her bargaining power over the division of the renegotiation surplus. Property ownership rights may also be used to determine the parties’ bargaining position in ex post renegotiation. Sanford J. Grossman and Oliver D. Hart, *The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration*, 94 *J. Pol. Econ.* 691 (1986); Oliver Hart and John Moore, *Property “Rights and the Nature of the Firm*, 98 *J. Pol. Econ.* 1119 (1990). Yet another approach is to permit bargaining power in renegotiation to be exogenously determined, but to set the contract quantity so as to balance the likelihood of renegotiation to a lower or higher value, in such a way as to

In sum, contemporary economic contract analysis indicates that expectation damages a) do not stimulate efficient ex ante investment, and b) are neither sufficient nor necessary to achieve efficient ex post trade.⁹³ Parties have the choice between more complete contracts that yield both efficient trade and efficient investment but can only succeed if there is no prospect of renegotiation, or simple contracts which the parties may renegotiate to achieve efficient trade without compromising investment efficiency.⁹⁴ This theoretical literature also raises doubts about the doctrinal constraints on liquidated damages by showing that penalties may serve to improve the efficiency of specific investment.⁹⁵ In the discussion that follows, we attack the penalty rule from a different perspective by demonstrating that it impedes the sale of insurance from promisees to promisors.

D. Commercial and Consumer Practices of Termination

The irrelevance of compensation in the economic theory of contracts is reflected in the patterns of commercial and consumer contracts in practice. These contracts reveal systematic attempts to contract away from the compensation principle, with varying degrees of success.

also cancel out under- and overinvestment tendencies. Aaron S. Edlin and Stefan Reichelstein, *Holdups, Standard Breach Remedies, and Optimal Investment*, 86 Am. Econ. Rev. 478 (1996).

⁹³But see Eric Posner, *Economic Analysis of Contract Law After Three Decades: Success or Failure?* 112 Yale L.J. 829 (2003):

“[It is possible that] on average pre-performance investment is not a significant issue, or, if it is, it is adequately controlled by the doctrine of mitigation. The rule of expectation damages is optimal because the perform-or-breach decision matters most, with specific performance reserved for cases where valuation problems are insurmountable. But this view is unsupported by any evidence.” (At 838-9)

⁹⁴See, e.g., Tirole, *supra* note —; Schwartz & Watson, *supra* note —.

⁹⁵An excellent review of the arguments for and against the penalty rule is Aaron S. Edlin & Alan Schwartz, *Optimal Penalties in Contracts*, 78 Chi.-Kent L. Rev. 33 (2003). Proponents of the penalty rule suggest that it prevents contracting parties from agreeing to inefficiently high damages in order to extract a larger portion of the surplus of a new entrant who arrives during the contract term. E.g., Tai-Yeong Chung, *On the Social Optimality of Liquidated Damage Clauses; An Economic Analysis*, 8 J. L. Econ. & Org. 280 (1992); Philippe Aghion & Patrick Bolton, *Contracts as a Barrier to Entry*, 77 Am. Econ. Rev. 388 (1987). The possibility that, when the seller has market power, the parties may use penalties to deter the entry of competitors into the seller’s market does not, by itself, justify the current penalty rule. Under current law, courts strike what they perceive to be penalty terms whether or not those terms were used to impede entry. A party should always be free to argue that any term would create a negative externality or perpetuate a market failure. But it is a mistake to treat as a sufficient proxy for these inefficiencies a liquidated damage clause that would overcompensate the promisee in expectation.

Remedial provisions can often be reframed as substantive terms.⁹⁶ The parties may contract for alternative methods of contract performance rather than providing for a primary obligation to perform and a secondary obligation to pay damages. Alternatively, one party may grant the other an explicit option⁹⁷ or the right to cancel upon payment of a fee or loss of a deposit.⁹⁸ A requirements or output contract permits one side to avoid the obligation to purchase or deliver larger quantities of goods.⁹⁹ In view of these substantive alternatives, breach damages are best seen as a subset of the larger category of termination provisions. These terms share the common feature that one party has the option to terminate the contemplated exchange by paying some amount or incurring some cost. That cost is unrelated to compensation. In practice, this termination fee is often undercompensatory (consider, for example, the number of buyers who

⁹⁶Restatement (Second) of Contracts §368 states that specific performance or an injunction will not be granted against a party who can substantially nullify the effect of the order by exercising a power of termination or avoidance. In contrast, §361 provides that the court can order an injunctive remedy in the face of a liquidated damages term.

⁹⁷For example, a unilateral promise invites the promisee to accept by rendering a performance rather than by a promissory acceptance. The promise creates an option contract once the promisee begins the invited performance. Restatement (Second) of Contracts § 45(1979). See generally Katz, *supra* note —.

⁹⁸Many such options are categorized doctrinally as alternative contracts. Traditional analysis has distinguished the alternative provision designed to secure performance of the primary promise (a liquidated damages clause) from two promised alternatives between which the promisor can choose, each an agreed exchange for the consideration given by the promisee (an embedded option in our terminology). *Garrett v. Coast & S. Fed. Sav. & Loan Ass'n*, 9 Cal. 3d 731, 738, 511 P.2d 1197, 1201 (1973); Restatement of Contracts §339 Comment f (1932). The former are at risk of being found unenforceable as a penalty, while the latter would be an enforceable alternative contract. The first Restatement of Contracts, in classic understatement, acknowledge that enforceable alternative contracts may easily be confused with invalid liquidated damage provisions. §334 Comment c (1932). Modern courts have been inclined to bring compensation to bear in the decision to enforce even alternative promises. *Supra* note —.

⁹⁹UCC § 2-306(1)(2003). Discretion under an output or requirements contracts is effectively constrained only by the duty of good faith. A number of courts have recognized that under demand in the case of a requirements buyer (or under supply for an output seller) need only be justified by a “valid business reason” to pass the test of good faith imposed by UCC §3-601(1). See e.g., *Empire Gas Corp. v. American Bakeries Co.*, 840 F.2d 1333 (7th Cir. 1988). Thus construed, a requirements contract can be characterized as an unconstrained option to take as small a quantity of the good as the buyer chooses. The buyer’s discretion is regulated in the over demand situation to prevent the buyer from opportunistically exploiting the contract price and reselling the goods in a favorable market. See *Id.*

Many other contracts grant one of the parties discretion to determine a significant exchange term. For example, the seller under an installment contract has discretion over the content of any installment. Breach does not occur unless nonperformance substantially impairs the value of the whole contract. UCC § 2-612(2003). UCC default rules also grant the buyer an option relating to assortment of goods and the seller an option relating to shipment. UCC § 2-311(2003).

can return goods for free with no questions asked). But sometimes buyers agree to make payments that are likely to be overcompensatory (both ex post and ex ante). Consumers, for example, pay cancellation fees for walking away from airline tickets or concert performances even where their seats are resold.

The success of these techniques in avoiding the regulation of liquidated damages is far from certain because of the broad reach and prominence of the compensation principle. Although doctrinal statements provide that underliquidation is subject to the same constraints as overliquidation,¹⁰⁰ courts have generally enforced undercompensatory damage provisions and termination options.¹⁰¹ As noted in Section I.B, the penalty rule is a more formidable obstacle.¹⁰² The linkage of a quasi-mandatory compensation principle with both the penalty rule and the expansive default of full expectation damages may have resulted from the anomalies of the penal bond and the early common law forms of action. But once entrenched, these rules have proved

¹⁰⁰See UCC §2-718, Comment 1(2003).

¹⁰¹Supra note 20. Thus, for example, commercial sellers have designed various limitation of remedy provisions such as the ubiquitous “repair and replacement” clauses. Richard Epstein observes that the warranty provisions in many sales contracts limit the buyer’s remedies for breach either to liquidated damages that are undercompensatory or to the repair or replacement of the damaged goods at the option of the seller. Richard A. Epstein, *Beyond Foreseeability: Consequential Damages in the Law of Contract*, 18 J. Legal Stud. 105, 114-21 (1989). Epstein also describes the standard limitation on damages in delivery services such as Federal Express, which undercompensate even the average consumer. Consistent with the analysis summarized above, Epstein concludes that, “[a]gainst the backdrop of express contractual provisions, there is ample reason to doubt that the expectation measure of damage of the classical common law maximizes the joint gains of the parties ex ante. If it did, we should expect to observe it frequently in practice, which is decidedly not the case.” Id.

¹⁰²Two of our students undertook a search of all federal and state appellate cases decided between January 1, 1988 and January 1, 2004 that invoke liquidated damages and the penalty rule. The search returned 333 recorded decisions where enforceability was at issue. A random sample of 109 of these cases were examined in detail. In 38% of the sample cases, the courts found that the stipulated damages provision in question was void as a penalty. Jason K. Binder & Michael Labriola, *An Empirical Analysis of Liquidated Damages Clause Jurisprudence in the United States From 1988 to 2003* (mimeo 2004). It is important to emphasize the significance of a roughly 40% rate with which contemporary courts refuse to enforce liquidated damages clauses. Assuming that none of the litigated contracts were unconscionable (there were no such holdings in the sample), then under accepted contract principles the parties should have been free to stipulate damages as they saw fit. Thus, the appropriate baseline from which to assess these findings is to assume that, but for the penalty rule, 100% of the contested clauses would be upheld as enforceable. Moreover, since lawyers drafting liquidated damages clauses are likely cognizant of the penalty rule, one would expect them to attempt to draft these provisions so as to pass the legal test for enforceability. Thus, the fact that nearly 40% of the litigated clauses fail the test is strong evidence of the continued vitality of the penalty rule.

remarkably durable despite the claims of scholars that they preclude (or at least constrain) parties from writing efficient contracts. Non-refundable deposits, prepayments of contract price and stipulated damages in licensing contracts are vulnerable to being struck down under the penalty rule.¹⁰³ Moreover, as we noted earlier, many courts scrutinize even alternative contracts such as take-or-pay provisions to ensure they have some link to the compensation of the seller.¹⁰⁴

II. EMBEDDED OPTIONS

The preceding discussion underscores the importance of improving our understanding of how contracting parties make their choices among alternative termination provisions and, specifically, how they select a termination fee. In Part II, we characterize termination provisions as options embedded in contracts and the termination fee as effectively the price paid for the option enjoyed by the promisor. We focus on the termination right held by a buyer of a good. In the case of termination by breach, this option price is the amount of damages owed by the breaching buyer. We then examine the factors that determine the option price to which the parties would agree. Although the protection of specific investment remains an important contract objective, our analysis shifts the focus to insurance or risk management. We conclude that contracts set option prices equal to either the ex ante or ex post loss from termination only where neither risk management nor promisee specific investment are important contracting objectives, thereby raising serious doubts about the compensation principle in the measure of contract damages.

¹⁰³ See e.g., *MCA Television Ltd. V. Public Interest Corp.*, 171 F3d 1265 (1999) (clause awarding the non-breaching party full payment of the entire contract price plus the right to revoke the breacher's license to broadcast held invalid as a penalty); *In re Admetric Biochem, Inc.*, 284 Bkr. R. 1 (2002) (deposit plus acceleration clause in commercial lease held void as a penalty where it constituted 20% of the total rent reserved under the lease); *Shree Ganesh, Inc. v. Days Inns Worldwide, Inc.*, 192 F. Supp. 2d 774 (N.D. Oh 2002) (liquidated damages clause in franchise agreement held void as a penalty).

¹⁰⁴ See note 22 supra and accompanying text. For discussion of take-or-pay clauses, see Goldberg, *Discretion in Long Term Quantity Contracts*, supra note 1; Goldberg, *The Net Profits Puzzle*, supra note 1; Victor P. Goldberg, *Bloomer Girl Revisited or How to Frame an Unmade Picture*, 1998 Wisc. L. Rev. 1051.

A. Termination Rights and Call Options

For expositional purposes, we focus on the termination right of a buyer of a good and assume that the seller's obligation is specifically enforceable. Perhaps the most intuitive way to think about termination is to view the buyer as purchasing the good for the price P and holding a *put* option on the good with exercise price, x .¹⁰⁵ In this paper, however, we use the equivalent characterization in which the buyer pays an *option price*, d , to acquire a *call* option to purchase the good with an *exercise price* of x . The sum of the option price and the exercise price is the contract price, $P = d + x$. For example, if the buyer's promise is enforced by damages, then d is the damages liability and x is the difference between the contract price and those damages.

Consider, for example, a contract under which the buyer purchases a widget for \$15 and can recover \$11 if she returns it within 30 days. For now, set aside the question of which party has possession of the widget during the 30-day period (by assuming the use-value is nominal) and the question of who holds the \$15 during that period (by assuming that the market discount rate is zero). Under the put-call parity rule, the buyer's position can either be characterized as (a) the combination of the widget together with a put option with exercise price \$11 or (b) a call option on the widget with exercise price \$11.¹⁰⁶ We prefer the call option characterization because it is more parsimonious and analytically revealing than the contract-plus-put combination. We can describe the call option by the pair (d, x) that isolates the price of the option in d : in the example above, the price of the call with exercise price \$11 is \$4. If we used the alternative characterization, the price of the put would be embedded in the contract price (the contract price of \$15 in the example above reflects the value of the put together with the value of

¹⁰⁵The payment of the termination fee or damages by the buyer or seller, respectively, is subject to a further option if there is a risk of insolvency. Each party can obtain a discharge from the obligation by surrendering their assets in a bankruptcy proceeding. Thus, the option enjoyed by a contracting party is essentially a compound option because it includes both the termination option in the contract and the bankruptcy option held by all debtors. Triantis, *Effects of Insolvency and Bankruptcy*, supra note – at 682-7. For the purposes of this article, however, we assume that both parties cannot avoid their debts by bankruptcy or dissolution.

¹⁰⁶Specifically, a call option is equivalent to the combination of holding the underlying asset, borrowing the exercise price and a put option on the same asset. E.g., Hans R. Stoll, *The Relationship between Put and Call Option Prices*, 24 J. Fin. 801 (1969). We assume a zero discount rate.

the widget).

The parties have the choice among a set of alternative option contracts, (d, x) . The values of d and x will tend to be inversely related, and the contract price will be higher the easier it is for the buyer to walk away from the contract.¹⁰⁷ Consider first a contract in which the buyer agrees to buy a widget for \$12 with no termination right, $(12, 0)$. Now, suppose that the seller offers the buyer an alternative contract under which the buyer makes a deposit, d , and holds an option to purchase the widget for \$1. The buyer pays d to acquire a call option with exercise price \$1. The deposit, d , is effectively the price the buyer pays for the option. If there is some possibility that the buyer will value the widget for less than \$1, she will be prepared to pay a deposit -- effectively an option price -- greater than \$11. Writing this option may also be costly for the seller,¹⁰⁸ but the parties will choose to trade the option if this cost is less than the value of the option to the buyer. If the seller has some bargaining power, she will be able to capture some of the surplus. It is therefore likely that the aggregate contract price will be greater than \$12 when the seller writes a call option with an exercise price of \$1. As a general matter, the sensitivity of the option price to changes in the exercise price is such that, for each dollar increase in the exercise price, the price of the option decreases by less than a dollar. Therefore, the aggregate contract price rises asymptotically with the exercise price as the parties approach $(0, P^*)$, where P^* is the maximum value for the contract price.¹⁰⁹

¹⁰⁷Some of the economic models of contracting analyze a contract that provides for two prices to be paid by the buyer. The buyer pays p_1 if there is trade and p_0 if there is not, where p_0 might be negative. The critical feature is the difference, $p_1 - p_0$, and the parties are otherwise free to set p_0 where they wish. See, e.g., Noldeke and Schmidt, *supra* note —; Hart and Moore, *supra* note —. “[O]nly the difference $p_1 - p_0$ matters for efficiency. Therefore, we have one degree of freedom in choosing p_0 which can be used to share the expected surplus ex ante between the parties.” Klaus M. Schmidt, *Contract Renegotiation and Option Contracts*, 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND LAW 433 (Peter Newman, ed. 1998). In the options analysis of property-versus-liability rules, Ian Ayres and Paul Goldbart analyze a similar choice among combinations of payments that might be made between plaintiffs and defendants in, for example, nuisance suits. By “convexifying” the available combinations, they enable the court to choose among distributional results without affecting allocative efficiency. Ian Ayres and Paul M. Goldbart, *An Option Theory of Liability Rules*, 100 Mich. L. Rev. 1, 28-34 (2001).

¹⁰⁸If, as we assume, the seller has no termination rights, the seller’s cost includes the fact that the seller cannot entertain better offers from other buyers.

¹⁰⁹In nominal (dollar) terms, the option price is most sensitive to changes in exercise price when the option is near or at the money (exercise price equal to the buyer’s valuation).

If we begin our analysis instead at the $(0, P^*)$ end of the spectrum, suppose that $P^* = x = \$18$. The seller effectively has given the buyer a free call option with an exercise price of \$18. We later discuss the possible motivations for giving such a free option.¹¹⁰ For now, note that if the seller offers to reduce the exercise price from \$18 to \$17, the buyer will be willing to pay an option price in the form of a deposit -- but in an amount less than \$1. Similarly, if the seller offers to reduce the exercise price further, the buyer will be prepared to increase her deposit, but by less than the reduction in the exercise price. Of course, the seller does not seek to maximize the contract price per se, but rather to maximize the difference between the option price and the cost of writing the option for the buyer.

The ways in which the price of the call option, d , is paid are as diverse as the forms of termination fees. The buyer may make a nonrefundable deposit, agree to a cancellation fee, or assume a commitment to pay damages in the event of breach. The common feature is that the buyer pays for the call option by incurring an initial cost specified by contract and can subsequently choose to incur an additional cost to execute the contract exchange.¹¹¹ If the option price is in the form of a commitment to pay damages upon breach, then the option price depends on the materialized loss from breach suffered by the seller. Consequently, the buyer faces a distribution of option prices during the term of the option. This uncertainty does not change the analysis. The option to breach and pay damages conventionally may be termed an exchange option: the right to exchange one asset of uncertain value (the damages liability) for another (the payoff from a completed exchange). For our expositional purposes, however, we will stick to the call option characterization whether the option price is fixed at the time of contract or uncertain.

Indeed, there may be other sources of uncertainty in the option price. The enforcement of

¹¹⁰These free options are common and can be explained either as cases where the seller incurs a trivial cost in writing the option (e.g., a reservation at a restaurant with a heavy street clientele) or where the seller uses the free option as a marketing tool (e.g., in the place of a free gift to potential customers). *Infra* note 146 and accompanying text.

¹¹¹The seller's option to terminate can be similarly presented. If the seller has a termination right, then she effectively incurs a liability for the termination fee and can pay the additional cost of delivering the good to the buyer in order to receive the contract price. The case of bilateral buyer's and seller's options is substantially more complex. See Triantis & Triantis, *Timing Problems*, *supra* note -- at 184-94.

breach damages may be less than perfect or the duration of the option may be uncertain. For example, many consumers purchase retail items where the return policy is unclear; retailers may be deliberately vague. The consumers ability to walk away from the deal may need to be discounted, therefore, according to the probability that the retailer will permit a refund of different amounts, and according to the distribution of negotiated or judicial outcomes in the event of any dispute. Practically, these factors are significant in many circumstances. We suggest below that the buyer may be willing to pay a premium to avoid these risks. This premium may be in the form of liquidated damages that will be struck down by a court – inappropriately, in our view – as supercompensatory and a penalty.

In the discussion that follows, we identify two sources of value of embedded call options: (a) preserving the real options given by nature (without the need for renegotiation between the parties) and (b) insuring the buyer against the risk of fluctuations in her valuation of the seller's performance. At the end of this Part, we invoke some basic insurance analysis to identify factors that determine the optimum (d,x) pairing. Consistent with conventional efficient breach theory, real options analysis suggests that expectation damages lead to the efficient exercise of real options. As we noted earlier, however, the embellishments to contract theory over the past twenty-five years have emphasized the significance of specific investments and the ability of parties to renegotiate their contracts. As a result, it is now clear that expectation damages are neither necessary nor sufficient for efficiency. The analysis that follows introduces important and widely applicable insurance considerations indicating that optimal termination provisions are highly context-contingent and are determined by the factors of risk management and market conditions, rather than the compensation principles of contract law.

B. Real Options and Ex Post Efficiency

In the next two sections, we distinguish between real options and written contract options. The exercise of real options yield value to the holder but impose costs on no one. In other words, no party *writes* real options; they are the product of nature. For example, a firm may have flexibility in choosing when to exploit a given opportunity. It can delay its investment in the project and hold the option to invest at a later date. Once it begins investing, the firm then holds options to accelerate, decelerate or even abandon the project. Consider the ability to delay the investment, effectively a call

option on the project. This option is valuable because the firm has the opportunity to take action to limit its losses as it receives new information about project payoffs. When a venture is carried out by contract rather than in an integrated firm, the parties may agree to termination rights in order to preserve the firm's valuable real options. The termination rights assign the exercise decision to one or both of the parties, depending on their comparative access to new information. At the same time, the contract should ensure that the holder of the option has the appropriate incentives to make efficient exercise decisions. Consistent with conventional efficient breach theory, expectation damages promote efficient option exercise by compelling the decision maker to internalize the costs borne by the other party. Thus, if only the optimal exercise of real options were relevant, the parties would be inclined to agree to expectation damages and to limit the parties' respective termination rights accordingly.

To illustrate this point, consider a firm whose research and development has yielded the design of a novel product for sale to consumers. To exploit this opportunity, the firm must build a machine to manufacture the product at a cost, c . The value of the machine, $\tilde{v}$, is uncertain because it varies with the likely demand for its product. If the firm is compelled to make the decision to build the machine immediately (at t_0), the firm would proceed if and only if the net expected value of the project were positive: $E_0(\tilde{v}) > c$.¹¹² However, the firm might be able to make some preliminary expenditure, s , that would enable it to delay its decision to pursue the project to the next period, t_1 . In other words, it can pay to preserve a call option on the project. During the term of the call, the firm might acquire information about $\tilde{v}$. For ease of exposition, suppose that all the uncertainty is resolved at t_1 . If the new information indicates that $v < c$, the firm would let the option expire; if it revealed that $v > c$, the firm would exercise the call by investing in the project. Thus, the call option allows the firm to retain the upside prospect while jettisoning the downside risk of loss. The option is valuable at t_0 , and the firm may pay s to preserve it, even if the project is on average unprofitable; that is, even if $E_0(\tilde{v}) < c$.¹¹³ It is important to appreciate that this flexibility is one given by nature, in the sense that its exercise produces value without imposing a cost on any other person.

¹¹²The subscript 0 indicates that the expected value is measured from the distribution at time 0.

¹¹³If c is also uncertain, the value of the option depends on the distributions of $\tilde{v}$ and $\tilde{c}$ not being strongly correlated; the option is especially valuable if they are negatively correlated.

Suppose instead that the firm decides to outsource the construction of the machine to another firm (the “seller”), and the parties agree to a fixed price payable on delivery, P . The buyer is the party who is likely to acquire information that resolves uncertainty in $\tilde{v}$, and therefore is also likely to be assigned the termination right and the associated call option.¹¹⁴ To ensure that the buyer has the correct incentives in exercising that option, her promise is enforced by expectation damages ($P-c$) that compel the buyer to internalize the cost of termination to the seller, its lost profit in this case. The buyer will walk away from the call (i.e. breach) only if its loss from trade ($P-v$) is greater than the damages liability ($P-c$): that is, where $v < c$.¹¹⁵

C. Written Embedded Options and Insurance

As summarized in Part I.C, ex ante investment efficiency may lead the parties to agree to less than compensatory damages in order to induce the seller to take into account the possibility that trade may not occur. We explore in this section another reason for departing from expectation damages. A second type of call option is not given by nature, but instead written by one party (in our analysis, the seller) in favor of the other (the buyer). The seller who writes an option bears a cost when the buyer exercises it (the option is said to be “in the money”). Thus, whenever a termination provision is not ex post compensatory, the seller will have written an option in favor of the buyer. Such written options are efficient if their ex ante value to the buyer exceeds the ex ante cost to the seller, and this may be the case because of the comparative risk-bearing advantage of the seller. The presence of these written embedded options explains why many contracts have termination provisions whose prices deviate from the compensation principle.

1. Two Insurable Risks in Executory Contracting

When insurance concepts are invoked in the analysis of contract damages, they typically

¹¹⁴The volatility of $\tilde{v}$ is not a sufficient reason by itself to assign the option to the buyer. It must also be the case that some of the uncertainty will be resolved by information produced during the term of the option and that the buyer has superior access to this information.

¹¹⁵Other real options may be similarly preserved. For example, if the seller has begun to invest in the machine, the buyer may hold the option to abandon by breaching. In that case, it is important that the seller be obliged to mitigate its damages. See Alexander J. Triantis and George G. Triantis, *Timing Problems in Contract Breach Decisions*, 41 J. L. & Econ. 163 (1998).

concern the risk of the promisor's breach. Thus, a buyer may be willing to pay a premium to be insured against increases in the seller's costs that lead the seller not to perform and the seller may be willing to accept that risk in return for the premium.¹¹⁶ There are several reasons why this trade may be advantageous to both parties. Most obviously, the buyer may be risk averse while the seller is risk neutral. In addition, the seller may enjoy a comparative advantage in reducing the risk of increases in its costs and thus may be better able to manage the risk for the buyer. This much is well known. But the buyer bears two further risks that have thus far been neglected in the economic analysis of contract: a) the risk of fluctuations in the seller's profit that may increase the buyer's liability for breach damages, and b) the risk of changes in the *value* to the buyer of the seller's performance. The buyer, therefore, may be willing to pay a further premium in order to be able to shift both the risk in the seller's cost and the risk in the value of performance. Once again, the parties will agree that the seller insure some or all of these risks either because the buyer is more risk averse or because the seller has a comparative advantage in managing the risk of changes in its costs or in the value of its performance to the buyer.

To illustrate this point, consider the example of the previous section that compared contracting to an integrated firm that exploits the same opportunity. Recall that the value of the project, $\tilde{v}$, is uncertain at t_0 and now suppose that the cost is also stochastic, $\tilde{c}$. The integrated firm bears the risk of exogenous shocks to the joint distribution of $\tilde{v}$ and $\tilde{c}$ that cause $\tilde{v}-\tilde{c}$ to rise or fall and thus might preserve its option to be able to walk away if new information at t_1 reveals that $v < c$. As in the previous section, compare the case where two parties pursue the same project by a fixed price contract with price, P . Under a specifically enforceable contract, the seller bears the cost risk ($P-\tilde{c}$) and the buyer keeps the value risk ($\tilde{v}-P$). The previous section suggested that the integrated firm's real option can be preserved if the contract is enforced by expectation damages. Each party shares in the value of this real option by being able to shed some of her respective downside risk (in particular, the risk that $\tilde{v}$

¹¹⁶E.g., A. MITCHELL POLINSKY, AN INTRODUCTION TO LAW AND ECONOMICS ch. 8. Like other commentators, Polinsky assumes that the seller's costs fluctuate (because of the risk that a new buyer will emerge to offer a higher price for the goods), but that the buyer's valuation is constant.

falls at the same time as $\tilde{c}$ rises) without passing any additional risk to the other party.¹¹⁷ When the parties agree to expectation damages, the seller does not write an option for the benefit of the buyer; it is given by nature.

Beyond the preservation of real options, however, the seller might also *write* an option in favor of the buyer that would further insure the buyer against the risks of the project. When the seller writes the option, the termination fee is not equal to the seller's realized loss from termination (i.e. expectation damages). Recall that under our notation, a regime of expectation damages gives the buyer a call option $(d, x) = (P - \tilde{c}, \tilde{c})$; that is, the buyer incurs a liability of $P - \tilde{c}$ and can exercise its call by paying the balance of the contract price, $\tilde{c}$. Note that both the option price and the exercise price are uncertain at the time the parties enter into their contract and these amounts will be judicially determined at a later date. Consequently, the buyer may be prepared to pay a premium in order to avoid the risk in the option price that is due to fluctuations in $\tilde{c}$. The seller may therefore write a call option with a fixed price (i.e., a liquidated damage provision) that the court would perceive as supercompensatory in the *ex ante* sense. In other words, rather than $(P - \tilde{c}, x)$, the parties may agree to $(E_0(P - \tilde{c}) + k, x) = (P - E_0(\tilde{c}) + k, x)$, where k is the premium paid by the buyer.¹¹⁸

Finally, recall that the parties also have the choice among different combinations of option price and exercise price, (d, x) . By raising the exercise price (x) and reducing the option price (d) below the seller's expected profit, $P - E_0(\tilde{c})$, the parties can shift a larger portion of the valuation risk ($\tilde{v}$) from the buyer to the seller. In this way, the buyer can escape more of the lower tail of the distribution by walking away from her call option when her valuation falls below the contract price. Of course, the buyer in return must agree to a higher exercise price. But if a court interprets the option price (d) as

¹¹⁷ This is the real option described in the previous section, whose value depends on the absence of positive correlation between the distributions of $\tilde{v}$ and $\tilde{c}$. Neither the seller nor buyer suffers from conditions that increase $\tilde{c}$ and decrease $\tilde{v}$, so that $c > v$.

¹¹⁸ This may explain why passengers pay significant penalties when they change their minds about flights they purchase from airlines. These penalties are premiums paid by travelers for insurance against declines in their valuation of booked flights (for example, where the purpose of the trip was a wedding that is cancelled), but they can lead to inefficient trade when the traveler values the flight less than a potential replacement passenger. See TAN *infra*.

stipulated damages imposed in the event of the buyer's breach, these damages will appear to be undercompensatory even though they are a fair price for the call option with that exercise price.¹¹⁹ The discrepancy arises from the fact that compensation measures the seller's lost payoff given that there is no exchange. The option price takes into account the exercise price that the seller will receive in those states of the world in which the option is exercised. From this perspective, true damages more nearly approximate the seller's cost of writing the option than the seller's loss when the option expires without exercise.

2. *The Advantages of Risk Reduction*

Some scholars believe that risk aversion can explain very little of the contracting that occurs in practice.¹²⁰ Although risk aversion may be accurate in describing consumers, it seems inappropriate where business entities are concerned, particularly in light of capital market asset pricing theory.¹²¹ The theory predicts that managers of business firms cannot increase firm value by buying insurance because investors are at least as capable of hedging or diversifying nonsystematic firm risks in their financial portfolios. Firms also cannot create value by hedging systematic risks because shareholders can choose their desired portfolio risk by altering their asset allocation between risky and risk-free assets.¹²² Nevertheless, corporate managers actively hedge interest rate, exchange or commodity price

¹¹⁹If we think briefly about the seller's option, we note that liquidated damages payable by terminating sellers are often pegged significantly below the compensatory level. See note 100 *supra*. For example, repair and replacement are common remedies for defective performance. Contracts also limit recovery for packages lost by couriers or luggage lost by airlines. Recently, distributors of electricity have issued interruptible electricity contracts that allow the distributor some flexibility to interrupt electric service to commercial customers, which they exercise in times of demand spikes. Some contracts provide for financial compensation at the time of the interruption, but others provide for a payment in advance. Neither payment is intended to reflect the loss suffered by the customer. See Ross Baldick, Sergey Kolos and Stathis Tompaids, *Valuation and Optimal Interruption for Interruptible Electricity Contracts* (Univ. of Texas Working Paper 2003).

¹²⁰ See Victor P. Goldberg, *Aversion to Risk Aversion in the New Institutional Economics*, 146 J. Inst. & Theor. Econ. 216 (1990) (arguing that an assumption of risk aversion deflects attention from more significant determinants and should be used only as a last resort). But see *Id.* at 218 (suggesting that insurance questions turn on whether a firm should outsource the task of minimizing the likelihood of the adverse contingency).

¹²¹ See, e.g., RICHARD A. BREALEY AND STEWART C. MYERS, *PRINCIPLES OF CORPORATE FINANCE* ch.7 (6th Ed. 2002).

¹²² See, e.g., Rene Stulz, *Mastering Risk I: Diminishing the threats to shareholder wealth*, Financial Times (April 25, 2000).

risks, and risk management is the topic of much literature both by serious academics and practitioners.

Reducing risk at the firm level has a number of advantages that cannot be replicated by investors and that might therefore lower the firm's cost of capital.¹²³ Under progressive tax rates, a firm's tax liability is lower if income is stable rather than volatile. In addition, the risk aversion of managers raises the cost of realigning their incentives by increasing the portion of compensation they derive from stock or stock options (particularly restricted stock). By dampening the variance in firm profitability, risk management reduces the cost of such compensation and allows firms to load up on incentive compensation.¹²⁴ Moreover, when exogenous risks are reduced, performance-based compensation targets more accurately the skill and effort of managers. Risk reduction also encourages shareholders to concentrate their investments in the firm rather than diversifying their portfolios, thereby improving their incentives to monitor management. In the same vein, volatility also increases the conflict between stockholders and debtholders over capital budgeting and investment decisions. If the probability of insolvency is high, shareholders are more likely to induce their managers to invest in unprofitable risky ventures (the overinvestment problem).¹²⁵ And, managers are more hesitant to invest in lower risk, profitable projects because the payoffs from these projects will accrue to debtholders in the insolvency states of the world (the underinvestment problem).¹²⁶ Reducing cash flow volatility and the probability of insolvency corrects both types of investment distortions.

Volatility in cash flow affects financing costs as well. It raises the probability of financial distress that would impose restructuring costs and disrupt relations with suppliers and customers. The higher risk of financial distress consequently increases the cost of debt capital and lowers leverage. Yet, debt financing has several advantages; most notably, interest payments are tax favored and debt

¹²³For a summary of the empirical articles supporting this proposition, see CHARLES W. SMITHSON AND CLIFFORD W. SMITH, JR., *MANAGING FINANCIAL RISK* 505-9 (1995).

¹²⁴E.g., Rene Stulz, *Optimal Hedging Policies*, 19 J. Fin. Qu. Anal. 127 (1984).

¹²⁵Michael C. Jensen & William H. Meckling, *Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure*, 3 Fin. Econ. 305 (1976).

¹²⁶Steward C. Myers, *The Determinants of Corporate Borrowing*, 5 J. Fin. Econ. 147 (1977).

may reduce managerial agency costs by compelling the disgorgement of free cash flow.¹²⁷ By increasing debt capacity, risk management restores the firm's ability to realize these gains. Volatility in cash flows also affects the availability of internal financing of new projects. If information asymmetries impede the raising of external capital, fluctuations in the available internal funds may prevent efficient capital budgeting.¹²⁸

The goal of risk management, however, is not simply the reduction of the firm's investment risk. After all, the cheapest way to minimize risk might be simply for the firm to hold U.S. Treasury Bills. Diversifying operations reduces risk, but there is considerable evidence that it distorts investment decisions by permitting inefficient cross-subsidization of projects.¹²⁹ A firm might also adapt to its operational risks by holding a buffer of cash or a line of credit, but this financial slack also may yield a temptation for managers to overinvest in unprofitable projects. Thus a firm may look instead to its capital structure and the design of its financial instruments to reduce the danger of insolvency. It may, for example, shorten the maturity of its outstanding debt or condition its interest obligations on the market prices of critical inputs or products.¹³⁰ Flexible manufacturing techniques can also lower the firm's vulnerability to market risks. For example, Hewlett Packard recently implemented a flexible assembly process that allows printers to be customized at the location of the customer rather than at its principal manufacturing plants.¹³¹ In short, risk management is a dynamic, portfolio strategy that includes all of the firm's financial, insurance, derivative and supply contracts.¹³²

¹²⁷Michael C. Jensen, *Agency Cost of Free Cash Flow, Corporate Finance and Takeovers*, 76 Am. Econ. Rev. 323 (1986).

¹²⁸Kenneth Froot, David Scharfstein and Jeremy Stein, *A Framework for Risk Management*, Harv. Bus. Rev. 91-102 (Nov-Dec 1994).

¹²⁹E.g., Philip G. Berger and Eli Ofek, *Diversification's Effect on Firm Value*, 37 J. Fin. Econ. 39 (1995); Hyun-Han Shin and Rene Stulz, *Are Internal Capital Markets Efficient*, 113 Q. J. Econ. 531 (1998).

¹³⁰See Corey Billington, Blake Johnson and Alexander J. Triantis, *A Real Options Perspective on Supply Chain Management in High Technology*, 15(2) J. App. Corp. Fin. 32 (2003).

¹³¹Alexander J. Triantis, *Real Options and Corporate Risk Management*, 13(2) J. App. Corp. Fin. 64-73 (2000).

¹³²See e.g., Christopher Culp, *Contingent Capital: Integrating Corporate Financing and Risk Management Decisions*, 15 J. App. Corp. Fin. 46 (Spring 2002).

D. Risk Management and Insurance in Thick- and Thin-Market Contracts

Forward contracts and explicit options are perhaps the most widely discussed tools of risk management. It is common to think of risk management narrowly in the sense of transactions in traded derivatives (particularly in foreign exchange, interest rates or commodities), where counterparties have a comparative advantage in hedging or diversifying the relevant risks. Yet, the market for derivatives is not complete and the bundle of risks facing a firm cannot be perfectly hedged. Insurance companies have stepped up their competition against financial instruments by providing industrial firms with tailored insurance against bundles of risks, including earnings insurance.¹³³ In addition, therefore, a firm may exploit opportunities to outsource risk in contracts with its customers and suppliers.¹³⁴ As discussed below, embedded options in these relationships play a key insurance role.

1. Forward Contracts and Option Contracts in Thick Markets

For ease of exposition, we adopt the well-known distinction between thick-market and thin-market contracts. The paradigmatic thick-market contract is traded in a liquid market at trivial transaction cost, and trading is motivated primarily by speculation or risk allocation. In a thick market, contracts are clearly divided into forward (or futures) contracts and option contracts. A forward contract is a simple purchase and sale contract in which the parties exchange exogenous risks, either because they have different expectations about future spot market prices or they have different abilities to manage the market risk. Forward contracts in thick markets are not option contracts, nor do they embed options the way we shall describe that thin-market contracts do. Neither party has any meaningful termination or cancellation rights because of the symmetry between the value of a thick market forward contract to the buyer and the performance cost of the seller. For the same reason, there is no real option of delay or abandonment for the parties to share. The substance of pure thick-market contracts renders inapt any discussion of compensatory damages: The enforcement of such forward

¹³³NEIL A. DOHERTY, INTEGRATED RISK MANAGEMENT CH. 14 (2000).

¹³⁴Recent developments in FASB standards have reflected the realization that such contracts, particularly if long-term, have a significant impact on a firm's liquidity and risk profile, and should be disclosed in financial statements. Pursuant to the directives of 401(a) of the Sarbanes-Oxley Act (2002) (Securities Exchange Act 13(j)), the SEC now requires registrants to disclose the quantity, price and duration of purchase obligations. 17 C.F.R. 229.303 (Item 303), (a)(5)(I), (ii)(D).

contracts should be simply the specific enforcement of the market risk allocation.

When options exist in thick markets, they are explicit. Like forward contracts, they are motivated by heterogeneous information, expectations or risk bearing abilities. A seller may write an option in favor of the buyer if risk is costly to the buyer and the seller has a comparative advantage in risk management, such as the opportunity to diversify or to hedge the market risk.¹³⁵ The seller may offer a wide range of alternative option contracts to the buyer. Option contracts are defined by the underlying asset (e.g., the commodity), the exercise price and the maturity date. For any given maturity and commodity (specifically, the probability distribution of the spot price of the commodity at maturity), the price of the option is a function of the exercise price. As before, we refer to this relationship by numerical pairings of option price and exercise price, (d, x) , and note in the case of a call option that there is a range of such pairings, ranging from $(d^{\wedge}, 0)$ to $(0^+, x^{\wedge})$, where $d^{\wedge}$ and $x^{\wedge}$ are the maximum values for the respective variables. To our knowledge, there is no systematic explanation of the factors that determine which pairings are selected by options markets or become the most liquid. Casual observation, however, reveals that commodities options tend to be issued near the money; the exercise prices are near the corresponding spot prices. It may be that options are too risky and difficult to value when the exercise price lies at the tail of the distribution of the underlying asset (e.g. are far out-of-the-money). Thus, they may be correspondingly more difficult for the writer to hedge.¹³⁶ Setting aside solvency concerns, therefore, the most common call option may look something like (d, v_0) , where v_0 is the spot price of the commodity at the time the option is issued.

The holder of an explicit call option contract may walk away from the purchase of the underlying commodity without making any payment other than the option price. The ex post loss incurred by the option writer does not affect the price. Moreover, the option price is not intended to be even ex ante compensatory in the sense envisaged by the legal regulation of liquidated damages. When

¹³⁵For example, the seller of a call option assumes the risk of the underlying asset value when it is out of the money; the buyer bears the risk while it is in the money. The parties might also agree to trade the call option if the seller (buyer) has private information indicating that the asset is overvalued (undervalued).

¹³⁶We note also that, perhaps due also to tax and incentive considerations, executive stock options are typically issued at the money.

courts examine the reasonableness of ex ante estimates of compensation, they think of the seller's loss given the buyer's breach (i.e., the buyer's call option is out of the money). However, the seller is willing to write the call option as long as the option price compensates the seller for the cost of writing the option. This cost is a function of the entire distribution of payoffs, in all states of the world. The option price is simply the price at which the seller agreed to bear the market risk shifted by the option, given its exercise price and maturity.

In a thick options market, the option price is determined by the value of the option to the marginal buyer, which is equal to the cost of the option to the marginal seller. If the option price is negotiated between writer and holder (even though the spot market may be thick), the price will fall somewhere between the value of the option to the buyer and its cost to the seller. Thus, for the inframarginal option seller in the thick market and for the option seller in the negotiated deal, the option price will appear supercompensatory if a court looked at whether it is a reasonable estimate of the ex post loss to the seller given breach. It is also supercompensatory in the sense that the seller may be overcompensated for its cost in producing any feature of a good or service delivered to the buyer. The penalty rule for liquidated damages does not, of course, apply to explicit option contracts. It does apply to bilateral contracts in thin markets, however, even if they contain embedded options that serve the same insurance purposes outlined above. Yet, even in thin markets, damages are effectively option prices that often have nothing to do with compensating the seller's ex post losses from breach.¹³⁷

2. Embedded Options in Thin-Market Contracts

As noted above, risk allocation is the principal purpose of thick-market contracts. Literal performance is often incidental and the parties may even choose to settle through monetary payment rather than physical delivery of the subject matter. Neither party, however, has a termination right. Accordingly, there is a sharp distinction between forward and option contracts in thick markets. A thin-market contract, on the other hand, contemplates physical performance, and one or both parties

¹³⁷ See Scott, *The Case for Market Damages*, supra note — at —.

may make investments specific to the contract that will be lost if performance does not occur.¹³⁸ The widely recognized purpose of thin-market contracts is to protect the parties' specific investments. Yet, thin-market contracts also embed options in the form of termination rights in order to pursue the objective of efficient risk management. Thin-market contracts are incomplete and contracting parties are often superior insurers because of their private information and ability to control relevant risks. In these cases, the option price and exercise price of the buyer's embedded call option are set so as to allocate optimally between the parties the risk of fluctuations in the buyer's valuation.

Consider two sets of contracts in the airline industry. First, we can express the ex post surplus yielded by a passenger flight in a highly stylized form as the difference between the value to the passenger and the cost to the airline. The passenger may avoid some of the lower valuations of the flight in a contract that permits her to breach and pay expectation damages. In the states of the world in which the passenger breaches, however, she still bears the risk of fluctuations in the airline's costs because they would be reflected in the passenger's expectation damages liability. Even with the common law limits of foreseeability, mitigation and uncertainty, the passenger would bear a risk with respect to which she has limited information and control. The parties may therefore contract for fixed liquidated damages. Moreover, as noted above, the passenger may be willing to pay a premium over the mean of the distribution of expectation damages. This liquidated damages amount is overcompensatory in that it exceeds reasonable estimates of the losses suffered by the airline when the passenger does not fly (e.g. the seat is sold to another passenger). This seems to explain why passengers pay penalties of \$100 when they cancel their flights and their seats are resold, often at a higher spot price. For example, on many routes the airplanes predictably fly at full capacity and yet the penalty is still assessed. Compensation is not the objective of the cancellation provision. The \$100

¹³⁸ The distinction between thick markets and thin markets is in some sense an artificial one. Most contracts fall between the two paradigms and are hybrids. It facilitates our analysis to adopt the artificial classification. As noted earlier, we define thick market contracts as those contracts for which risk allocation rather than investment is the motivating purpose underlying the exchange. Thus, by "thin-market" contracts we mean to include all relational contracts where one or both parties contemplate making a relation-specific investment. The investments we have in mind would include the production of specialized goods, the development of human capital specific to a particular deal, or research to acquire information about future costs or prices.

reflects not only the expected cost to the airline, but also the value of the option to the passenger.¹³⁹

Of course, some passengers desire more insurance: They wish to transfer a larger portion of the risk associated with their valuation of the flight. These parties might choose a higher priced fare with a lower cancellation fee. This alternative is characterized in our analysis as a higher exercise price (x) and lower option price (d). The airline might well incur a loss from the passenger's decision to cancel, but our analysis emphasizes that the cancellation fee is the price paid for the call option on the flight rather than compensation for the airline's lost profit from termination.

The airline, in turn, passes some of the risk of fluctuations in passenger demand to its suppliers, including its aircraft manufacturers. For example, an aircraft manufacturer may be able to diversify the risk of fluctuating passenger demand better than the airlines if the manufacturer has a sufficiently diverse customer base. Moreover, the manufacturer also enjoys a comparative advantage in salvaging the aborted exchange by reconfiguring and reselling the planes.¹⁴⁰ Thus, the manufacturer may grant the airline an option to terminate during at least a portion of the manufacturing period.¹⁴¹ The price for

¹³⁹ We find some appreciation of this argument in cases in which courts try to distinguish between liquidated damages, which are subject to the penalty rule, and provisions for alternative performance, which are not. As we noted above in our discussion of the Blockbuster litigation (see TAN 11-12 *supra*), when courts enforce alternative performance provisions, they recognize the value of options and that buyers may therefore be willing to pay a premium over the cost to the sellers of providing such options. In *Pickens v. Blockbuster, Inc.*, 2004 WL 339594 (Cal. App), customers claimed that the late fees were unenforceable liquidated damages because they were excessive: they were higher than the rental rate per diem. In affirming summary judgment for Blockbuster, the court found that consumers might rationally decide to hold on to the video and pay the late fee. *Id.* We argue that all termination rights should be viewed similarly: liquidated damages are the price paid for the call option that a rational buyer might exercise in some states of the world. As in the case of any commodity or service, the price reflects the value of the option to the buyer and thus may exceed the cost incurred by the seller. The courts, therefore, ought not interfere with the parties' freedom to set these termination prices.

¹⁴⁰ There are other examples of termination options that rely on the seller's comparative advantage in salvaging an inefficient exchange. In book publishing, for example, the consignment contract (providing the retailer a right of return) has evolved as the norm in the industry. Among other reasons for the free option to terminate, the publisher can reallocate inventory more efficiently because of specific knowledge about demand in other locations and because of economies of scale in coordination. Kandel, *The Right to Return*, *supra* note – at 353-55.

¹⁴¹ "We had found that we could double-book some delivery positions, on the basis that not all options were exercised.... Was there a way that we could diversify some of this risk by double booking options from two airlines that were perceived to be countercyclical? Our worldwide customer base suggested that we could." John Stonier (Marketing Director Airbus Industrie), *The Change Process*, in TOM COPELAND AND VLADIMIR ANTIKAROV, *REAL OPTIONS: A PRACTITIONER'S GUIDE* 47 (2001).

the option will be negotiated between the two parties and will lie somewhere between its value to the buyer and cost to the seller, depending on their relative bargaining power. Once again, the option price is unlikely to have a strong connection to a court's assessment of the manufacturer's foregone profit when the airline allows its option to expire without exercise.

The parties may also prefer that the seller use general manufacturing methods rather than cheaper specialized methods, so that the buyer can realize benefits from the option. The flexibility gains from general investments may thus outweigh the losses from foregone specific investment.¹⁴² Specifically, the value to the holder of the option might exceed the cost of using general rather than specialized production methods. For example, Airbus manufactures an increasing number of different aircraft using the same technology platform and common parts, and can thereby sell valuable options to its airline customers to convert their orders into orders for other aircraft.¹⁴³ The same flexibility could support an option to terminate because the platform could more readily service other contracts.

In sum, the purpose of thin-market contracts is not only to optimize relation-specific investments. In addition, thin-market contracts often serve an insurance or risk-management function. Thus, parties to thin-market contracts will trade off the benefits of specific investment (which requires commitment) against the risk management benefits of general investment (which requires flexibility). Pursuing this objective further complicates the structuring of termination rights.

¹⁴²Conventional contract theory examines simply the relative benefits of specific investment over general investment. More recent scholarship recognizes that investment is specific or general to varying degrees. E.g., Eric Posner, Alexander Triantis and George Triantis, *Investing in Human Capital: The Efficiency of Covenants Not to Compete* (Working Paper 2003). A seller who uses general manufacturing techniques may be able to profit by selling the resulting flexibility in the form of embedded options. John Stonier and Alexander J. Triantis, *Natural and Contractual Real Options: The Case of Aircraft Delivery Options*, in A. MICALIZZI AND L. TRIGEORGIS, EDS., *REAL OPTIONS APPLICATIONS* 159-195 (1999).

¹⁴³ Airbus allows buyers to choose between the aircraft in its family of A318, A319, A320 and A321 with a very short lead-time. These aircraft are built on the same production line and have many common components, so the decision as to which aircraft is built can be delayed. Airbus explicitly markets contract options and offers guidance to buyers in the valuation of these terms. John Stonier and Alexander J. Triantis, *supra* note --; John Stonier, *The Change Process*, in Tom Copeland and Vladimir Antikarov, *Real Options: A Practitioner's Guide* 47 (2001). "This facilitated our [Airbus'] ability to allow customers to defer, until as late as possible, the decision as to which size of aircraft within the family they would actually take delivery of." Stonier, at 41.

E. Pricing Embedded Options in Thin Market Contracts

The pricing of embedded call options is the price at which flexibility is sold by one party to another. It is likely to be quite heterogeneous across circumstances and this observation motivates our argument in Part III against a damages default rule that purports to set a “price” for breaching and the penalty doctrine that constrains the parties’ freedom to do so. The pricing of termination options is itself an interesting avenue for analysis and, in the following section, we offer a preliminary outline of the relevant factors.

1. Payment and Possession During the Option Term

When a typical explicit call option is written in a thick market (e.g., a commodities option), the option holder (the buyer) pays the option price, but does not advance the exercise price until she decides to exercise the option. This is not always the case with the embedded call options we have described. In some cases, the buyer pays the entire contract price (P) that includes both option price and exercise price, and is entitled to a refund of the exercise price when either she surrendered her option or it expires. In other cases, the buyer pays nothing up front and simply incurs the contingent liability imposed by a court when she walks away from the contemplated exchange. Or, the buyer may leave a deposit equal to a fraction of the contract price or agree to stipulated damages. The significant difference among these possibilities is the extent to which the seller receives money to hold and invest during the option term. If the seller holds some of the exercise price, the buyer foregoes the interest on this sum and must include this lost income in the option price. Thus, for example, a money-back guarantee is not a free call option since the buyer gives up the income on the refundable portion during the option term.

A typical holder of an explicit call option also does not enjoy possession of the underlying asset, and thereby its value in use, until she exercises her option to acquire the asset. Recall that the value of an option comes from the resolution of uncertainty through the production of new information during the option term. For many thin-market goods whose value is subjective rather than purely market driven, the seller will be inclined to give possession and use of the good to the buyer in order to encourage information production. These goods are sometimes referred to as “experience goods.” Rent-to-own centers rely on this phenomenon in making sales and explicitly lease the good to the

prospective buyer during the term of the option. Yet, it is important to bear in mind that parties who let their options expire owe the duties of a bailee to return the good in reasonable (or better) shape. And, in some cases, the use value of the good during the option period may roughly offset the seller's income from the buyer's up-front payment.

2. The Relevant Factors: Exogenous Risks, Moral Hazard and Adverse Selection.

Parties write embedded options in part to exploit comparative advantages in bearing exogenous risks. As with any service, insurance will be provided if the risk bearing cost to the provider is lower than the value to the insured. Thus, if an embedded call option simply entailed the transfer of exogenous risk from buyer to seller, the seller would write the option that maximized the surplus created by the difference between the option's value to the buyer and its cost to the seller. The seller would agree to assume the range of risk over which it has a comparative risk-bearing advantage. As noted earlier in Part II.C, these measures are functions of the multi-variable risk profiles of each party. Given that the underlying exchanges occur in thin markets, the option price is the product of bargaining between the seller and buyer, who will divide the surplus. Thus, it is likely that the price the buyer pays for the option will be higher than the cost of the option to the seller. The seller's cost will be a function of the volatility of the buyer's valuation, as well as the length of the option term.

Given the close connection between options and insurance, it should not be surprising that the determinants of the structure of embedded options should also invoke familiar considerations about moral hazard and adverse selection. The parties cannot completely control for these information problems because the courts cannot verify ex post the cause of any realized loss. In the face of such concerns, the parties must weigh the various factors in choosing whether to include an embedded option and, if so, its exercise price and option price. In the traditional insurance contract, the relevant private information and hidden action lies with the insured party. In the context of an embedded option in a sales contract, however, there are concerns about adverse selection and moral hazard on both sides.

3. Seller's Private Information and Control

We discuss first the seller's private information and its control over the buyer's valuation. The seller's advantage in this respect reinforces the efficiency of shifting the risk to the seller and leads to a

pairing of a lower option price and higher exercise price. In cases where the buyer's volatility is entirely due to her ignorance of information known to the seller or due to the seller's actions (e.g., the inherent quality of the good), we might thus observe a free option.

Even in the absence of superior risk bearing capacity, the seller may provide insurance to the buyer in the form of a call option if the seller has private information about the good that affects the buyer's valuation. For example, assume the buyer's value depends only on the quality of the good, which the seller knows perfectly. A simple contract would face a significant adverse selection problem if the range of possible quality was large. The buyer might offer a price associated with the average value of the product, but this may drive from the market the high quality sellers, leading the buyer to further discount the price. Warranties are well known as signaling devices that bridge the information asymmetry between seller and buyer.¹⁴⁴ Their enforcement, however, requires a court to determine whether the warranty has been breached. The costs of verifying breach may be large relative to the value of the signal, either because the evidence is indeterminate or because the warranty standard is opaque.¹⁴⁵ Therefore, the buyer may be given instead the right to return the good or cancel the deal without the burden of proving breach or incurring prospective litigation costs.¹⁴⁶

The ubiquity of sellers' private information is the most common explanation for the return

¹⁴⁴ See generally George Akerloff, *The Market for "Lemons: " Quality Uncertainty and the Market Mechanism*, 84 Q. J. ECON. 355, 366 (1970).

¹⁴⁵ In this respect, we note the vague default standards for warranty liability under the Uniform Commercial Code. Under the Code's implied warranty of merchantability, for example, the buyer, in order to establish a breach of warranty, must show that the good is not "fit for the ordinary purposes for which such goods are used." See UCC § 2-314(2)(2003).

¹⁴⁶ See Katz, *supra* note -- ("The usual warranty of merchantability gives a moderately strong signal of quality, but an unconditional guarantee of satisfaction sends an even stronger signal").

In some cases where the seller has private information, the buyer may be able to acquire the information at some cost. But if the investigation requires an investment that the buyer will not recover if she fails to purchase the product, then the seller might hold-up the buyer during subsequent negotiations over price. There are several ways in which the buyer's investment in research may be protected. One way is for the seller to give the buyer an option on the good at a specified exercise price.

policy of retailers of goods, particularly those whose value may not be transparent.¹⁴⁷ If the only source of volatility in the buyer's valuation were the quality of the equipment, the seller could ensure that no buyer would wish to exercise her right of return. The seller thus would be prepared to issue the option for free. The buyer's valuation of the good is rarely due entirely to factors within the seller's private knowledge, however. To the extent that there are exogenous risks, such as technological innovations that affect the buyer's valuation, the seller bears a cost in writing the option. She may therefore be inclined to impose a fee or, in the alternative, to bear the cost in order to benefit from the favorable signal. One might think of the right to return in this respect as falling within a category of gifts which sellers give to prospective buyers. Instead of giving the prospective buyer a free dinner, a round of golf or a tote bag, the seller gives a free option. When the exogenous risk increases, the seller may change her strategy to either begin charging an option price or not offering the option at all. For example, hotels and resorts are more likely to levy cancellation charges in high than low season even though the likelihood of finding a substitute customer is greater in the former case.¹⁴⁸ Our analysis suggests that the cancellation charge is not compensation for lost revenue, but rather the fee for the option held by the customer. The call option is more costly to write (and probably also more valuable to the buyer) in high season because customer demand tends to be more volatile than in low seasons where the truncation of the distribution at zero is more significant.¹⁴⁹ In turn, in low seasons the volatility in the value of the reservation is more likely to be the product of the seller's private information. For

¹⁴⁷For example, sellers of electronics equipment have significant private information relative to their customers concerning the capabilities of the products they sell. Other examples of seller private information where a right of return is common include oriental rugs, works of art—such as paintings and sculpture—and similar goods where value is a combination of provenance and aesthetic values.

¹⁴⁸See Katz, *supra* note — at 20-2.

¹⁴⁹To illustrate, consider a ski resort with 100 rooms. On average, it fills 75 rooms in high season and 25 rooms in low season. Compare the effect of an adverse event such as bad weather on the demand in both seasons. In the high season, the resort could lose up to 75 paying customers; in the low season, its losses are bounded at 25 customers, regardless of the severity of the adverse contingency.

One of our students observed that inspection clauses in residential home purchase agreements in Washington, D.C., became tighter or disappeared when that market heated up several years ago. D.J. Moore, *Understanding Real Estate Hot Markets* (mimeo December 2002). We speculate that the cause was the higher volatility rather than heightened demand. To see why volatility matters rather than absolute price levels, consider a hot market in which the prices have jumped but are expected to remain constant for the next year. Under these circumstances, the seller's cost of writing the option is relatively small and the seller may do so to signal its private information about the quality of the property.

example, the seller will have private information concerning the quality and range of services available during the off-season.

Suppose that the value of the good to the buyer depends on effort or investment by the seller that is not contractable. In the parlance of economics scholarship, the buyer's valuation is a function of the seller's cooperative investment. For example, assume that the buyer commissions a portrait of herself and her children by a well-known artist whose prior work demonstrates great skill in capturing the essence of his subjects. The artist might be induced to make the efficient level of investment by giving the buyer a call option with an exercise price equal to the value that results when the seller exerts the efficient level of effort. As noted earlier, however, economists have shown that this solution is undone when the buyer can force the renegotiation of the exercise price and can thereby hold-up the seller with respect to her specific cooperative investment.¹⁵⁰ If the seller could precommit not to renegotiate, then the option serves the purpose of inducing the efficient investment or effort by the seller. Retailers, for example, have largely been successful in preventing renegotiation by constraining the authority of their sales agents to modify sale prices. Specifically, an agent who receives a return is not authorized to give a partial refund of the original purchase price.

4. Buyer Private Information and Control

We now turn to consider the buyer's private information and control that has the same effect as adverse selection and moral hazard in insurance. If buyers are heterogeneous in the volatility of their valuations, sellers may not write costly call options. For example, suppose that buyers are fickle to varying degrees and a seller cannot discriminate among buyers. The seller would charge an option price that reflected the average sensitivity of its buyers. The less fickle buyers would therefore cross-subsidize the options of the more fickle buyers and the less fickle might exit as a result. A competitor retailer (an entrant) may see an opportunity to deny its customers the right of return in order to attract with lower prices the business of the less fickle buyers, thereby leaving the original retailer with a more fickle pool and higher costs that must be recovered by raising prices. There are many possible equilibria, depending on conditions such as the risk aversion of consumers in the market and the

¹⁵⁰ See note 90 *supra*.

volatility of their tastes over time. The entrant might draw away all buyers, leaving an industry without returns. Or, there may be a separation of the market in which the competitors coexist; for example, some restaurants take reservations while their competitors do not. Finally, the original firm may be able to hold on to most of the buyers because of its comparative ability to bear the buyers' valuation risks or because of the presence of significant seller side private information or control. The seller, however, might offer different combinations of option price and exercise price in order to screen her buyers for the volatility of their valuations.

The risk in the buyer's valuation may also be a function of the buyer's efforts and investment. The insurance written by the seller who gives an option gives rise to moral hazard in the buyer's actions, both during and before the contract period. Of course, where the buyer has possession of the good during the option term, she will not have the efficient incentive to take care of it. Moral hazard comes also in more subtle forms. During the contract, the buyer who is protected by insurance will be motivated to forego measures that dampen (or take actions that increase) the volatility of her valuation of the product in order to maximize ex post the value of her option. Thus, for example, a buyer from a retailer with a policy of free returns may underinvest in search activity to determine the suitability of the product to her tastes. Alternatively, when the buyer is an agent of her family, the buyer may fail to invest efficiently in learning whether the product suits the tastes of other family members.¹⁵¹ Because sellers cannot discriminate against buyers on the basis of the risk of return, buyers have the incentive to underinvest in actions that will reduce volatility. The seller, however, can require the buyer to bear some portion of the cost of returns through the expenditure of time in exercising the option even though it is nominally free. Moreover, some returns are limited to a "store credit" which further constrains the buyer's choice and thus motivates the buyer to internalize some of the cost of exercising the option.

5. Summary

In the preceding discussion, we have argued that the embedded options that are common in

¹⁵¹This observation is distinct from Richard Craswell's theory of the effect of damages on search. His focus is on a party's pre-contract search for the appropriate contracting partner. See Richard Craswell, *Offer, Acceptance and Efficient Reliance*, 48 Stan. L. Rev. 481 (1996). We are focusing on the buyer's effort in figuring out the value of the good itself: e.g., whether the family really needs the television.

commercial and consumer contracts are a function of the trade-offs arising from the interaction of comparative advantages in bearing exogenous risks and the problems of hidden action and hidden information between sellers and buyers. Parties are more likely to agree to embedded options and at lower option prices where the seller has superior ability to bear exogenous risks and enjoys private information about and control over the contract good (and where the buyer has relatively little control). Given the heterogeneity of sellers, buyers and the subject matter of exchanges, the price of termination options vary widely. As we have discussed, these prices can be expected to deviate greatly from the compensation that matters in contract damages: the amount necessary to put the seller in the position she would have been in if the exchange were completed. We turn now in Part III to a consideration of the ways in which that central insight informs our understanding of optimal damages default rules.

III. OPTION THEORY AND OPTIMAL DAMAGE DEFAULTS

This article focuses on the default damages provisions in contract law.¹⁵² The state must decide whether to specify a “majoritarian” default term or whether to set a “bargain-forcing” default that induces parties to create their own damages term.¹⁵³ As we discussed in Part I(B) above, courts and legislatures over the past 150 years have created a majoritarian default of full expectation damages.

¹⁵²It has long been recognized doctrinally that, in principle, determining the optimal remedies for breach of contract is no different from the analysis courts use to allocate any other contractual risk. See, e.g., the statement by Justice Holmes in *Globe v. Landa*, *supra* note —.

¹⁵³ A “majoritarian” default rule is one that purports to provide a contract term that the broadest number of bargainers would have stipulated for themselves had their contracting costs been lower. These defaults are justified, therefore, to the extent that they are widely suitable to many contracting parties, thus eliminating both the resource cost and the error cost of negotiating over the proposed term. Robert E. Scott, *A Relational Theory of Default Rules for Commercial Contracts*, 19 J. Legal Stud. 597, 608-9 (1990). On the other hand, a “bargain forcing” default is one that encourages the parties to bargain explicitly over the term in question by penalizing one party (or both) should they remain silent. These rules are not set to reflect the ultimate risk allocation preferred by most bargainers. The most familiar of these defaults are so-called “information-forcing” default rules that induce one party to share important information with the other. *Id.* at 609-610. We prefer the broader bargain-forcing characterization in this paper because it embraces both information forcing defaults as well as other related contract defaults, such as the indefiniteness doctrine. The latter rule encourages the parties to bargain over heterogeneous factors, such as contract price, at the penalty of having the agreement held unenforceable. For discussion, see Robert E. Scott, *A Theory of Self-Enforcing Indefinite Agreements*, 103 Colum L. Rev. 1641, 1645-59 (2003).

Although various authors have questioned the merits of the default,¹⁵⁴ no one thus far has doubted the case for a majoritarian default damages provision or the salience of the compensation principle in shaping the preferred default rule. Yet, the theory of embedded options advanced in this article implies that optimal contract damages have little to do with compensation. Instead, they are determined by the cost to the seller of writing the option, the value of the option to the buyer, and the market in which they transact. Optimal damages, therefore, are as context dependent as the price of goods in sales contracts. Thus, the wisdom of having a majoritarian default needs to be considered explicitly.

A. Conditions for Efficient Damages Defaults.

The case for default rules in contracts rests on the assumption that state institutions, such as courts and legislatures, sometimes can design contract provisions at lower cost than the parties could themselves. The source of this advantage may be economies of scale. Private parties who bear the cost of creating efficient terms to address contracting problems cannot capture the full social gain because they cannot collect a fee from later parties who might copy those terms.¹⁵⁵ As a result, contracting parties may underinvest in contract design. The efficiency of default provisions relies on the satisfaction of two key conditions. First, parties must have the freedom to contract for their own terms because agreements occur in heterogeneous circumstances and the state cannot feasibly write contracts that suit all environments. Second, although the state can better internalize the benefits of default provisions to all parties, the economies of scale available to the state depend on the degree to which the

¹⁵⁴See Charles J. Goetz & Robert E. Scott, *Enforcing Promises: An Examination of the Basis of Contract*, 89 Yale L. J. 1261, 1281-86 (1980). Recently, Richard Craswell summarized the difficulty in selecting an economically efficient default in his critique of the categories of compensatory damages advanced by Fuller and Perdue. After reviewing the various decisions made by contracting actors (search, precautions, reliance, trade), he suggests that “[t]here is... no reason to suppose that the totality of economic effects will *always* favor an award of expectation damages...[W]henver different amounts of damages would be optimal for each of the different incentives to be optimized, the measure that is optimal when *all* of the relevant incentives are considered will often be some hybrid or intermediate number [of the expectation, reliance and restitution interest measure].”

Craswell nevertheless believes that the optimal damages would have some significant link to compensation. Craswell comes to a similar conclusion in his critique of Fuller and Perdue’s three-way classification of expectation, reliance and restitution interests. Craswell, *Against Fuller and Purdue*, *supra* note -- at 109-10.

¹⁵⁵For analyses of how legal practice innovations spread, see Gerald F. Davis and Henrich R. Greve, *Corporate Elite Networks and Governance Changes in the 1980s*, 103 Am. J. Sociology 1 (1997); Gerald F. Davis, *Agents Without Principles? The Spread of the Poison Pill through the Intercompany Network*, 36 Admin. Sci. Q. 583 (1991); Charles J. Goetz & Robert E. Scott, *The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms*, 73 Calif. L. Rev. 261, 291-293 (1985).

needs of contracting parties are homogeneous and observable by the state.¹⁵⁶

The first condition of an efficient default -- the ability of parties to contract for their own, different term -- is generally satisfied in contract law, but it is violated in the case of damages for breach. As noted in Part I, a liquidated damages provision is void as a penalty unless the damages are “reasonable in light of the anticipated or actual harm caused by the breach and the difficulty of proof of loss.”¹⁵⁷ Under this rule, parties must structure their liquidated damages provisions, including alternative termination mechanisms such as pre-paid deposits, so that they accord with the compensation principle. This regulation of liquidated damages would be less problematic if, as courts conventionally assume, parties had no legitimate reason for deviating from the compensation norm. To the contrary, however, when damages are viewed as prices of embedded options that provide insurance to promisors, there is good reason why they might be either under- or overcompensatory. From this perspective, the penalty rule is a form of price regulation that is incongruent with the function of contract law.

The second condition -- the state’s cost-efficiency advantage in designing defaults -- is violated with respect to a wide range of contract terms. In light of the heterogeneity and complexity of commercial and consumer contracts, the economies of scale in fashioning default terms are generally outweighed by the informational advantage enjoyed by the parties themselves. Simple default rules are often inadequate in a complex world.¹⁵⁸ Not surprisingly, therefore, a relatively small number of default *rules* have evolved at common law. A much larger set of default *standards* that address a wide range of contracting contexts, have been proposed in Article 2 of the UCC and the Restatement of Contracts. Contract law typically provides broad standards when the efficiency conditions for default rules are not met. The strategy of adopting default standards rather than rules, however, gives rise to a

¹⁵⁶For discussion, see Schwartz & Scott, *Contract Theory*, supra note -- at 598-601.

¹⁵⁷ See UCC §2-718(2003); Restatement (Second) of Contracts §356(1979).

¹⁵⁸Moreover, state-created default rules may not even be feasible where there is asymmetric information. Defaults must condition on information that the enforcing authority is able to observe. A default rule that conditions on unverifiable information would create moral hazard. Consequently, parties will routinely contract out of these possible defaults. For discussion, see Schwartz & Scott, *Contract Theory*, supra note -- at 605-08.

variety of undesirable effects noted below.

In the following discussion we elaborate our argument that the default damages rule in contract law violates the fundamental conditions of efficient defaults. We make several proposals for reform. First, thick-market contracts simply allocate market risk and therefore should be enforced by a default based strictly on market damages. Second, in many thin-market contracts, breach damages are effectively a price paid by the promisor for insurance against fluctuations in her valuation of the contract. As discussed in Section II.E., this price varies widely depending on the source of the volatility in valuation, and on a variety of other factors that have no connection to compensating the promisee for her losses. Contracting parties should have the freedom, therefore, to agree to liquidated damages of their choosing. Indeed, because prices reflect benefits to buyers, it is not surprising that they can be supercompensatory to sellers. Third, given the heterogeneity of the option price or insurance premium, any majoritarian damages default in thin-market contracts is unlikely to satisfy the second efficiency condition. Thus, in the absence of a negotiated option, our analysis argues for the efficiency of a bargain-forcing default. The alternative bargain-forcing approaches are either to specifically enforce thin market contracts that lack options to terminate or to adopt a default that gives the buyer a free option to terminate.

B. Optimal Damages in Thick-Market Contracts.

Why do parties enter into fixed-price contracts for future delivery of goods that are traded in a thick market? After all, one can always acquire the goods on the spot market at the prevailing price without negotiating a forward contract. In the pure thick market contract, there is no specific investment, no efficient breach, no information asymmetry, and no moral hazard. As we discussed in Part II.C, parties enter fixed price contracts in these markets simply to shift market risks. In the most liquid markets, the parties often do not contemplate physical delivery, but only a closing of their respective positions through the payment of money. In this light, there is no reason for the court to take any action other than to enforce the risk allocation by awarding market damages.¹⁵⁹

¹⁵⁹ The following analysis of optimal damage measures in thick market contracts draws on the text in Scott, *The Case for Market Damages*, *supra* note – at 1160-1179.

Consider a contract to deliver at a future date 1,000 tons of a stipulated commodity at a contract price of \$200,000. Suppose that the seller later contracts to fill the order from a third-party supplier for \$150,000. The buyer then breaches and, at the time of performance, the market price of the commodity is \$100,000. The third party supplier subsequently releases the seller from his contract. The market measure of damages is the \$100,000 difference between the market and contract price. The buyer argues, however, that the seller's loss is only \$50,000 in foregone profits and that this compensatory amount is the appropriate measure of damages.¹⁶⁰

Before we determine the optimal default rule to govern the buyer's breach, we must first understand the motivations that lead parties to write this contract. If the buyer only wanted to guarantee a supply of the commodity at the date of delivery, the parties could have contracted to pay the market price on that date. The fixed price contract, however, has effectively shifted the risk of price fluctuations. After the contract the seller bears the risk of price increases and, in turn, has purchased the reciprocal opportunity to hold the buyer to the risk of a price decline. The seller can deal with this contract risk in a number of ways. The seller may choose to self-insure and bear the entire risk internally. If so, the seller may wait and purchase the contract goods on the spot market before the contract delivery date. Alternatively, the seller can lay off the risk, as in our example, by entering into a forward or option contract to acquire the goods from a third party at a fixed price. In the example, the seller purchased a call option from the third party. We presume that the seller paid something for it. If the court imposed on the buyer a liability equal to \$50,000 in foregone profit rather than the \$100,000 market damages, it would permit the buyer to appropriate (without paying) the benefit of the seller's contract with the third party. This policy will deter such third party contracts in the future and, as a result, perhaps even similar initial contracts between sellers and buyers. The courts, therefore, should not read options (e.g., to terminate and pay foregone profits) into forward contracts such as the one at issue in the example. Forward contracts entail clear risk allocation that is enforced by granting the seller damages based on the full \$100,000 market risk.

¹⁶⁰This facts of this hypothetical are loosely drawn from *Nobs Chemical, USA, Inc. v. Koppers Co.*, 616 F.2d 212 (5th Cir. 1980).

Driven by the compensation principle, however, most courts have declined to award market damages in similar contexts.¹⁶¹ An especially attractive focal point has been § 1-305 of the UCC, which expressly incorporates the compensation principle.¹⁶² This provision seems to trump the various arguments for market damages. Thus, the conclusion reached by many (but not all)¹⁶³ courts is that market damages are not appropriate where the market fluctuation deviates substantially from the ex post economic loss to the promisee. These holdings, however, are based on a misunderstanding of the motivation for thick-market contracts, which is simply the allocation of risk. The use of the compensation principle to depart from market damages in thick market contracts is inconsistent with the parties risk-management goals. In these thick-market contexts, the courts should simply enforce the parties' risk allocation. If parties prefer an alternative risk shifting mechanism offered by option contracts, the court should similarly simply enforce the risk allocation that they have bargained for.

C. The Case Against Compensation in Thin-Market Contracts

Most contracting does not occur in the paradigmatic thick market context. The issue of optimal damages rules is more complex, therefore, when we turn to thin-market contracts (or hybrid thick-thin markets). Recall that these are contexts in which one or both parties contemplate making a specific investment in the contract. The central insight of the embedded options approach is that in thin or hybrid markets there will be a wide variety of termination provisions and that the measure of damages for breach is merely a subset of this larger family of termination options. Like other termination fees, contract damages are essentially the nonrefundable portion of the contract price. That portion is nonrefundable because it is the price paid for the termination rights which, by definition, the buyer does not recover. In our analysis, it is the price of the buyer's call option, and we would normally expect the

¹⁶¹See *Nobs Chemical, USA, Inc. v. Koppers Co.*, 616 F.2d 212 (5th Cir. 1980); *Allied Cannery & Packers, Inc. v. Victor Packing Co.*, 209 Cal. Rptr. 60 (1984); *H-W-H Cattle Co., v. Schroeder*, 767 F.2d 437 (8th Cir. 1985); *Coast Trading Co. v. Cudahy Co.*, 592 F.2d 1074 (9th Cir. 1979); *Union Carbide Corp. v. Consumers Power Co.*, 636 F. Supp.1498 (E.D. Mich. 1986).

¹⁶² "The remedies provided by this Act shall be liberally administered to the end that the aggrieved party may be put in as good a position as if the other party had fully performed...." UCC §1-305(1) (2003) (formerly §1-106(1)).

¹⁶³*Tongish v. Thomas*, 840 P.2d 471 (Kan. 1992) (compensation principle does not trump market damages simply because contract-market differential exceeds promisee's economic loss); *Transworld Metals, Inc. v. Southwire Co.*, 769 F.2d (2d Cir. 1985 (same)).

price of the option to lie somewhere between the option's value to the buyer and its cost to the seller.

1. The Costs of the Compensation Principle

a. The penalty rule. Once we see the damages award as essentially a price for a service (insurance against fluctuations in buyer's valuation), the penalty doctrine becomes a major impediment to efficient contracting. Contracts scholars have previously demonstrated that the premise underlying the regulation of liquidated damages—that parties have no good reason to write damages clauses that depart from the compensation principle-- fails for several reasons.¹⁶⁴ Options analysis supplies an even more devastating critique of the penalty rule and, generally, of the doctrinal mandate that liquidated damages adhere to the compensation principle. If damages are merely termination fees and thereby option prices, and if parties use termination rights to allocate risks, those fees/prices do not even purport to reflect the loss suffered by the seller from the buyer's option to terminate. This is patently true with respect to the courts' focus on ex post loss and on ex ante reasonable estimates of loss given breach. Indeed, the sale of embedded contract options will often yield a profit to the seller above the cost of providing insurance to the buyer against fluctuations in the buyer's valuation of the contract performance, in much the same way that the seller of a car will make a profit on the sale of options such as a sunroof and CD player. The option fee is a function of the value of the option to the buyer and not a function of the ex post loss suffered by the seller. Thus, there is no penalty in any meaningful sense of the word.

Price regulation is not the domain of contract law. There is no reason why a seller should not be able to profit on providing insurance to her buyers.¹⁶⁵ Given the heterogeneity of call options and

¹⁶⁴ Some of the criticism of the penalty rule has come from within the compensation norm. For example, as a result of the doctrines of foreseeability and uncertainty, expectation damages are limited to the losses from breach that the promisee expects to be able to verify to a court. A liquidated damage clause that incorporates observable but nonverifiable values thus will be vulnerable to a penalty claim even when the clause accurately measures the promisee's lost expectation. As a consequence, sophisticated parties are discouraged from using liquidated damage clauses even when these clauses would otherwise be optimal. See Goetz & Scott, *Liquidated Damages*, supra note -- at --. Other authors suggest that supercompensatory damages may be necessary to achieve efficient relation-specific investment. See note 95 supra.

¹⁶⁵ Ian Ayres and Barry Nalebuff discuss an unfair trade practice action brought by the state of Connecticut against a car rental company who charged a fee of \$150 if the rented car is driven over 80 mph for more than two minutes. They note that the state did not challenge a provision in the rental contract charging \$5/gallon of fuel if the

contexts, as well as their complexity, courts would be far less suited to judging the fairness of option prices than the sunroof and CD players mentioned above. Indeed, the value of a call option to a buyer is a function of the variance of the buyer's valuation, which is significantly more difficult to assess than mean values. It seems better policy to permit parties to bargain explicitly over call option prices.¹⁶⁶

b. The lost volume confusion. The costs of the compensation principle are not limited to the inefficiencies of the penalty rule. Consider the paradigmatic lost volume case in which the buyer's termination deprives the seller of a sale, so that the seller's economic loss systematically exceeds market damages.¹⁶⁷ In these cases, scholarly debate has focused on how much of the seller's selling costs or overhead were "consumed" by the breaching buyer and whether the default measure of damages ought to be the full profit lost by the seller (which may be overcompensatory) or incidental damages (which may be undercompensatory). But the focus on lost volume and selling costs is a red herring. Rather, the choice between market damages and lost profits is a choice between alternative termination provisions, or embedded call options. The parties may choose among pairs of option prices and exercise prices, (d, x) . If the exercise price is very large, the option price (reflected in the breach damages) may be very small. The parties do not intend that the option price reflect the seller's loss from the terminated sale, but only that it exceed the seller's cost in writing the option. Near the other end of the range, they might agree that the buyer can terminate without being liable for any damages or forfeiting any deposit; in this case, the option price is simply equal to the income from a deposit during

renter failed to return a full gas tank. They suggest that courts may view terms that are framed explicitly in terms of price (such as the \$5 per gallon of gas) more charitably than they view terms that are framed as "penalties." New York Times.

¹⁶⁶ In this vein, however, there may be a relatively weak vestige of a case against supercompensatory liquidated damages. To some buyers, notably consumers, the term "damages" may suggest that the seller represents his loss from breach as equivalent to that amount. If this amount were supercompensatory, the buyer might be induced to overestimate the cost of the option to the seller and thereby overpay for it. At most, this argument speaks in favor of striking down penalties only when explicitly framed as liquidated damages.

¹⁶⁷ See e.g., *Neri v. Retail Marine Corp.*, 285 N.E. 2d 311 (N.Y. 1972). The lost volume claim is particularly salient when the market damage measure (the difference between the contract price and market price) is zero. Sellers argue that, although they did in fact resell the contract goods at the contract price, the second buyer would have purchased anyway. Therefore, if the breaching buyer had performed the contract, the seller would have realized two profits rather than just one. Thus, the seller claims the lost profit from the breached contract as compensation for his lost expectancy. SCOTT & KRAUS, *CONTRACT LAW AND THEORY* at 1091-1105.

the option period. Indeed, as we discussed in Section II.E, the parties might rationally choose even a free option either as a mechanism for addressing the seller's private information and moral hazard in performance or as a marketing tool. In sum, whether a given volume seller would have chosen to write an option to a buyer and at what price the option would be offered simply cannot be determined a priori. In any event, the termination provision is unlikely to have much to do with compensating the seller for the lost sale.

The compensation principle (and the penalty rule) corrupts the way courts apply damages in lost volume cases in a more subtle way. The rule prevents sellers of special order goods from writing efficient option contracts that require the buyer to pay a nonrefundable deposit. The express terms of UCC §2-718(2) effectively limit the deposit that can be retained by the seller to either \$500 or the amount permissible as liquidated damages under the penalty rule.¹⁶⁸ In response to the regulation of deposits, many sellers then sue for lost volume damages.¹⁶⁹ While scholars have criticized the courts for granting lost volume damages in this circumstance, the impulse of the courts may well be to provide the seller a "make up" for the invalidation of the contracted-for termination fee.¹⁷⁰ In turn, the precedential effect of these decisions may lead courts to grant lost volume damages in other contexts, where it is likely that the parties would have preferred a low cost option.

2. Expectation Damages Cannot be Justified as a Majoritarian Default

The compensation principle has given rise to a number of other doctrinal controversies that similarly miss the contract goals of risk allocation. Lost volume sales is just one area in which the measure of expectation damages are a point of controversy. There are many others. Consider, for

¹⁶⁸UCC §2-718(2)(2003) provides that "Where the seller justifiably withholds delivery of goods because of the buyer's breach, the buyer is entitled to restitution of any amount by which the sum of his payments exceeds

(a) the amount to which the seller is entitled by virtue of terms liquidating the seller's damages in accordance with [the penalty rule], or

(b) ...twenty percent of the value of the total performance for which the buyer is obligated under the contract or \$500, whichever is smaller.

¹⁶⁹ The offset for lost volume damages is implicitly invited by UCC §2-718(3).

¹⁷⁰See, e.g., *Rodriguez v. Learjet*, 946 P.2d 1010 (1997); *R.E. Davis Chemical Corp. v. Diasonics, Inc.*, 826 F.2d 678 (1987).

example, the cases in which a buyer (or lessee) purchases the right to mine the seller-lessor's land in return for a royalty on extracted ore and a promise to regrade the land upon conclusion of the lease. Assume the buyer-lessee breaches the agreement by failing to regrade the land. Courts have divided (bitterly) over the proper damages default.¹⁷¹ Should the lessor be entitled to cost of performance damages when the economic loss from breach (as measured by the diminution in market value) is substantially less than the cost of regrading? Once again, the focus on expectation damages as a majoritarian default is misplaced. These grading contracts, just as lost volume sales, each have idiosyncratic circumstances that are critical to the choice of call option/damages measures. As the parties trade off the various insurance and information factors identified in Part II, they choose among a wide range of option prices and, if the termination is a breach provision, an equally wide range of damages measures. Only occasionally will the option price reflect the amount of compensation for the lessor's loss. The equivalence of option price and compensation is even less likely when the parties also seek to provide incentives for efficient relation-specific investment. Any attempt to generalize damage measures across these different contexts is doomed. Yet, default rules covering each of the various combinations of option and exercise prices would simply be too expensive to create and to enforce; there would be almost as many legal rules as there are sets of contracting parties. Thus, broad categories are unlikely to improve much on the current default damages rule.¹⁷²

Although expectation damages cannot be justified as a majoritarian default rule, it is tempting to argue that a default damages measure for even a smaller plurality of parties would at least save the costs of contracting in those cases. Indeed, expectation damages might be the most plausible candidate

¹⁷¹Compare *American Standard, Inc. v. Schectman*, 439 N.Y.S. 2d 529 (1981) with *Peevyhouse v. Garland Coal & Mining Co.*, 382 P. 2d 109 (1962).

¹⁷²Schwartz and Scott have argued that an efficient default rule is one that is relatively simple in form and suitable for a wide variety of contracting parties. The criterion of simplicity is a function of institutional competence. Courts cannot conduct investigations into the efficiency properties of all possible rules and rule combinations. See Schwartz & Scott, *Contract Theory*, supra note —, at 598-601. The suitability criterion is even more difficult to satisfy in connection with breach remedies because, as we have shown, contracting parties are exceptionally heterogeneous regarding termination provisions. Schwartz and Scott are similarly critical of dealing with heterogeneity through standards rather than rules. They argue that standards increase the uncertainty of litigation, reduce the likelihood of settlement and encourage parties to increase their expenditures on litigation. *Id.*, at —. Richard Craswell has shown that the courts have over time effectively created a default standard based on the broad categories of over-expectancy, expectancy or under-expectancy. See Craswell, *Against Fuller and Purdue*, supra at —. regarding termination provisions.

on this ground because of its historical and doctrinal salience. But writing defaults for minority preferences imposes costs on all those parties who would prefer to contract out. These costs are not trivial. The gain of having a rule appropriate for a plurality of parties yields a lower social gain than one satisfying a majority. Against this limited gain, we must set out the added cost to the majority who must contract away from the default rather than draft their contract on a blank slate. Courts tend to regard state-created defaults as presumptively fair or efficient and this institutional bias raises the cost of contracting out.¹⁷³ This is particularly true in the case of the expectation damage measure because contract doctrine explicitly identifies the compensatory purpose of damages and the correspondingly acceptable rationale for liquidated damages--the difficulty of verifying the promisee's loss. The same historical and doctrinal salience of expectation damages that make it an attractive plurality default is likely to cause it to be even stickier than most other default provisions. Thus, expectation damages impose a cost on the majority of contracting parties who wish to adopt any among a range of liquidated damages as termination options.¹⁷⁴

D. Bargain-Forcing Defaults.

The foregoing section demonstrated that expectation damages, and the compensation principle in general, fail as the basis of a majoritarian default rule. Expectation damages are often at odds with the intention and needs of the contracting parties, and therefore parties will incur transaction costs in their efforts to contract away from the inefficient default. The key question, of course, is not whether the majoritarian default of expectation damages is inefficient (which it seems clearly to be) but whether it is inferior to any other alternative default that might be proposed. One option might be to extend the

¹⁷³See, e.g., *Hayward v. Postma*, 312 Mich. App. 720, 724, 188 N.W. 2d 31,33 (1971) (parties must use clear and unequivocal language to shift liability for risk of loss from seller to buyer); *Caudle v. Sherrard Motors Co.*, 525 S.W. 2d 238, 240 (Tex. Civ. App. 1975) (same); *Davis v. Small Business Inv. Co.*, 535 S.W. 2d 740, 744 (Tex. Civ. App. 1976) (contractual provision purporting to allocate to debtor the burden of "all" expenses incurred in preserving collateral not an "agreement otherwise" sufficient to opt out of UCC §9-207(2)(a)). Moreover, judicial interpreters may be reluctant to give the express language of the contract a meaning that conflicts with the relevant default. See *Nanakuli Paving & Rock Co. v. Shell Oil Co.*, 664 F.2d 772 (9th Cir. 1981) (merger clause that excludes evidence of prior dealings does not bar evidence of usage of trade to alter the price term in the contract). For discussion, see Goetz & Scott, *The Limits of Expanded Choice*, *supra* note — at —.

¹⁷⁴The same criticism may be leveled at the regulation of termination rights under the requirement of good faith. See Goldberg, *Discretion in Long-term Open Quantity Contracts*, *supra* note —.

market damages default that applies in thick markets to the thin market context as well.¹⁷⁵ This proposal has several advantages. First, it eliminates the inevitable boundary problems of distinguishing between thin- and thick-market transactions. Second, market damages have the attractive feature of being a default rule rather than a default standard.¹⁷⁶ But this suggestion flounders on the difficulty of verifying the relevant “market price” in transactions in which the market is thin, by definition. The inability to verify market prices would produce the very same litigation cost problems that commended the proposal in the first place.

In the absence of a thick market, therefore, a preferable approach is to establish a bargain-forcing default that both encourages parties to negotiate option prices on their own and does not create burdensome judicial measurement costs when they fail to do so. Thus, if the parties are not explicit about an embedded option or termination right, the courts should not read one into the contract. One way to force the parties to bargain explicitly is to adopt specific performance as the general default rule.¹⁷⁷ The argument for a specific performance default follows from the fact that a binding agreement to exchange goods or services for a price in thin markets implies that the parties have chosen to form a legally enforceable contract in order to credibly motivate the seller-promisee to undertake a relation-specific investment.¹⁷⁸ Specific enforcement of thin-market investment contracts thus enables the buyer-promisor to make credible promises to pay the contract price.

¹⁷⁵Despite the high degree of homogeneity among thick-market contracts, the market damages default serves the principal contracting goals in these markets; namely, to allocate market risk. Indeed, market damages default may be a misnomer because in essence they specifically enforce the parties’ contract.

¹⁷⁶Supra note 172 (regarding costs of default standards).

¹⁷⁷Granting parties the right to contract for specific performance (other than for personal service contracts) would return the common law to its historical roots described in Section I.B. The device of the penal bond was a means by which parties through the 18th century could, in effect, contract for specific performance. The demise of the penal bond therefore eliminated parties’ ability to contract for specific performance. The case for specific performance is best understood by visualizing the termination option as creating a contract plus a put. Where the parties fail to negotiate the exercise price of the put, the “no option” solution would be logically compelled and the buyer would be specifically required to pay the full purchase price for the goods or services.

¹⁷⁸If the buyer’s promise to pay the contract price were not enforceable (say, because the buyer could exercise a free option to terminate), sophisticated commercial parties would decline to make the specific investment because of the risk being held up by the buyer once the investment costs were sunk. Supra note 90.

To be sure, specific performance may lead to overinvestment. Moreover, as we argued in Part II.D above, the parties will often want to trade off some of the benefits of specific investment in order to reap the risk management advantages of an embedded option. But the heterogeneity of option prices implies that only the contracting parties are able to determine the optimal trade off between the benefits of specific investment (which requires commitment) and the risk management benefits of general investment (which requires flexibility). Thus, we would expect that, in equilibrium, many (if not most) parties would respond to the specific performance default by explicitly setting out termination rights in their agreements. The specific performance default forces the parties to bargain explicitly over option prices whenever such flexibility is value maximizing for both parties. Moreover, the specific performance approach is well suited for contracts between commercial parties because it closely parallels thick market contracting. As we observed earlier, thick market contracts are clearly segmented into forward or future contracts on the one hand and option contracts on the other. If a contract does not specify an option, none will be implied. In this way, a specific performance decree functions to enforce the allocation of market risks just as the award of market damages does, but without the need to verify market prices to a court.

In the case where the buyer is the promisor, a specific enforcement default is straightforward. It authorizes an action by the seller for the price of the contracted goods or services.¹⁷⁹ In this context, the issue of defective or inadequate performance is not relevant: either the buyer has paid the price or he will be ordered to do so. On the other hand, where the promisor is the seller, the proposed default is somewhat more complicated. Specific performance would be available to order the seller to perform an executory promise (other than for personal service contracts). In addition, a buyer would have recourse to existing default rules governing implied and express warranties to recover money damages whenever the seller's performance failed to meet the contract requirements.¹⁸⁰ A specific performance default does not eliminate the need to calculate damages in cases of nonconforming performance. But the rule

¹⁷⁹See e.g., UCC § 2-709.

¹⁸⁰ Thus, for example UCC §2-714 would still be applicable. Under this provision, the buyer who takes possession of non-conforming goods may recover damages for any non-conformity in the seller's performance. Under 2-714(2) the measure of damages for breach of warranty is the difference between the value of the goods delivered and accepted and the value they would have had if they had been as warranted.

does reduce judicial measurement costs in resolving disputes over executory contracts. Moreover, a general rule of specific performance eliminates the presumption that breach of an executory contract requires compensation, a presumption that has made the expectation damages default rule so sticky. It thus reduces transactions costs for those parties who are motivated to bargain over option prices. The embedded options approach thus provides yet another reason supporting the existing arguments in favor of the default of specific performance.¹⁸¹

A specific performance default may not be appropriate in consumer transactions, however. Especially in the case of standardized goods, for which a merchant seller has not made a specific investment, requiring the consumer buyer to pay the full purchase price and accept delivery of unwanted goods would serve no efficiency purposes and could work unnecessary hardship on consumer buyers. Merchants also are more likely to be superior bearers of risk because of their ability to spread such risk among their customers and to control the quality of their products. Moreover, as a bargain- forcing rule, specific performance is ineffective against consumers who typically do not negotiate the terms of their contracts. Merchants typically draft consumer contracts, and the legal rule might better focus the incentive to provide explicitly for termination options on the seller, the drafter of the contract.

We propose, therefore, that absent an specific agreement respecting termination rights, the default rule provides consumer buyers free options: the right to walk away from the executory exchange, but to hold the merchant seller to the deal. The default would motivate those sellers who prefer to offer buyers a priced option or no option at all to contract out of the default in their standard forms or, in the case of retail sellers, by posting appropriate notification at the point of sale.

For other sellers, such as retail sellers of electronic goods or merchants who rely on direct mail sales,

¹⁸¹For a persuasive argument in favor of a default remedy of specific performance, see Alan Schwartz, *The Case for Specific Performance*, 89 Yale L. J. 271 (1979). Specific performance appears to be routinely available as the default remedy in most civil law regimes. See BARRY NICHOLS, *FRENCH LAW OF CONTRACT* 211-13 (1982); GUNTHER H. TREITEL, *REMEDIES FOR BREACH OF CONTRACT: A COMPARATIVE ACCOUNT* 51 (1988); Deborah E. Townsend, *The Foreign Economic Law of the People's Republic of China: A New Approach to Remedies*, 24 Stan. J. Int'l L. 479, 485 (1988). See also, United Nations Convention on the International Sale of Goods, Article 46(1) which provides that "the buyer may require performance by the seller...unless the buyer has resorted to a remedy which is inconsistent with this requirement." For discussion, see Steven Walt, *For Specific Performance Under the United Nations Sales Convention*, 26 Texas Int'l L.J. 230-31 (1991).

the free option default would merely replicate their existing terms of sale. To be sure, the free option default for consumers requires courts to craft a set of associated rules governing the duration of the consumer buyer's call option. A convenient analogy exists with the buyer's right to reject defective goods under the UCC. Section 2-602 provides that a rejecting buyer must exercise the right of rejection within a reasonable time, cannot thereafter exercise ownership inconsistent with the seller's rights and must hold the rejected goods for the seller with reasonable care.¹⁸² Applying a similar standard to the terminating buyer would protect both the buyer's option and the seller's interest in the underlying goods or services.

Insofar as contract enforcement against the consumer buyer is concerned, the indefiniteness doctrine of contract law offers an appropriate doctrinal home for the free option default.¹⁸³ Unless the parties price a call option, courts would lack "a reasonably certain basis for giving an appropriate remedy" to a merchant seller.¹⁸⁴ The free option consumer default is similar to the indefiniteness rule that courts apply when parties fail to provide a contract price for services for which there is no established market price.¹⁸⁵ Under the common law, courts declare executory agreements void for

¹⁸²UCC §2-602 (2001). In turn, a failure to reject in time or a wrongful exercise of ownership inconsistent with the seller's interests constitutes an acceptance under § 2-606.

¹⁸³ Conventional wisdom holds that modern courts will enforce indefinite contracts by filling contractual gaps with general standards of reasonableness and good faith. See e.g., 1 ARTHUR CORBIN, CORBIN ON CONTRACTS §95 at 400 (1963). This conventional wisdom is misleading, however. A systematic review of the case law shows that American courts continue to dismiss claims for breach of contract on the grounds of indefiniteness, often without granting any relief to the disappointed promisee. See Scott, *Self-Enforcing Indefinite Agreements*, supra note — at ---.

¹⁸⁴ See UCC §2-204(3) (indorsing the indefiniteness doctrine when courts lack the basis for imposing an appropriate remedy). The indefiniteness rule would apply only where the agreement was executory and the buyer declined to pay the contract price and retain the contract goods or services. On the other hand, if the buyer affirmed the agreement and exercised its option to buy the goods or services, then the conduct of the parties would establish the existence of an enforceable agreement binding on the parties. See e.g., Restatement (Second) of Contracts §§ 4, 30(1)(1979), UCC §2-207(3)(2003).

¹⁸⁵The Uniform Commercial Code authorizes courts to provide a reasonable price where the parties have not specified a price. UCC §2-305(1)(2003). This provision, however, simply follows the line of cases holding that price terms in sales contracts could be supplied from evidence of market prices. *Varney v. Ditmars*, 217 N.Y. 223, 111 N.E. 822 (1916). But where there is no relevant market price, courts continue to decline to enforce such agreements on the grounds of indefiniteness. *Cobble Hill Nursing Home, Inc. v. Henry and Warren*, 74 N.Y. 2d 475, 548 N.E. 2d 203 (1989). For discussion, see Scott, *Self-Enforcing Indefinite Agreements*, supra note — at 1645-59. 1645-59 (2003).

indefiniteness if the parties fail to specify readily verifiable measures of performance in realized states of the world. The indefiniteness rule might be understood as a kind of global bargain-forcing default, one that uses the threat of non-enforceability to encourage parties to specify the solution to certain contingencies themselves. Viewed from this lens, a free-option, bargain-forcing default in thin or hybrid consumer markets would require merchants to address termination and breach as unified issues.

CONCLUSION

Many consumer and commercial contracts contain explicitly negotiated termination provisions. We have analyzed these provisions as embedded options that serve a valuable risk management function. We have focused attention on the buyer's right to terminate as creating a call option and providing a type of insurance against decreases in the buyer's valuation of the exchange. Parties in different circumstances thus may agree to a wide range of option-price and exercise-price combinations. This perspective allows us to analyze contract damages as embedded call options that serve similar insurance purposes. Given that contract damages are option prices, the inquiry dictated by the penalty rule – whether stipulated damages reflect the actual or expected loss to the seller when the buyer walks away from the exchange – is wholly inapt. The price of the option is a function of its value to the buyer and its cost to the seller. Only in unusual circumstances will that option price reflect the ex post loss suffered by the seller if the option is exercised.

The option price is set at or above the cost of the option to the seller, depending on the exercise price. A court might ask instead whether stipulated damages are a fair price for the option in light of its exercise price. But that would take the courts into the realm of price regulation, a task they appropriately eschew in other contexts.¹⁸⁶ The implicit premise is that prices are extraordinarily heterogeneous and parties have far superior information than the courts in this regard. We argue not only that contracting parties should be free to set the prices for their embedded options, but also that the

¹⁸⁶ Antitrust law polices markets in which prices are determined, but there is otherwise no broad attempt by the law to regulate prices.

law should not provide default damages rules that effectively establish prices for the parties when they fail to do so for themselves. Rather, in commercial transactions, the courts should specifically enforce contract exchanges in order to force parties to bargain over termination rights and embedded options. In light of the fact that merchants typically draft contracts in consumer transactions, we propose that they be encouraged to specify termination rights explicitly by enforcing a contract default that would give consumers free options, the right to terminate without charge.

In many ways, our proposals reflect nostalgia for contract law of the early 19th century, before the compensation principle led the courts to expand the reach of expectation damages to include lost profits and before they began to aggressively strike down penalty liquidated damages. In this light, our project in this article is to trim back the unruly branches of the common law. One might argue that our position ignores the dramatic economic changes that have occurred over the past two centuries. But to the contrary, our approach is justified by the concomitant increase in the heterogeneity of contractual contexts, in the rapidity with which parties can access and process new information, and in the appreciation for the benefits and tools of risk management.