THE OPTION ELEMENT IN CONTRACTING

Avery Wiener Katz[*]

MOST contractual arrangements are either structured as op-
tions or include options as important elements. As a result,
many of the major doctrines of contract law effectively operate to
create or to set the terms of such options. For instance, it has long
been recognized that a contract that is enforceable only through
monetary liability operates in practice as an option, because as a
legal matter the promisor retains the power either to perform or to
breach and pay damages. Similarly, the doctrine of promissory es-
toppel, which attaches liability to precontractual statements in
cases where they are reasonably relied upon, effectively grants an
option to the relying party to enforce the promise or not as she
finds convenient. Similar options arise where contracts are void-
able—but not void—for reasons of mistake, lack of capacity, or
fraud.

   Despite this connection, the law of contracts has often treated
options quite differently from other contractual transactions. Op-
tion contracts with an explicit zero premium were not enforceable
under the traditional common law, for instance, and even today are
only enforceable if the contracting parties undertake special for-
malities.[1] Conversely, the characterization of a transaction as an
option contract can have the effect of relieving parties from doc-

---

[1] See Restatement (Second) of Contracts § 87(1) (1981); U.C.C. § 2-205 (2004).

trinal limitations on their contractual freedom, such as the duty to mitigate damages or the rule that holds excessively high liquidated damages void as penalties.

Such differential treatment is somewhat challenging to explain from an economic viewpoint, both because all contracts resemble options in the aforementioned sense, and because contracts that are nominally structured as explicit options can be close economic substitutes for contracts that are nominally structured as unconditional. For example, a party who is willing to accept a zero premium option with a given exercise or "strike" price should be equally willing to accept a positive premium option with a correspondingly lowered strike price or extended option term. Conversely, a party who is willing to accept a zero premium option with a given term should be willing to accept a positive-premium option with a correspondingly longer term. Why, then, should the law draw a sharp distinction between zero premium and positive premium options?

Similarly, a party who demands a substantial prepaid deposit or liquidated damage clause—economically equivalent to a high option premium and a low strike price—should be willing to agree to a lower deposit or liquidated damages figure in exchange for a correspondingly higher purchase price. What factors, then, apart from a desire to avoid the constraints of the common law penalty doctrine, determine how parties choose between these possible alternatives? And how should the law distinguish, if at all, between front loaded option contracts on the one hand, and large deposits that are formally styled as options for the purpose of evading the penalty doctrine on the other?

Such questions arise not just out of doctrinal puzzles but out of transactional problems as well, even when the relevant transactions are unconstrained by legal strictures. Sellers often fail to use option-based pricing policies in circumstances in which doing so would be perfectly feasible and would appear to serve their interest. For example, it has been argued in defense of the practice of resale price maintenance ("RPM") or the awarding of lost-volume damages that sellers need to charge an above-marginal-cost price on retail output in order to cover the cost of precontractual or overhead sales expenses. But such arguments assume that it is infeasible or unprofitable to charge customers for the seller's sales

investments up-front through a cover charge or entrance fee that is equivalent to a straightforward option. In most cases there are no obvious barriers to doing so, but in the retail and wholesale context such arrangements are rare, with the exception of a few superstores organized as private purchasing clubs.

In order to begin to address such questions, it is necessary to set out a substantive account of the efficient design of option contracts—one that explains how contracting parties should strike the balance among option premium, option term, and exercise price, in order to maximize the expected surplus from exchange. This Article will present such an account; it will show that the tradeoff between these various aspects of option contracts can affect the parties' incentives to acquire and disclose information, to make relation-specific investments, and to take efficient precautions against breach. The appropriate balance between option premium, option term, and exercise price, accordingly, ultimately depends on the relative importance that the parties attach to these various incentives.

This Article will also show how option contracts can be profitably used in ways that do not necessarily improve the efficiency of the underlying transaction but that advantage the parties using them at the possible expense of others participating in the market. For example, options can be used to signal or to screen for private information in the context of adverse selection and can also be used for purposes of price discrimination. To this extent, it may be socially desirable to regulate the use of such arrangements if such regulation can be accomplished without unduly sacrificing their efficiency advantages.

Some parts of my account will be familiar to readers conversant with the economic literature on contracts or antitrust. The contribution of this Article lies not in creating new theory but in synthesizing theoretical analyses from a variety of literatures and specialized bodies of knowledge in order to develop a useful taxonomy of the considerations that are relevant in deciding whether to structure a contractual relationship in the form of an option. Organizing such insights into a more systematic conceptual framework helps us to integrate and synthesize disparate bodies of practical knowledge relating to various commercial and legal fields, including sales, information licensing, construction, financial instruments, and so on.

2004]   *The Option Element*   2191

Such a synthesis, as I have argued elsewhere, enables insights from one field to be translated and analogized for the purposes of critiquing and improving transactional planning in others.[2]

The organization of this Article will proceed as follows: Part I will demonstrate the importance of the analysis by surveying the range of legal problems that reduce to the question of option design. Part II will present a conceptual account of option design that demonstrates the basic relationship among the three fundamental elements of an option premium, exercise price, and the length of time that the option is open—and will explain how tradeoffs among these three elements can affect the efficiency of the underlying transaction. Part III will relate those tradeoffs to the overall incentive structure created by the contract; it will show how the choice among option premium, exercise price, and option length can influence the parties' incentives to perform, to take precautions against breach, to mitigate damages if necessary, and to invest in their relationship and in relation-specific assets. Part IV will discuss how the analysis presented in the prior two Parts relates to the legal questions that motivated the initial inquiry, as well as to some illustrative transactional problems.[3]

I. THE IMPORTANCE OF OPTION CONTRACTS

In this Part, I survey the range of legal problems for which an option-contract analysis is useful. Section A discusses the various black letter doctrines under which contracts nominally labeled as options are treated differently from contracts generally. Section B relates option-contract analysis to the general area of contractual

---

[2] See Avery Wiener Katz, An Economic Analysis of the Guaranty Contract, 66 U. Chi. L. Rev. 47, 114 (1999); Avery Wiener Katz, The Economics of Form and Substance in Contract Interpretation, 104 Colum. L. Rev. 496, 512 (2004) [hereinafter Katz, Economics of Form].

[3] This Article does not attempt to present a systematic account of option theory or options pricing; several good introductions to those subjects are already available for readers who wish to consult them. See, e.g., Richard A. Brealey & Stewart C. Myers, Principles of Corporate Finance (7th ed. 2003); John Cox & Mark Rubinstein, Options Markets (1985); Avinash K. Dixit & Robert S. Pindyck, Investment Under Uncertainty (1994). The focus of this Article is on the use of options to promote efficient contract design and to foster greater understanding of the principles of contract design among the lawyers and policymakers whose actions and decisions regulate contractual planning.

remedies. It discusses and critiques the familiar observation, commonly credited to Justice Holmes, that the default rule under which the remedy for breach of contract is limited to monetary damages in effect turns all contracts into option contracts, and then extends this observation to a variety of specific doctrinal rules in the remedial area. Finally, Section C shows how an option-based perspective raises interesting practical questions about business behavior and transactional planning, as well as about doctrinal applications.

A. Special Doctrinal Treatment of Option Contracts

1. Consideration and Mutuality

It is a basic principle of the common law that promises are generally not legally enforceable unless they are given in exchange for consideration—some payment, performance, or counterpromise that flows back to the promisor or his designee.[4] While the precise meaning and rationale of the consideration doctrine have long been debated, one commonly accepted component of the concept is the element of bargain—that is, promises should presumptively be enforceable if they are made as part of a deliberate and arm's-length economic exchange. On this conceptual account, most option contracts should qualify as bargains in that they are extended as part of an exchange process and operate as one of the critical terms and conditions of the exchange.

Traditional common law courts were nonetheless reluctant to view an important set of option contracts—those with zero premium—as full-fledged exchanges. They reasoned that the holder of such an option did not give anything to her grantor unless and until the option was exercised. For example, in *Wickham & Burton Coal Co. v. Farmers' Lumber Co.* the court refused to enforce an agreement under which a seller of coal agreed with one of its customers to furnish at a fixed price and over a fixed six-month term whatever amount of coal the buyer wished to purchase, on the grounds that the agreement neither required the buyer to pay to

---

[4] In order to avoid ambiguity when discussing generic option transactions, I adopt the convention in this Article of referring to offerors or grantors of options ("optionors") with male pronouns, offerees or grantees ("optionees") with female pronouns, and contracting parties generally with neutral pronouns.

keep this option open nor did it bind the buyer to any minimum amount.[5] In the court's view, the fact that the buyer retained the option to order no coal at all rendered any promise it made illusory and hence insufficient to provide consideration.[6] Similar results were reached in cases in which one party retained an option to terminate the contract at any time (since that party could if she wished escape all liability by exercising her option)[7] and in cases in which buyers promised to purchase their requirements or sellers promised to sell their entire output of a particular commodity (since the ostensibly bound party could choose to have no requirements or output during the relevant time frame).[8]

The evident fact that such promises arose out of arm's-length exchanges entered into for purposes of risk and incentive allocation, however, was not lost on courts and commentators. Many strove to find doctrinal limits on the optionee's discretion that would render her promise other than illusory. In the famous case of *Wood v. Lucy, Lady Duff Gordon*,[9] for instance, Judge Cardozo found consideration by reading into an exclusive agency contract an implied promise of best efforts on the part of the promisee. Similarly, other courts have implied a duty to exercise discretion in good faith, as in *Mattei v. Hopper*,[10] where the purchase of a parcel of commercial real estate was made subject to an independent broker obtaining leases satisfactory to the buyer. Still others have stretched to find consideration in contractual or statutory notice requirements, even where those requirements did not substantially burden the op-

---

[5] 179 N.W. 417 (Iowa 1920).

[6] Id. at 420.

[7] See, e.g., Miami Coca-Cola Bottling Co. v. Orange Crush Co., 296 F. 693, 694 (5th Cir. 1924) (holding that a licensing agreement granting the plaintiff the exclusive right to manufacture "Orange Crush" under defendant's trademark was void for lack of mutuality because the license contained a provision allowing the plaintiff to cancel at any time).

[8] See, e.g., Oscar Schlegel Mfg. Co. v. Peter Cooper's Glue Factory, 132 N.E. 148, 149 (N.Y. 1921) (holding that a contract whereby a glue factory agreed to sell to a jobber his glue requirements at a certain price without any consideration paid by the jobber and without any promise by him to take any glue lacked mutuality and was not enforceable).

[9] 118 N.E. 214, 215 (N.Y. 1917).

[10] 330 P.2d 625, 627 (Cal. 1958).

tionee's exercise of discretion.[11] On the statutory front, the Uniform Commercial Code ("U.C.C.") today implies a general duty of good faith in the performance of all contracts within its scope and imposes more specific duties of good faith to limit the discretion that can be exercised under a requirements or output contract,[12] a sales contract that leaves the price or other particulars of performance to be fixed by one of the parties,[13] or an acceleration clause.[14]

Notwithstanding these developments, the illusory promise rule remains part of black letter doctrine to this day. As Restatement (Second) of Contracts Section 77 provides: "A promise or apparent promise is not consideration if by its terms the promisor or purported promisor reserves a choice of alternative performances unless . . . each of the alternative performances would have been consideration if it alone had been bargained for."[15] There remains a legally enforceable mechanism, however, for offerors who want to bind themselves to a contract while leaving their offeree with full discretion whether to go forward with the deal. Although the formal device of nominal consideration has been generally discredited as a method of creating contractual obligation, Restatement (Second) of Contracts Section 87(1) authorizes it for option contracts in particular.[16] Similarly, in sales contracts, U.C.C. Section 2-205 allows such "firm offers" to be made by a signed record but limits

---

[11] See, e.g., Lindner v. Mid-Continent Petroleum Corp., 252 S.W.2d 631, 632 (Ark. 1952) (holding that an oil company's option to cancel the lease of a filling station operator upon ten days' notice did not qualify as an at-will termination provision, and that at the very least the lessee was bound to pay rent for ten days, which constituted sufficient consideration to support the finding of a contract between the lessor and lessee); Gurfein v. Werbelovsky, 118 A. 32, 33 (Conn. 1922) (holding that although a buyer retained the right to cancel an order before shipment, the seller's right to ship any time within three months, existing for however short a time after acceptance, carried with it the right to ship at the time of acceptance, and that this constituted sufficient consideration to establish a valid contract).

[12] U.C.C. § 2-306 (2004).

[13] Id. §§ 2-305(2), 2-311(1).

[14] Id. § 1-309.

[15] Restatement (Second) of Contracts § 77 (1981).

[16] Id. § 87(1) ("An offer is binding as an option contract if it (a) is in writing and signed by the offeror, recites a purported consideration for the making of the offer, and proposes an exchange on fair terms within a reasonable time; or (b) is made irrevocable by statute.").

their enforceability to a period not to exceed ninety days.[17] Thus, offerors who wish to bind themselves to zero premium options can do so provided that the option does not extend over too long a term and provided that they observe the requisite formalities.

From an economic viewpoint, the puzzling aspect of this doctrinal survey is neither that the courts held to formal accounts of consideration in the face of the commercial reality of the underlying deals, nor that it took so long to find a formal escape hatch. Rather, it is why parties frequently found it in their interest to enter into such arrangements, when by changing their deal slightly they could have had a legally enforceable contract. To illustrate more precisely, consider a firm offer that gives the offeree the option to purchase a promissory note for $500, with the option to be exercised within no more than one year. This option is a very close substitute for another option that sets the strike price at $450 rather than $500, but that also requires the offeree to pay $25 upfront to hold the option open. Assuming that there is approximately a 50% chance that the option comes "into the money" (that is, that it becomes worth exercising because the market price rises above the strike price), the $25 down payment compensates for the extra $50 of profit that the offeree will gain in the event that she exercises the option.[18]

An option that is traded for a positive premium was and is fully enforceable, subject to other possible defenses such as mistake or unconscionability. Why didn't parties who wanted to make firm of-

---

[17] U.C.C. § 2-205 (2004) ("An offer by a merchant to buy or sell goods in a signed record that by its terms gives assurance that it will be held open is not revocable, for lack of consideration, during the time stated or if no time is stated for a reasonable time, but in no event may the period of irrevocability exceed three months. Any such term of assurance in a form supplied by the offeree must be separately signed by the offeror.").

[18] Of course, the reduction in strike price slightly increases the chance that the option will be worth exercising, so the expected value of the option rises by slightly more than $25; the basic lesson is that the reduction in exercise price can be priced out upfront. Similarly, if the parties did not want to alter the exercise price of the option, they could increase its term, since the length of time that the option is open affects the likelihood that it will at some point come into the money. Thus, a zero premium option to purchase a promissory note within one year, for instance, would be a good substitute for a $25 option to purchase the same note at the same price within two years, if the time extension increased by an additional 50% the probability that the option comes into the money.

fers simply collect payment from their offerees, and in exchange for that payment sweeten the other terms of the exchange? One possible answer, of course, is that they wanted to make such offers but did not want to make them legally binding, but such an explanation does not account for subsequent parties' use of the formal devices established by U.C.C. Section 2-205 or Restatement Section 87(1).

2. Offer and Acceptance

As the phrase "firm offer" indicates, option contracts also implicate contract formation issues, as the question of revocability typically reduces to the question of whether an option has been granted. And as with consideration, doctrinal treatment of this question has evolved over the years. Case law and commentary prior to the first Restatement of Contracts suggested, consistent with a strictly formalistic view of consideration, that offers for unilateral contracts could be revoked at any time before performance was completed.[19] The first Restatement moderated this position, taking the view that an offeree who tenders or partially performs the requested performance is entitled to enforce the offeror's promise so long as she completes performance within a reasonable time.[20] The second Restatement, in contrast, advances this right of the offeree so that it begins at the onset of performance; it also explicitly labels the resulting obligation an option.[21]

---

[19] See, e.g., Petterson v. Pattberg, 161 N.E. 428, 430 (N.Y. 1928) (holding that defendant, who had promised to accept a cash payoff of a mortgage held on the plaintiff's realty, revoked the promise as the plaintiff stood on defendant's doorstep ready to tender cash in his pocket); I. Maurice Wormser, The True Conception of Unilateral Contracts, 26 Yale L.J. 136, 136–37 (1916) (presenting a notorious hypothetical in which the defendant, who has promised to pay plaintiff $100 for crossing the Brooklyn Bridge, withdraws the offer once plaintiff has walked halfway across).

[20] Restatement of Contracts § 45 (1932) ("If an offer for a unilateral contract is made, and *part of the consideration* requested in the offer is given or tendered by the offeree in response thereto, the offeror is bound by a contract, the duty of immediate performance of which is conditional on the full consideration being given or tendered within the time stated in the offer, or, if no time is stated therein, within a reasonable time." (emphasis added)).

[21] Restatement (Second) of Contracts § 45(1) (1981) ("Where an offer invites an offeree to accept by rendering a performance and does not invite a promissory acceptance, an option contract is created when the offeree *tenders or begins the invited performance or tenders a beginning of it*." (emphasis added)).

The irrevocability of an option contract, however, is not its only distinctive feature under the law of offer and acceptance. In contrast to ordinary offers, for instance, options remain open by default until their originally announced expiration date, even in the case of an explicit rejection or counteroffer by the offeree, on the theory that the optionee is entitled to enjoy the full speculative value that she was promised.[22] On the other hand, contract terms governing the time and manner of option exercise are applied strictly, in contrast to the more liberal treatment accorded to ordinary acceptances. As the comments to Restatement Section 25 explain, "any relaxation of terms would substantively extend the option contract to subject one party to greater obligations than he bargained for."[23]

Finally, the traditional distinction between unilateral and bilateral contract comes down to the existence of an option; specifically, a bilateral contract binds the offeree unconditionally to perform, while a unilateral contract leaves her the option to choose whether or not to perform. Legal commentators have generally recognized that the doctrinal distinction serves this functional role, but they have nonetheless vacillated over the years on the question of which category of obligation should be presumptively favored.[24] Contemporary doctrine muddies the waters by ostensibly making the legal characterization of the obligation turn on the parties' intentions, but then imposing presumptions that make whatever obligation exists begin at the earliest possible moment.[25]

---

[22] See id. § 37 ("[T]he power of acceptance under an option contract is not terminated by rejection or counter-offer, by revocation, or by death or incapacity of the offeror, unless the requirements are met for the discharge of a contractual duty."). For a critique of the Restatement rule, see Michael J. Cozzillio, The Option Contract: Irrevocable Not Irrejectable, 39 Cath. U. L. Rev. 491 (1990).

[23] Restatement (Second) of Contracts § 25 cmt. d (1981).

[24] Compare Restatement of Contracts § 31 (1932) (interpreting ambiguous offers as offers for bilateral contracts), with Restatement (Second) of Contracts § 32 (1981) (allowing an ambiguous offer to be accepted by either performance or a counterpromise as the offeree chooses); see also U.C.C. § 2-206(a) (2004) ("[A]n offer to make a contract shall be construed as inviting acceptance in any manner and by any medium reasonable in the circumstances."). The second Restatement even eschews the use of the terms "unilateral contract" and "bilateral contract," because of "doubt as to the utility of the distinction, often treated as fundamental, between the two types." Restatement (Second) of Contracts § 1 cmt. f (1981).

[25] Compare Restatement (Second) of Contracts § 62(1) (1981) (beginning of performance amounts to an acceptance by performance (hence a unilateral contract) in

Whether a given offer is interpreted as unilateral or bilateral, however, and whether the law should impose a presumption in favor of one or the other, cannot be answered without having some sense of why parties would use such arrangements and what transactional purposes they serve. Giving the offeree a binding option allows her to speculate at the offeror's expense and therefore only makes economic sense when her valuation of that speculative opportunity exceeds his. But the Restatement does not tell us when this is likely to be the case, and U.C.C. Article 2, true to its standard interpretative philosophy, tells us that the answer turns on an all-things-considered examination of the circumstances. If we are to cut through this thicket of doctrinal alternatives, we need a more structured account of when and when not to infer—or from the parties' viewpoint, to establish—an option contract.

3. Performance, Breach, and Damages

a. Anticipatory Repudiation

In ordinary contracts as interpreted under modern legal doctrine, promisors have a duty not to create unreasonable doubt about their contractual performance, both because certainty of performance is part of what the promisee has bargained for and because excessive doubt disrupts the promisee's ability to prepare for performance and to make appropriate reliance investments. Accordingly, a party who repudiates a contractual promise in advance of the time for performance, either by indicating an intention not to perform or by undertaking an act that renders performance infeasible, breaches the contract, thus discharging his counterparty's remaining duties and entitling her to bring an immediate action for damages.[26] Similarly, a party who does not go so far as to repudiate, but who creates sufficient uncertainty about his future performance, entitles his counterparty to demand adequate assurance of performance and, if reasonable, to suspend her own performance

---

cases where an offeror explicitly invites an offeree to choose between acceptance by performance or by promise) and id. § 62(2) (such an acceptance also operates as a promise to complete performance), with id. § 45(1) (where an offeror clearly requests acceptance by performance, the offeror becomes bound as soon as the offeree tenders or begins the invited performance but the offeree does not similarly become bound).

[26] See id. § 253. The analogous provision under sales law is U.C.C. § 2-610 (2004).

in the meantime.[27] In this case, the request for assurances also enables the insecure party to avoid wasteful reliance expenditures and to begin to mitigate damages in the event that assurances are not forthcoming.[28]

In the option-contract setting, however, the logical rationale of the anticipatory repudiation and adequate assurances doctrines does not apply. An optionee has no duty to go through with the deal; on the contrary, she has bargained for the right to speculate, which remains valuable to her until the last possible moment. Similarly, the grantor of the option has not bargained for certainty, but has instead agreed to bear whatever uncertainty results from the optionee's exercise of discretion. But the substantive justifications for these doctrines—the need to promote reliance investments in cases where the contract is going to be performed, and the need to encourage flexible and prompt mitigation in cases where it is not— remain whether or not the contract is denoted an option.

b. Duty to Mitigate

In the usual case, a party who suffers a breach of contract is expected to take reasonable actions to mitigate losses—that is, to seek out a substitute arrangement and to deduct the amounts so earned from the damages payable on the original contract. Any losses that could have been avoided by mitigation, furthermore, cannot be recovered as part of the damages payable by the breaching party. The mitigation principle is a fundamental principle of contract law, one that is commonly justified on diverse normative principles, including fairness (the aggrieved party owes a duty of cooperation and ought not run up losses even at the expense of a deliberate breacher), efficiency (it is wasteful to encourage the aggrieved party to run up losses for which she is the least-cost avoider), and corrective justice (any losses that could have been avoided by mitigation were caused by the aggrieved party and not by the breacher).

In contracts where the promisor has bargained for an option on the time or services of the promisee, however, the duty to mitigate

---

[27] See Restatement (Second) of Contracts § 251 (1981); U.C.C. § 2-609 (2004).

[28] See generally Richard Craswell, Insecurity, Repudiation, and Cure, 19 J. Legal Stud. 399 (1990).

does not apply.[29] For example, in a so-called "pay-or-play" contract, common in the entertainment industry, the promisor promises either to employ the promisee for a given job or to pay an equivalent salary.[30] In a "take-or-pay" contract, common in extractive industries such as coal or natural gas, the promisor promises either to take a given quantity of output or to pay for that quantity even if none is taken.[31] A promisor who makes such a promise and then chooses the payment option cannot insist that the promisee make other arrangements or deduct any amounts earned in the event she does. The formal argument in favor of such a result is that if the optionee exercises one of the options available to her under the contract, no breach has occurred. No breach means that no duty to mitigate on the part of either party ever arises.

Whether such a result makes sense in substantive terms depends, of course, on whether it promotes efficiency, fairness, or some other normative objective. There is some evidence that the doctrine accords with commercial practice; for instance, Professor Goldberg finds that the "pay-or-play" contracts entered into between Hollywood studios and movie producers or directors typically provide that there is no obligation for the artist to mitigate damages resulting from her removal from the project.[32] A full explanation, however, requires an account of why some contracting parties would do away with the duty to mitigate while others would not.

c. Liquidated Damages and Penalties

Structuring a contract as an option can also help the parties evade the penalty doctrine, which provides that parties may not contract for liquidated damages in an amount exceeding a fair estimate of the aggrieved party's lost expectation. Specifically, this can be done by providing for an unconditional payment that is characterized as the purchase of an option, which can subsequently be exercised by tendering an additional sum. (The take-or-pay con-

---

[29] See Victor P. Goldberg, Bloomer Girl Revisited or How to Frame an Unmade Picture, 1998 Wis. L. Rev. 1051, 1066–70.

[30] Id. at 1052, 1055.

[31] See Scott E. Masten & Keith J. Crocker, Efficient Adaptation in Long-Term Contracts: Take-or-Pay Provisions for Natural Gas, 75 Am. Econ. Rev. 1083 (1985).

[32] Goldberg, supra note 29, at 1070–72.

tract is an extreme example of this arrangement, in which the additional sum is zero.) If the purchaser subsequently decides not to go through with the transaction, she simply elects not to exercise the option. As a result, there is no breach, and hence no opportunity for the penalty doctrine to come into play.[33]

If the prepayment is characterized as a deposit or performance bond rather than the purchase of an option, in contrast, the law will limit the extent to which it can be made nonrefundable. For instance, in sales contracts, U.C.C. Section 2-718 provides that a breaching party is entitled to the return of any deposit she has made to the extent that the deposit exceeds the aggrieved party's compensable damages.[34] And in the common law setting, the breacher may have an analogous claim for restitution.[35]

Recent scholarship has called into question whether the penalty doctrine can be justified on economic grounds.[36] If these scholars are right, it is desirable that there should exist a device that allows the parties to evade the penalty doctrine. But if they are right, why don't contracting parties use this device all the time? And if they are not, why should legal doctrine provide this loophole?

B. Remedial Rules Generally

Even apart from the special treatment accorded to option contracts in particular, the entire law of contract remedies is usefully analyzed in terms of options because in the ordinary case breach of contract gives rise only to a claim of monetary damages, not to specific performance. As Justice Holmes famously remarked:

---

[33] See Nat'l Fuel Gas Distribution Corp. v. Pa. Pub. Util. Comm'n, 464 A.2d 546, 558 n.8 (Pa. Commw. Ct. 1983) (declining to apply the penalty doctrine to take-or-pay contract); accord Lake River Corp. v. Carborundum Co., 769 F.2d 1284, 1292 (7th Cir. 1985) (approving in dicta the immunity from the penalty doctrine enjoyed by take-or-pay clauses).

[34] U.C.C. § 2-718 (2004).

[35] The restitution argument would be that the forfeited deposit results in unjust enrichment because it puts the aggrieved party in a better position than contractual performance would have done. But see Lawrence v. Miller, 86 N.Y. 131, 139–40 (1881); Maxton Builders, Inc. v. lo Galbo, 493 N.Y.S.2d 825 (N.Y. App. Div. 1985), aff'd, 502 N.E.2d 184, 189 (N.Y. 1986) (involving earnest money deposits in real estate transactions treated as presumptively nonrefundable to the extent that they complied with trade usage).

[36] For a survey of arguments, see Aaron S. Edlin & Alan Schwartz, Optimal Penalties in Contracts, 78 Chi.-Kent L. Rev. 33 (2003).

> Nowhere is the confusion between legal and moral ideas more
> manifest than in the law of contract. . . . The duty to keep a con-
> tract at common law means a prediction that you must pay dam-
> ages if you do not keep it—and nothing else. If you commit a
> tort, you are liable to pay a compensatory sum. If you commit a
> contract, you are liable to pay a compensatory sum unless the
> promised event comes to pass, and that is all the difference. But
> such a mode of looking at the matter stinks in the nostrils of
> those who think it advantageous to get as much ethics into the
> law as they can.[37]

In option terminology, we can restate Holmes's point by saying
that the promisor holds a call option to buy her way out of the con-
tract by paying a strike price equal to the value of court-awarded
damages.

Holmes's option theory of contracts has been widely criticized
over the years by commentators who have argued that breaking a
promise is not merely a social inconvenience, but a deontological
wrong.[38] Some of these critics have also attacked the modern-day
economic elaboration of Holmes's theory: the so-called theory of
efficient breach. But much of this criticism misses the mark, be-
cause not even the most uncompromising deontologist would ob-
ject if the parties were explicitly to write Holmes's call option into
their contract. If they did, the party calling the option would not be
breaking any promise; rather, she would simply be exercising a
right freely and consensually granted to her by the promisee.
Whether breach of contract can be viewed as a wrong, accordingly,
depends on whether the parties have contracted with the intention
of incorporating such an option into their agreement.

As Professor Shavell has observed, the legal regime merely sets
up a default term for contracting when it establishes a rule of dam-
ages.[39] Under a regime where specific performance is the nominal
default, as in some of the civil law systems of Europe, the parties
can contract around this default rule by writing a payment option

---

[37] O.W. Holmes, The Path of the Law, 10 Harv. L. Rev. 457, 462 (1897).

[38] See, e.g., Charles Fried, Contract as Promise: A Theory of Contractual Obligation 16 (1981); Daniel Friedmann, The Efficient Breach Fallacy, 18 J. Legal Stud. 1 (1989).

[39] Steven Shavell, Contracts, *in* 1 The New Palgrave Dictionary of Economics and the Law 436, 438 (Peter Newman ed., 1998).

into their contract. Under the common law rule that makes expectation damages the default, parties can in principle escape the default by writing in a different damages term or even a term calling for specific enforcement. A deontological critic, however, should have no reason for preferring one default rule to another. Absent some substantive account of human flourishing that explains which kind of agreements human beings should properly enter into and that justifies interfering with their individual decisions in this regard, respect for the parties' autonomy requires only that their voluntary agreements should be enforced.[40] Accordingly, the aspect of common law doctrine most vulnerable to attack on autonomy grounds is not the default rule of expectation damages, but rather those doctrines that impinge on the parties' freedom to contract around this default rule by arranging for liquidated damages or specific performance as they freely choose.[41] Indeed, an efficiency-based default rule arguably promotes party autonomy by adopting a rule of interpretation that is likely to correspond to the way in which most contracting parties actually would want their agreement to be enforced.[42]

---

[40] See Richard Craswell, Contract Law, Default Rules, and the Philosophy of Promising, 88 Mich. L. Rev. 489, 528 (1989) (elaborating this defense of efficiency analysis against the objections of deontological critics).

[41] Recent statutory innovations under the newly amended Article 2 of the U.C.C., if adopted by the states, may have the effect of expanding the parties' contractual freedom in this way. The new version of § 2-710 extends to sellers the option, previously available only to buyers, of collecting consequential damages in appropriate cases; new § 2-716 directs courts to consider the parties' ex ante consent to specific performance as a factor in the decision whether to issue such an award; and new § 2-718 eliminates the requirements of difficulty of proving loss and inconvenience of obtaining an adequate remedy for commercial parties seeking to enforce liquidated damages clauses. U.C.C. §§ 2-710, 2-716, 2-718 (2004).

[42] Professor Shavell writes:

> This understanding of damage measures as a device to induce the behaviour that the parties would have specified in more complete contracts sheds light on the notion, held by many legal commentators, that contract breach is immoral, as it constitutes the breaking of a promise. That belief is often incorrect, it is submitted, and might fairly be considered to be the opposite of the truth. The view that a contract breach is the breaking of a promise overlooks the point that the contract that is breached is generally an incomplete contract, and that the breach is what the parties *want* and would have specified in a complete contract. In the example of the simple incomplete contract calling for a desk to be produced, the seller who finds that his production cost would be [substantially higher than the buyer's losses from non-delivery] will commit breach under the expectation measure. But in so doing, he will be acting precisely as would have

From an options perspective, then, the effect of remedial doctrine is simply to set the terms of the promisor's inherent call option. A legal regime or individual agreement that awards higher or more extensive damages will raise the option's strike price, making it more costly for the promisor to exercise the option and less likely that she will do so. As with all options traded at arm's length, however, the change in the strike price will affect the ex ante exchange value of the option and hence the price at which the parties are willing to enter into the contract initially.[43] Changes in the damage rule, accordingly, result in changes in the structure of the call option embedded in every contract, and in the risk and incentive properties associated with it.

The same is true of the rules that set the times when a breaching party is deemed to have breached and when damages are to be measured. The doctrine of anticipatory repudiation, for instance, shortens the term of the promisor's option to breach and pay damages by forcing her in certain circumstances to exercise that option well before the scheduled time of performance. Similarly, the time at which the promisee's damages are measured in repudiation cases determines whether or not she gets an option to speculate against the contract. If damages are measured at the time of repudiation, as they are under U.C.C. Section 2-713 when a seller repudiates an obligation to deliver goods,[44] the parties' positions are closed out immediately and there is no further speculative opportunity for either party. If damages are not measured until the time of performance, however, as they are under U.C.C. Section 2-708 when a buyer repudiates an obligation to purchase,[45] the aggrieved seller receives an option to enforce the contract if prices move against her, but can forego enforcement in the event that prices move in her favor. The positive value of this option leads to an associated increase in the seller's expected damages from breach.[46]

---

been set out in a complete contract, and it is that contract which is best regarded as the promise between the parties that ought to be kept.

Shavell, supra note 39, at 439.

[43] See generally Paul G. Mahoney, Contract Remedies and Options Pricing, 24 J. Legal Stud. 139 (1995).

[44] U.C.C. § 2-713 (2004).

[45] Id. § 2-708.

[46] See Thomas H. Jackson, "Anticipatory Repudiation" and the Temporal Element of Contract Law: An Economic Inquiry into Contract Damages in Cases of Prospec-

Finally, a similar interpretation can be given to the various doctrines that regulate the so-called "self-help" remedies of contract law—rejection, revocation, the demand for adequate assurance, suspension of one's performance, and rescission—because such remedies are all optional on the aggrieved party's part. These rules affect the value of the promisor's option just as the rules governing monetary remedies do. For instance, the sales law doctrine of acceptance, which limits the period of time during which a buyer is entitled to reject non-conforming goods, reduces the value of her option to reject. This limitation thus affects both the seller's incentive to perform and the ex ante price of the goods that are traded.

II. A THEORETICAL ACCOUNT OF OPTION DESIGN

The preceding discussion may have persuaded the reader that the option element in contracting is important and even pervasive, but it raises more questions than it answers. In order to begin to address these questions, this Part develops a simple model of the option contract that shows how the basic aspects of an option relate to and can be traded off against each other.

A. Three Essential Terms: Option Premium, Strike Price, and Option Life

The model is simple, and I have already hinted at its basic intuition in my discussion of some of the foregoing examples. Specifically, I model the idealized option contract as consisting of three main terms: the option premium (denoted as P), which is the unconditional amount that the optionee must pay upfront in order to acquire the right to exercise the option; the exercise or strike price (denoted as S), which is the amount that the optionee must pay conditionally in the event that she chooses to exercise her rights,

tive Nonperformance, 31 Stan. L. Rev. 69, 103–04 (1978) (explaining the relationship between time of measuring damages and the expected value of damages and criticizing the rule of U.C.C. § 2-708 on the grounds that it gives the aggrieved seller an unbargained and inefficient speculative opportunity). Revisions to U.C.C. § 2-708, recently promulgated by the American Law Institute and the National Conference of Commissioners on Uniform State Law but not yet adopted as the law of any state, would eliminate the current asymmetric treatment of buyers and sellers in this regard, measuring damages at repudiation for either party, in accordance with Jackson's argument. See U.C.C. § 2-708(1)(b) cmt. 4 (2004).

and the option life (denoted as T), which is the period of time for which the option remains open. A firm offer extended under U.C.C. Section 2-205, for instance, would have a P equal to zero, a T equal to ninety days, and an S equal to whatever price at which the offeror stood ready to sell.

Similarly, a prepaid airline reservation that requires the passenger to charge a $400 ticket to her credit card at the time the reservation is issued, but that allows her the option to reschedule her trip to a later date with a change penalty of $100, gives the passenger a call option on the reserved seat with P=$100, S=$300, and T equal to the time between the making of the reservation and the latest time that the airline is willing to credit the remaining value of the ticket to another flight. The $100 payment is the unconditional premium, P, because the passenger must pay it whether she exercises the option or not; and $300 is the strike price, S, because if she actually decides to take this flight, she loses the opportunity to apply the $300 credit to a later trip. (This calculation assumes that the value of applying the $300 to a later trip is actually worth $300, which it would be to a passenger who travels frequently and will use the credit in ordinary course. If the value of the credit is only $250, for instance, because there is a one in six chance that she will never rebook on this particular airline, then the option premium, P, equals $150 and the strike price, S, equals $250.[47])

As the airline example illustrates, the designation of P as the "up-front" premium is intended merely to indicate that its payment is unconditional. It is possible for P to be financed on seller credit, so long as the seller has a reliable way to collect in the event that the optionee chooses to exit. Credit cards are often used as such a collection mechanism. In the case of vacation deposits described

---

[47] Alternatively, we could characterize the ticket not as granting a call option on the right to keep this particular reservation, but as an outright sale of the trip, bundled with a put option in which the passenger holds the right to force the airline to buy back the reservation in exchange for a credit to be used on another flight. In this case we would have to determine what fraction of the payment is properly allocated to the sale of the ticket and what fraction is properly allocated to the sale of the put option. Such a determination would require us to know the going price for an unconditional sale of a seat, a fact that the hypothetical description in the text does not disclose. The two characterizations are equivalent in economic terms; this is an instance of a feature of options that financial economists label "put-call parity." See, e.g., Michael S. Knoll, Put-Call Parity and the Law, 24 Cardozo L. Rev. 61, 72 (2002). Accordingly, we are free to choose one characterization or the other at our convenience.

below,[48] for instance, many innkeepers will agree to wait to submit a charge against the customer's account until the actual date of the reservation if they are given the contractual right to make the charge in the event that the customer cancels her reservation. Similarly, a seller could collect a larger deposit than she plans to keep, refunding part of the excess following cancellation. The actual timing of payment is a secondary complication that I abstract from in this discussion; the variables P, S, and T, accordingly, should be interpreted in terms of the parties' actual opportunity costs from entering into and exercising rights under an option contract, rather than the payments that may happen to flow between the parties at any particular point in time.[49]

B. The Relationship Among Option Premium, Strike Price, and Option Life

In a competitive market, or indeed in any arm's-length exchange, these three terms of the option contract will stand in a regular relationship with one another. Specifically, the option premium, P, and the option life, T, will vary directly with each other, so that an increase in T will be associated with an increase in P. Similarly, the strike price, S, and the option life, T, will vary directly, so that an increase in T will be associated with an increase in S. The option premium, P, and the strike price, S, however, will vary inversely—an increase in S will lead to a decrease in P, and vice versa.

These relationships among P, S, and T follow directly from a calculation of the economic value and economic cost of an option. More specifically, the expected value of an option to its holder, which I denote as E(V), depends on the possible distribution of values, V, that the underlying good or service might take on, and

---

[48] See infra Section IV.B.1.

[49] In addition, the framework is general enough to encompass contracts that we would in common parlance describe as unconditional. An idealized unconditional sale of a widget for one dollar that is specifically performable in equity, for instance, can be interpreted as carrying along with it an exit option with T=0 (the buyer has no time after entering into the sale to change her mind) or S=∞ (the price of refusing to buy, and hence resisting the enforcement power of the chancellor, is infinite). In practice, even unconditional contracts are less than fully enforceable, so that S may be high but not infinite, or T may be very small but not zero.

on the probabilities associated with each of those possible values. If the underlying good turns out to be worth more than the strike price, S, at any time during the period that the option is open, the optionee will want to exercise the option and will in this case receive net gains V-S. (It is conventional to refer to this situation by saying that the option is "in the money.") If the underlying good turns out to be worth less than the strike price, S, for the entire option term, the optionee will prefer to let the option expire, and will receive ex post net gains of zero. The total expected value of the option thus equals the discounted value, averaged over all possible realizations of V and over the time horizon, T, of the greater of V-S and zero.[50] To take an extreme example, on the one hand, an option that one knows one will never exercise is worthless. On the other hand, an option that one might want to exercise always has positive value, even if the probability of exercise is very low.

Calculating the expected cost of an option to the optionor is slightly more complicated. Denote the value that the optionor attaches to the underlying good as W. This value, W, need not be identical to the optionee's value, V, although in general the two parties' values will be positively correlated in the statistical sense for most ordinary commodities. In all cases where the optionee elects to exercise the option (that is, in all cases where V>S), the optionor will lose an amount equal to W-S (that is, he will lose his own use-value, W, but will receive the strike price, S, as compensation). Note that it is possible for the optionee's decision to exercise the option to make both parties better off, if S is sufficiently high or W is sufficiently low. In all cases where the optionee elects to pass on the option, the optionor will lose zero. Thus the total expected cost of the option, denoted as E(W), equals the discounted value, over the time horizon, T, and over all possible realizations of V where V>S, of the greater of W-S and zero.

In any arm's-length exchange, and certainly in a competitive market, the premium, P, paid for an option should be positively related to both expected value and expected cost, and in general will lie in a range bounded by the two if they diverge. But expected

---

[50] Risk-averse parties will discount these possible outcomes in proportion to their distaste for variation in their incomes. I am ignoring this complication for the present discussion.

value will vary directly with the option term, T, and inversely with the strike price, S. Decreases in S will raise E(V) by making the exercise of the option more attractive to the optionee for any particular realization of V, and by making it more likely in marginal cases that V will exceed S so that the option will be in the money. Increases in T will raise E(V) by making it more likely that at some point during the option term, the optionee's valuation, V, will actually exceed the strike price, S.[51] The optionor can therefore keep the option equally attractive to the optionee by adjusting P, S, and T according to this relationship.

The relationship between E(W), S, and T is more complicated because of the possibility that the optionor's valuation, W, can fall below S, so that the optionor prefers ex post for the optionee to exercise the option.[52] This analysis is simplified, however, if we observe that the parties will never want to choose an S in the range where E(W) varies directly with S, because by reducing S and raising P, both optionor and optionee can be made better off. (The increase in P and decrease in S necessarily make the optionor better off, and given the aforementioned relationship between S and E(V), for any decrease in S we can always find some increased value of P that leaves the optionee no worse off than under the previous contract, and possibly better off.) Thus, in the range of

---

[51] More formally: For any given random process that generates values of V over time, there is some probability, p_0, that the option comes into the money (that is, that V>S) at some point between now and a given subsequent time, t_0. Suppose that this does not happen—an event that occurs with probability 1-p_0. Then if the option life is extended to a later time, t_0+t_1, there is an additional probability, p_1, that the option comes into the money at some time between t_0 and t_1. Accordingly, the total chance that the option is in the money at some point between t_0 and t_1 equals $p_0+p_1(1-p_0)$.

[52] Consider first the relationship between E(W) and S. A decrease in S has two possibly offsetting effects on E(V). First, it will raise the optionor's cost, W-S, for any particular realization of V. This "price effect" clearly increases the overall expected cost, E(V). The decrease in S, however, will also increase the chance that the optionee wishes to exercise (the "quantity effect"). If, in those marginal cases affected by the decrease, the value of W-S is positive so that the optionor prefers the optionee not to exercise, then the quantity effect also increases the expected cost of the option. If, however, the value of W-S is negative in those marginal cases affected by the decrease, then the quantity effect offsets the price effect and could in some cases lead to an overall reduction in E(W). Note that the relationship between S and E(W) is more likely to be positive when S is high, because then the option will rarely be in the money, and the quantity effect will dominate the price effect. When S is low, in contrast, the option will usually be in the money, and the price effect will dominate the quantity effect.

values among which the parties would actually want to choose, $E(W)$ and S must be inversely related. Similarly, while there may be values of the option life, T, such that T varies inversely with $E(W)$,[53] the parties will never want to choose a T in that range of values, because by raising both T and P simultaneously, both optionor and optionee can be made better off. Thus, in the range of values among which the parties would actually want to choose, $E(W)$ and T will be directly related.

Finally, while I do not explicitly incorporate such factors into my analysis below, it is worth noting that exogenous market conditions and other aspects of the underlying exchange—such as the market rate of interest and the level of price volatility associated with the underlying commodity—will also affect the relationship among P, S, and T. In contracts where the payment of the premium is delayed until the optionee decides whether to exercise, a relatively high interest rate will benefit the optionee and cost the optionor by increasing the value of float that is extended. To compensate, the parties will have to increase the strike price, S, for any given level of premium, P. Similarly, a commodity that is subject to significant price fluctuations is more likely, given any particular values of S and T, to wind up in the money, and so an option to trade that commodity must, in an arm's-length exchange, be associated with a higher option premium, P. This is why we do not typically observe zero premium firm offers extended in ordinary commodities or securities markets, as the market value of such offers is strictly positive due to the expected variation in future price. In the discussion below, I abstract from these concerns for reasons of expositional simplicity; but as a general matter, the reader should understand that increases in interest rates or market volatility will have a qualitatively similar effect to increases in the time horizon, T.[54]

In what follows, I also abstract from the further complication that in bilateral contracts where both parties have promised to per-

---

[53] This would be the case if an extension of the option term, T, led to an increase in the chance that the optionee would exercise, and the optionor wanted to increase this chance because his valuation, W, was likely to be less than the strike price, S, for this range of T.

[54] A longer time horizon means that the discounted value of receiving funds at the expiration of the option period is reduced; this is equivalent to an increase in the interest rate. Similarly, with a longer time horizon there will be more opportunity for prices to fluctuate; this is operationally equivalent to an increase in volatility.

form, there are actually two options at work in that the buyer can escape the contract by paying the seller's expectation damages and the seller can also escape the contract by paying the buyer's expectation damages.[55] To consider symmetric options would needlessly complicate the exposition of my argument, however, and once its basic intuition is understood, the reader should be able to extend its logic to the bilateral case in straightforward if tedious fashion.

C. Efficient Option Design

Given this functional relationship among P, S, and T, there exists a continuous range of possible option contracts among which the parties might choose, ranging from those with low P and high S to those with high P and low S, and from those with low P and T to those with high P and T. Similarly, the parties can choose between nominally unconditional contracts (those with $T=0$ and $S=\infty$, and some value of P) and a continuum of more or less equivalent option contracts (those with $T>0$ and $S<\infty$, and a somewhat higher value of P). How, then, can and should the parties choose among these possible contracts?

Phrased in the most general terms, the answer stems from the fact that while both the optionee's expected value $E(V)$ and the optionor's expected cost $E(W)$ vary directly with the option life, T, and inversely to the strike price, S, they may vary at different rates. If so, the parties will differ in their willingness to trade one of these terms off against another, and there will be mutual profit in a contract that adjusts terms efficiently between them. For instance, suppose that extending the life of the option by six months would increase the optionee's expected value by $500 and would increase the optionor's expected cost by only $200. In that case it would be efficient to extend the option life, and we would expect that in an arm's-length exchange the parties would have every incentive to do so.

Such an abstract answer, however, does not explain why the parties' willingness to trade off option premium, option life, and strike

---

[55] Indeed, in cases where a contracting party decides to exit the deal, his counterparty's economic losses will include the lost value of her own exit option. See Alexander J. Triantis & George G. Triantis, Timing Problems in Contract Breach Decisions, 41 J.L. & Econ. 163, 166 (1998).

price might differ in actual transactions. Accordingly, the following Part of this Article grounds such tradeoffs in the overall incentive structure created by the contract. As we will see, the reason why one party might place relatively high value on a particular element of an option turns on a variety of transactional factors, including incentives to perform and to take precautions against breach, the value of relation-specific investments, and the distribution of information between the parties.

III. DETERMINANTS OF COMPARATIVE ADVANTAGE IN OPTIONS TRADING

For purposes of organizing the analysis of this Part, it is useful to divide the various reasons why parties might wish to choose one option over another into three categories: efficiency considerations, nonefficiency considerations, and mixed considerations. Such a taxonomy is necessary if we are interested in investigating the appropriate legal treatment of option contracts, because if the parties' choices are motivated primarily by efficiency considerations, then absent some important countervailing public value, the law should defer to those choices and enforce the contract as written. If option design is instead driven by nonefficiency considerations such as a seller's desire to gain market power or to foreclose competition, nonenforcement and perhaps affirmative regulation of contracting behavior is in order. We start with nonefficiency considerations, not because they are necessarily more likely to motivate such transactions, but because they are somewhat more straightforward to explain and analyze.

A. Nonefficiency Considerations

1. Bounded Rationality

The psychological literature on economic behavior has shown that individual human beings engaged in economic exchange do not in fact always act as the idealized maximizers of neoclassical economic theory but in many instances follow heuristic rules of thumb when making decisions or are subject to imperfect cognitive

procedures when processing information.[56] As a result, choices be-
tween economic alternatives can be sensitive to the way in which
those choices are framed.

Such factors could influence parties' choices between the three
variables, P, S, and T, if they affect people's perceptions of or atti-
tudes toward those variables disproportionately. For example, one
fairly robust finding of the behavioral economics literature is that
people tend to give more decisional weight to factors that are more
salient or memorable to them. It is plausible that for some option-
ees, the up-front payment, P, may be more salient than the subse-
quent exercise price, S, especially if the initial payment must be
paid out of pocket rather than financed. If so, parties designing the
option will have an incentive to back load the total payments under
the option, lowering P and raising S past the point that would oth-
erwise be transactionally efficient. A similar effect would be pro-
duced if people's inability to resist immediate temptation led them
to discount future costs and benefits at a rate that exceeded their
long-run relative value—a phenomenon sometimes labeled "hy-
perbolic discounting."[57]

Of course, in order for this factor to distort the choice among P,
S, and T, the two parties to the exchange would have to be un-
equally subject to it, because just as it is especially salient for the
optionee to have to make an unconditional payment of P and espe-
cially tempting to put off the pain by back loading part of the total
cost into the subsequent strike price, S, it is especially salient and
tempting for the optionor to receive that payment, P, rather than
waiting until some future time at which he might or might not re-
ceive S. In cases where one party is more sophisticated than the
other or more able to invoke procedures that rationalize its deci-
sionmaking, however, we would expect the sophisticated actor to

---

[56] See generally Daniel Kahneman & Amos Tversky, Judgment Under Uncertainty:
Heuristics and Biases, *in* Judgment Under Uncertainty: Heuristics and Biases 3, 18–20
(Daniel Kahneman et al. eds., 1982); Christine Jolls et al., A Behavioral Approach to
Law and Economics, 50 Stan. L. Rev. 1471, 1476–77 (1998).

[57] See David Laibson, Golden Eggs and Hyperbolic Discounting, 112 Q.J. Econ. 443,
445–46 (1997); Richard H. Thaler & George Loewenstein, Intertemporal Choice, *in*
Richard H. Thaler, The Winner's Curse: Paradoxes and Anomalies of Economic Life
92, 96–98 (1992). A less formal account of the phenomenon can be found in Jon El-
ster, Ulysses and the Sirens: Studies in Rationality and Irrationality 71 (1979).

design the option so as to delay its benefits under the option while accelerating the benefits of its less sophisticated counterparty.

A similar phenomenon might occur not because of bounded rationality as such, but because of imperfect information. A buyer offered an option contract might have better information about the up-front payment, P, than about the strike price, S, (for example, hotel and airline rates are typically published in generally available schedules or tourist books, while their cancellation policies are less widely advertised). A traveler who hears about the price of a particular room or flight, but who lacks information about the cancellation penalty, will reasonably assume that this seller's cancellation policy is the same as that of most other sellers and make a purchase decision on that basis. The seller's incentive under such an informational structure, however, is to debase the latent terms and improve the patent terms of the exchange as much as possible.[58] In such cases, there is a colorable case for legal rules, such as the illusory promise rule and the limits on liquidated damages, that push parties toward providing more unconditional compensation in the form of P, and less conditional compensation in the form of S—but only if we think that regulators have the ability to choose the tradeoffs more efficiently than an unregulated market.

2. Market Exclusion

The bounded rationality/imperfect information story sketched in the previous Section explains why parties might disproportionately back load compensation by reducing the option premium, P, while inflating the strike price, S. It is also possible, however, that they would wish to do the reverse and front load the payment for anticompetitive purposes. Such a strategy is effective because an optionee who has already paid most of the price of a good up-front as an option premium, and who needs only to pay a small additional balance in order to get the good, will be much less likely to switch to a competitor who enters the market after the option premium has been paid. Setting up exchanges in the form of prepaid options,

---

[58] See Avery Katz, The Strategic Structure of Offer and Acceptance: Game Theory and the Law of Contract Formation, 89 Mich. L. Rev. 215, 288–89 (1990); Avery Katz, Your Terms or Mine? The Duty to Read the Fine Print in Contracts, 21 RAND J. Econ. 518, 525–28 (1990).

or more generally, front loading payment to a greater extent than would otherwise be profitable, may be a way of excluding competitors from the market. Several economic commentators have shown how liquidated damages clauses can be used in this way,[59] and the traditional non-Chicago antitrust account of resale price maintenance is also based in part on this concern.[60] Recent work on consumer "lock-in" in imperfectly competitive markets has also shown that this could be a more general phenomenon.[61]

In principle, the anticompetitive pricing of options contracts could justify some form of regulation under the antitrust laws. Whether this possibility justifies any particular legal policy or presumption in the area of contract doctrine is less clear. Courts deciding contracts cases are unlikely to possess information sufficient to judge whether anticompetitive effects on third parties outweigh the benefits of option arrangements to the contracting parties. Moreover, such third parties would lack standing to challenge such terms in the contract law setting. As the recent literature on liquidated damage clauses has shown, the individual consumer who enters into a lock-in arrangement benefits from the lock-in, since in order to induce the consumer's agreement, the seller with market power will in general share some of the gains from exclusion with the consumer.[62] Absent the emergence of an actual competitor with whom the consumer wishes to deal or some other independent reason she wishes to exit from the relationship, the consumer will have little incentive to challenge the anticompetitive clause.

---

[59] See Philippe Aghion & Patrick Bolton, Contracts as a Barrier to Entry, 77 Am. Econ. Rev. 388, 388–92 (1987); Tai-Yeong Chung, On the Social Optimality of Liquidated Damage Clauses: An Economic Analysis, 8 J.L. Econ. & Org. 280, 283–89 (1992); Kathryn E. Spier & Michael D. Whinston, On the Efficiency of Privately Stipulated Damages for Breach of Contract: Entry Barriers, Reliance, and Renegotiation, 26 RAND J. Econ. 180, 183–86, 195–96 (1995).

[60] See, e.g., 8 Philip E. Areeda & Herbert Hovenkamp, Antitrust Law: An Analysis of Antitrust Principles ¶ 1604 (2d ed. 2004); Robert Pitofsky, In Defense of Discounters: The No-Frills Case for a *Per Se* Rule Against Vertical Price Fixing, 71 Geo. L.J. 1492–93 (1983).

[61] See, e.g., Paul Klemperer, Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade, 62 Rev. Econ. Stud. 515, 517–18, 520–23, 535–36 (1995).

[62] Aghion & Bolton, supra note 59, at 388–92.

3. Regulatory Arbitrage

Finally, contracting parties may alter the structure of their option contracts in order to take advantage of arbitrage possibilities created by other regulatory regimes to which they are subject. A prominent illustration is provided by the widespread practice of tax planning. Under the Internal Revenue Code, similar transactions can be taxed differently depending on whether they are structured as options or unconditional obligations.[63] Much creative work in the tax area in recent years, both by practitioners and by scholars, has focused on the arbitrage possibilities left open by tax law, and how best to deal with those possibilities. Similarly, disparate accounting treatment of options may lead parties wishing to improve the appearance of their financial condition to use the option form to shift funds between current and future income and expenses.[64] A third example is provided by the privileged treatment of leases under U.C.C. Article 9 and the federal bankruptcy code. A disproportionate amount of litigation has occurred over the issue of whether an equipment lease with an option to purchase or renew at a favorable price at the end of the lease term amounts in substance to either a sale, subject to U.C.C. Article 9's filing requirements and to the automatic stay imposed when the lessee files for bankruptcy, or a "true lease" that is outside the scope of Article 9 and free from any claim of a bankruptcy trustee.[65] This Article does not attempt to provide any analysis of such issues, except to observe that they provide a common reason for structuring options in one way or another and that they can in general lead to option contracts that are designed inefficiently from the viewpoint of maximizing social

---

[63] For instance, parties to an option contract do not generally realize gain or loss upon the receipt or payment of an option premium, instead deferring realization until the optionee's contract rights have either been exercised or expire. The analogous transaction set up as an outright sale, in contrast, is treated as a realization event. See generally Bruce Kayle, Realization Without Taxation? The Not-So-Clear Reflection of Income From an Option to Acquire Property, 48 Tax L. Rev. 233 (1993).

[64] Similarly, disparate accounting treatment of options may lead to parties wishing to improve their apparent financial condition to use the option form to shift funds between current income and assets, or between expenses and liabilities. See David M. Schizer, Executives and Hedging: The Fragile Legal Foundation of Incentive Compatibility, 100 Colum. L. Rev. 440, 454 (2000).

[65] See, e.g., In re Marhoefer Packing Co., 674 F.2d 1139 (7th Cir. 1982).

value or even of maximizing the non-arbitrage gains from exchange.

B. Efficiency Considerations

What, then, are the efficiency reasons for parties to trade off option premium, strike price, and option term in a particular way? There are several possible reasons that I classify in two categories: risk allocation and incentives.

1. Differential Beliefs About the Future or Differential Risk Aversion

Most simply and straightforwardly, parties often engage in speculative exchange when they have different beliefs about what the future will hold. Trading options and other financial instruments can then be a way of betting on the future and hedging against the risk of other events. This is, indeed, the main motivation for the market in purely financial options, where parties bet on price movements of commodities and other underlying fundamental assets. Most standard accounts of finance markets view such trading as a socially desirable mechanism for sharing information and minimizing risk through financial diversification. To the extent that option contracts are motivated by such speculative concerns, a similar normative analysis will apply.

Similarly, the design and pricing of options may be motivated by differences in risk aversion. When an option is sold, the seller receives an unconditional payment and the buyer receives a risky asset that has positive value if her private valuation, V, exceeds the strike price, S, and zero value otherwise. Under many circumstances, the parties may differ in their willingness to hold this particular risk; for instance, the optionor may be relatively immune to risk because of the size of his portfolio or the frequency with which he trades, so that it is profitable for him to insure the optionee. Such a rationale applies in both purely financial settings and the context of real contractual exchange alike, and it may explain why large sellers are willing to offer requirements contracts and large buyers such as grain elevators are willing to offer output contracts.

To the extent that risk aversion and the value of insurance vary over time and over particular ranges of potential asset values, they can explain not only why parties enter into options, but why they

might trade off option premium, strike price, and term in a particular way. For example, a seller may be willing to insure the risk that the buyer's valuation for the good being exchanged falls below some very low level (that is, to the point where the purpose of the contract would be frustrated) because this risk is low and is likely to be buyer-specific. He would not be willing to insure the risk that the buyer's valuation falls only moderately, however, because this risk has greater probability and is more likely to be the result of general market shifts. Optimal risk allocation would then require the seller to bear the former portion of risk but not the latter.[66]

2. Ex Post Incentives

Risk and speculative motives help explain the existence of options trading in both financial and nonfinancial markets. More interesting from the viewpoint of contract theory, however, are a set of factors that go to the basic incentives of the parties to undertake costly efforts that will increase the expected surplus from their exchange. Such factors are not of prime importance in financial options, where the parties are small relative to the market and are typically not in a position to affect the relevant risks, but they are fundamental in the ordinary contractual setting. In this regard, options can affect the parties' incentives to perform, to take precautions against breach, or to make reliance investments in their relationship.

Specifically, an option separates the incentives to manage upside and downside risk and splits those incentives between the contracting parties. Buying an option gives the buyer incentives to invest in actions that increase upside risk or that are valuable in the event that such risk materializes because in that instance the buyer will want to exercise the option, at which point the underlying asset and all the costs and benefits associated with it will belong to her. It is on this precise logic that executive stock options are claimed to im-

---

[66] Cf. Richard A. Posner & Andrew M. Rosenfield, Impossibility and Related Doctrines in Contract Law: An Economic Analysis, 6 J. Legal Stud. 83 (1977) (using similar risk aversion story to explain impracticability and frustration doctrines of contract law).

prove the incentives of corporate officers to act in ways that in-
crease the market value of the firm.[67]

The difference between an option and an outright sale in this re-
gard is that the seller of an option retains the incentives to invest in
actions that are valuable in the event of downside risk, because in
that event the option will not be exercised and the optionor will
remain the owner.[68] Thus, options will be economically advanta-
geous if the parties have some comparative advantage in managing
the different types of risk—for instance, if the seller's actions are
more important in ensuring that the underlying asset satisfies some
basic standard of merchantability, while the buyer's actions are
more important in maximizing either the potential resale value of
the underlying good or its synergistic value when combined with
other complementary assets.

Additionally, options will be a useful incentive device if the op-
tionee has limited available funds or faces a high cost of liquidity.
The price necessary to acquire an option will usually be substan-
tially less than the cost of buying the underlying asset outright, so
the optionee can more easily afford it. On the other hand, the po-
tential for upside risk may provide almost as strong a motivation
for the optionee to act efficiently, especially if the parties think that
downside risk is low. In such circumstances, options are a more ef-
ficient way to provide high-powered incentives than outright sales.

The incentive effects of options have been recognized and ap-
plied in specific contexts by a number of contributors to the law
and economics literature, but their general implications for con-
tract law have not been drawn out. For instance, in his well-known
article on the problem of providing multiple parties to a transac-
tion with simultaneous incentives to take appropriate precautions
against loss (what he calls "double responsibility at the margin"),

---

[67] See, e.g., David Yermack, Do Corporations Award CEO Stock Options Effec-
tively?, 39 J. Fin. Econ. 237, 242–47 (1995) (discussing the incentive rationale for ex-
ecutive stock options). But see Lucian Arye Bebchuk et al., Managerial Power and
Rent Extraction in the Design of Executive Compensation, 69 U. Chi. L. Rev. 751
(2002) (criticizing incentive argument for executive stock options and arguing that the
prevalence of such compensation arrangements is better explained by managerial
power and rent extraction).

[68] This division of incentives is a feature of both put and call options—the only dif-
ference being the identity of the residual owner in the event that the option is not ex-
ercised.

Professor Cooter identifies the call option as one of the standard legal devices that can be used to accomplish such double responsibility.[69] The underlying intuition for his argument is that if the option is properly priced, the optionor will want to take optimal precautions to guard against the event that the optionee fails to exercise the option, and the optionee will want to take optimal precautions to make the most of performance because she anticipates that once the optionor has taken precautions, she will then want to exercise the option.[70] In a similar spirit, Professors Nöldeke and Schmidt have shown how options can be used to solve the problem of efficient promisee reliance,[71] Professors Edlin and Hermalin have shown how options can be used to promote efficient relational investment,[72] and Professors Masten and Crocker have shown in particular how take-or-pay contracts—which as we have seen amount in substance to an option contract—help improve relational incentives in the natural gas industry.[73]

The foregoing account explains why contracting parties would want to use options, but how does it explain the way in which they choose between the basic elements of option premium, strike price, and option life? The answer is straightforward and can be explained as follows. Consider a proposed option contract with given values of P, S, and T, and imagine a marginal change in T that would increase P and decrease S. Such a change would marginally shift incentives to the optionee (and away from the optionor) by increasing the chance that the optionee will want to exercise the option. This shift is desirable if the marginal value to the contract of the optionee's incentives is greater than the marginal value of the optionor's incentives.

Similarly, imagine a marginal change in the option's terms that would increase both P and T. This change would also increase the

---

[69] Robert Cooter, Unity in Tort, Contract, and Property: The Model of Precaution, 73 Cal. L. Rev. 1 (1985).

[70] Id. at 3–4, 22–23.

[71] Georg Nöldeke & Klaus M. Schmidt, Option Contracts and Renegotiation: A Solution to the Hold-up Problem, 26 RAND J. Econ. 163 (1995); Georg Nöldeke & Klaus Schmidt, Sequential Investments and Options to Own, 29 RAND J. Econ. 633 (1998).

[72] Aaron S. Edlin & Benjamin E. Hermalin, Contract Renegotiation and Options in Agency Problems, 16 J.L. Econ. & Org. 395, 404–09 (2000).

[73] Masten & Crocker, supra note 31, at 1084–88.

chance that the optionee will exercise the option and thus would marginally shift incentives from the optionor to the optionee. It would also have the effect of providing additional incentives for the optionor to take precautions against low-value realizations for a longer period of time. (For a concrete example, consider a retailer that offers a money-back guaranty, and thus effectively commits itself to provide post-sales services or technical support in order to induce the buyer not to take advantage of the guaranty.) In this way, the parties can settle on the efficient option for their specific situation by adjusting P, S, and T until the marginal value of enhancing the optionor's incentives in each dimension is exactly balanced by the marginal value of enhancing the optionee's incentives.

C. Mixed Explanations

The above discussion suggests some fairly unambiguous recommendations for public policy regarding option contracts: To the extent that the choice to use an option contract and the tradeoff among its fundamental terms is motivated by efficiency concerns, the courts should enforce those contracts, notwithstanding traditional doctrinal barriers such as the illusory promise rule or the penalty doctrine. Similarly, when determining whether ambiguous contracts are options or not (for example, in distinguishing between an offer for a bilateral contract and an offer for a unilateral contract), courts should keep in mind the substantive economic factors that go to whether an option would be efficient in the particular setting. To the extent that the use of options is driven by inefficient motivations such as the desire to exclude competitors or to extract excess profits from boundedly rational consumers, courts should not defer to the parties' private decisions. What makes these problems harder and more interesting is that there are also motivations for using option contracts that are neither clearly efficient nor clearly inefficient. What public regulators should do in such areas depends on their views regarding the burden of error in balancing private autonomy against the general social interest. In this Section, I discuss three types of motivation that can be characterized in this way: information signaling, price discrimination, and coordination with other participants in a thin market.

1. Ex Ante Information Signaling

The way in which an option assigns responsibility for upside risk to the optionee and responsibility for downside risk to the optionor also makes it useful as a device for signaling private information. Often, potential contracting parties have information relevant to the value of an exchange that they cannot communicate credibly at reasonable cost. For instance, a seller of a brand new product may believe with high confidence that the product is a good one but will find it difficult to convey credibly that belief to a wary buyer because sellers of low-quality products have a self-serving incentive to make the same assertion. One common way to deal with this problem is for the seller to enter into a legally enforceable warranty, and to build the cost of servicing the warranty into the contract price. Because honoring the warranty would be more costly for a low-quality seller than for a high-quality seller, the fact that the seller is willing to offer the warranty without excessively marking up the price is a strong and credible signal of product quality.[74]

Granting an option to purchase is another way for such a seller to signal product quality. Indeed, a warranty is itself a kind of option—namely, an option to return the goods if they do not measure up to the promised quality. The usual warranty of merchantability gives a moderately strong signal of quality, but an unconditional guaranty of satisfaction sends an even stronger signal. Some retailers that have special reason to assure buyers of the value of their wares—for example, mail-order companies such as L.L. Bean and Lands' End—even offer a guaranty that allows buyers to return goods for any reason and at any time. (In our framework, this unusual guaranty would translate into a put option with $T=\infty$.)

Conversely, a party with private information about the quality of its likely performance will find it difficult to find a contractual partner willing to make a nonrefundable up-front payment or counter-performance. Committing to up-front performance, even in exchange for a concession on the contract price, is too risky because of the chance that the party who receives the commitment is a low-quality partner. Thus, parties who lack information about

---

[74] See, e.g., Sanford J. Grossman, The Informational Role of Warranties and Private Disclosure about Product Quality, 24 J.L. & Econ. 461, 470–77 (1981); George L. Priest, A Theory of the Consumer Product Warranty, 90 Yale L.J. 1297, 1303 (1981).

each other, or who are in new relationships, are unlikely to find it profitable to trade using options with high P and low S. The problem here is adverse selection; a partner who demands an up-front unconditional commitment is not to be trusted.

The reason why I classify informational signaling under the heading of mixed considerations, even though such signaling is sometimes the only way for parties to engage in an exchange in the context of adverse selection, is that such signaling operates as an informational externality in the market. The fact that the parties to an exchange find it in their interest to signal through the use of an option or other contractual terms does not necessarily imply that such terms are socially beneficial because they can have an adverse effect on other market actors not party to the exchange. To the extent that a contract term allows a seller to improve its sales prospects by signaling its reliability, for instance, part of this gain comes at the expense of its competitors. After the signal, this seller will be perceived as more reliable than average, but the others in its risk pool will, as a direct consequence, be perceived as less reliable than average. Unless those other sellers offer similar terms, they will lose sales.

Such competition through signaling is not necessarily efficient in contrast to ordinary price competition, which is generally efficient. Sellers who signal through the choice of contract terms may find that the more signals that are sent, the more their signaling value is diluted because sending a signal no longer serves to distinguish one from one's competitors. The result may be a rat race in which all parties are worse off than if none had tried to signal in the first place.[75]

The proper policy response to such external effects, if they exist, is not obvious. Depending on the particular circumstances, it may be possible to improve social welfare either by subsidizing information acquisition or by taxing it.[76] Because we do not otherwise seem

---

[75] The classic formulation of this argument can be found in Michael Spence, Job Market Signaling, 87 Q.J. Econ. 355, 361–68 (1973) (analyzing equilibria in labor markets where employees send costly signals of their quality); see also George Akerlof, The Economics of Caste and of the Rat Race and Other Woeful Tales, 90 Q.J. Econ. 599 (1976) (discussing how signals can upset market equilibria).

[76] See, e.g., Severin Borenstein, The Economics of Costly Risk Sorting in Competitive Insurance Markets, 9 Int'l Rev. L. & Econ. 25, 32–33 (1989) (demonstrating the ambiguity of optimal policy response in the insurance setting).

to have a strong social policy of regulating exchanges to guard against the externalities produced by excessive signaling, however, and because it seems unlikely that there is any special reason to adopt such a policy in the area of option contracts in particular, it is probably most reasonable to treat such motivations as legitimate ones in the private contractual area, and to defer to whatever option contracts the parties choose to design.

2. Price Discrimination

Offering an exchange in the form of an option may also be a way of separating out buyer groups with different elasticities of demand and charging them different prices (or, in the case of buyers who offer options, a way to discriminate between different groups of sellers). If buyers who value the flexibility of an option are willing to pay a higher price for the underlying product than buyers who do not, offering one contract with a low option premium and a high strike price, and another with a high option premium and a low strike price, operates to offer a selective discount to the low demand sector of the market. Such a strategy requires, of course, that individual buyers be unable to assign their contract rights to each other, because such trade would undermine the price discrimination scheme. For contracts for the sale of services, however, or for goods whose location and use is easily monitored, restricting resale is often feasible.

A standard example of this pricing strategy is found in the airline industry. Airlines typically sell at least two types of tickets: unrestricted use tickets, which can be credited toward the price of a ticket on another flight or even cancelled without restriction, and advance purchase tickets, which are nonrefundable and cannot be credited toward the price of another ticket without penalty. The former tickets are substantially more expensive, but can be purchased at any time up until the flight leaves the gate; the latter are cheaper but can only be purchased some time in advance of travel.

The standard account that industrial organization economists give of this pattern of prices is that it is a way for the airlines to discriminate between business and personal travelers. Business travelers have more inelastic demand for airline travel and place high value on the option to change plans. Thus they are willing to buy tickets with low P (that is, a small unconditional payment) and high

S (that is, a high price paid in the event they actually decide to travel). Personal travelers are less willing to pay (they have the alternative of going by car or substituting a local vacation) but often have greater flexibility in scheduling. Thus they are willing to buy tickets with relatively high up-front P so long as the subsequent payment S is small.[77]

In this example, the airline's price discrimination strategy may well promote economic efficiency, because it helps to cover the overhead cost of running the airline and maintaining a travel network. By offering both kinds of tickets, the airline manages to extract a larger portion of the business travelers' surplus from exchange while still serving the personal sector at a price that is closer to its actual marginal cost. The airline covers its overhead and increases total quantity supplied at the same time.

The reason why I classify price discrimination under the heading of mixed considerations, however, is that it is difficult to generalize about whether it increases or decreases economic efficiency on balance. In some instances, as when two-part pricing helps a firm with market power to recover fixed costs without sacrificing marginal sales, price discrimination is efficient.[78] In other cases, however, price discrimination can reduce market efficiency.[79] The industrial organization literature suggests that it is possible in principle to distinguish cases in which banning price discrimination would be socially desirable, but the information on which such a judgment would depend is difficult to acquire and few commentators are confident that courts are up to the task. Thus, viewing the matter from the perspective of efficiency alone, as opposed to norms of distributional fairness or equal treatment, we should defer to private option contracts that are designed and entered into for this motivation. As I argued above with regard to the issue of informational signaling, we do not have a strong social policy of regulating exchanges to guard against price discrimination in gen-

---

[77] For a more formal analysis, see Lars A. Stole, The Economics of Liquidated Damage Clauses in Contractual Environments with Private Information, 8 J.L. Econ. & Org. 582, 591–93 (1992).

[78] See Walter Y. Oi, A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly, 85 Q.J. Econ. 77, 78–81 (1971).

[79] See generally Hal R. Varian, Price Discrimination, *in* 1 Handbook of Industrial Organization 597, 631–34 (Richard L. Schmalensee & Robert D. Willig eds., 1989).

eral. There is thus probably little basis to apply such a policy to the category of option contracts in particular.

3. Coordination

Finally, if markets are thin, it might be in contracting parties' interests to structure their exchange as an option in order to coordinate with other parties who are doing the same in their own contracts. To illustrate this possibility, consider a resort hotel that operates in a remote vacation area along with a small number of competitors. If all the other hotels in the area have adopted pricing policies that include large nonrefundable deposits, there will be relatively few cancellations and hence few rooms available at the last minute. Travelers will accordingly anticipate that their chances of finding a room on a walk-in basis are slim, and so few will undertake to travel to the area without firm reservations. Given the low number of potential walk-ins, a cancelled reservation will result in a significant loss expected for the hotel, justifying a steep deposit. Conversely, if other hotels allow cancellation without penalty, there will be more rooms available at the last minute, thus encouraging more customers to travel without reservations, leading to a thicker walk-in market, leading to lower expected losses for the hotel following cancellation, leading to a lower required deposit.

What is going on here is a network externality that results from the fact that some minimum number of transactions is required to support an effective market for walk-in business. Given this externality, there can be multiple equilibria in this market, with the specific outcome depending on the configuration of expectations and on the participants' past behavior.[80] In such settings, it is difficult to draw robust conclusions regarding efficient behavior. Private attempts to coordinate pricing policies could improve efficiency, but they could also lead to cartel-like behavior and price-fixing. This motivation is thus a mixed one, possibly justifying regulatory intervention to discourage cartelization or to help parties coordinate on an equilibrium that they cannot reach privately.

---

[80] See W. Brian Arthur, Self-Reinforcing Mechanisms in Economics, *in* 5 The Economy as an Evolving Complex System 9, 10–13 (Philip W. Anderson et al. eds., 1988); Michael L. Katz & Carl Shapiro, Network Externalities, Competition, and Compatibility, 75 Am. Econ. Rev. 424, 427–28 (1985).

IV. APPLICATION TO DOCTRINAL AND TRANSACTIONAL PROBLEMS

The previous Parts of this Article set out a simple theoretical framework intended to unify our understanding of the various uses of option contracts. Because this framework incorporates a variety of economic and commercial considerations that may well cut in opposite directions in individual cases, however, any attempt to draw specific conclusions from that framework must be tentative at best. Indeed, it is for this very reason that I have elsewhere argued that public lawmakers are not often in a particularly good position to issue strong prescriptions regarding such tradeoffs, and that private parties should be allowed the leeway to choose their favored arrangement absent credible suspicion of externalities or some other market failure.[81] Nonetheless, the foregoing discussion does suggest some basic heuristic principles that might be used to assist both public regulation and private planning. As an illustration of these principles, in this Part I return to a number of the doctrinal puzzles that motivated the original inquiry, and analyze them in terms of the theoretical framework developed above. Then, in order to show that the framework has value in transactional settings as well, I use it to analyze two illustrative business practices—vacation deposits and resale price maintenance—that might otherwise appear anomalous but that can be explained more easily when viewed from an option perspective. A caveat is in order at the outset, however: The following analysis should not be taken as a definitive explanation of the particular phenomena at issue because I have undertaken no detailed empirical surveys in this regard. Rather it is offered as a guide to and recapitulation of the basic lessons of this Article.

A. Distinguishing Doctrinally Between Standard Contracts and Option Contracts

As we saw in Part I above, many legal doctrines apply differently to contracts that are explicitly denominated as options, even though from a conceptual viewpoint every contract that is enforceable through money damages is effectively an option. In order for this doctrinal framework to be coherent it is thus necessary to iden-

---

[81] See generally Katz, Economics of Form, supra note 2, at 506–14.

tify some substantive reasons for treating options differently and some criteria for determining which contracts qualify for treatment as options and which do not. Because the substantive reasons for differential treatment are presumably as varied as the specific doctrines that provide for it, this Article does not attempt to provide an exhaustive list of such explanations. Instead, I focus for illustrative purposes on three specific legal applications: contractual penalties, requirements contracts, and firm offers.

1. Contractual Penalties Versus Options

Assessing the use of options as a device to implement contractual penalties is a complicated matter. There can be many explanations for such transactions because penalties themselves can be used for various purposes: to motivate specific investment, to signal information, to facilitate price discrimination, or to exclude subsequent competitors.[82] If the intended penalty is motivated by the parties' joint pursuit of economic gain, and there are no third-party externalities that result, it is unclear why one would want to distinguish between options and penalties from a public regulatory viewpoint. If the main explanation for restricting contractual penalties is bounded rationality, however, a formal legal distinction between the two devices may make sense. Courts and commentators arguing for strict judicial supervision of liquidated damages have often stressed the possibility that contracting parties might, out of excessive optimism or trust, discount the possibility of breach or the likelihood that such damages would ever have to be paid.[83] Casting a penalty clause in option clothing, however, probably makes it clearer to the less sophisticated party that the payment is a basic contractual obligation, especially if the penalty takes the form of a prepaid deposit. The use of an option or a deposit can therefore mitigate the concern that the party who makes the deposit might fail adequately to appreciate the risk she undertakes in

---

[82] For a survey of explanations, see Edlin & Schwartz, supra note 36.

[83] See, e.g., Robert A. Hillman, The Limits of Behavioral Decision Theory in Legal Analysis: The Case of Liquidated Damages, 85 Cornell L. Rev. 717, 727 (2000) (citing Jeffrey B. Coopersmith, Comment, Refocusing Liquidated Damages Law for Real Estate Contracts: Returning to the Historical Roots of the Penalty Doctrine, 39 Emory L.J. 267, 268 & n.6 (1990) (citing Charles T. McCormick, Handbook on the Law of Damages § 147 (1935))).

doing so. Having to part with cash or other property in hand is a powerful corrective to such myopia.

Additionally, to the extent that one is sympathetic to the substantive purposes of the penalty doctrine, it may still be worth distinguishing between options and deposits that are substantively equivalent to penalty clauses and have no other plausible economic purpose, and options and deposits that may have the effect of penalties but that are also motivated by other significant economic concerns. As an illustration, consider the role played by deposits in the typical commercial sales contract for new construction, high-rise luxury condominiums in New York City. Such contracts typically provide for nonrefundable down payments in the amount of 25% of the overall purchase price, even though in some cases 25% might well be an overestimate of the developer's lost expectation should the purchaser back out of the deal.[84] (Of course, it might also be an underestimate, as this market is a particularly volatile one, experiencing local short-run price swings of 25% or greater several times in recent years.[85])

In this context, the nonrefundable deposit operates as an important device for allocating price risk in a volatile and speculative market. Like financial derivative instruments more generally, it makes it possible for investors to spread risk more flexibly by splitting and selling off portions of both upside and downside risk to other market participants. The purchasers in such contracts are often themselves brokers or speculators who plan on reselling the condominium units to residential purchasers after the project is completed if marketing to the general public is feasible. Spreading risk in this way reduces the overall cost of financing commercial real estate developments, with associated benefits to the developers, brokers, and ultimate purchasers. Thus, even if the deposit has the effect, or even is partially motivated by the goal, of penalizing a canceling buyer, its other economic benefits may justify exempting

---

[84] See Uzan v. 845 UN Ltd. P'ship, 778 N.Y.S.2d 171, 172 (N.Y. App. Div. 2004) (enforcing defendant's retention of 25% nonrefundable deposit following plaintiff's default, on grounds that the deposit was negotiated at arm's length and reflected the standard percentage in a volatile market for preconstruction luxury condominium units).

[85] Douglas Elliman, Manhattan Market Report, 1994–2003, at 5, 15, 21, 28, 32 (2004), available at http://www.millersamuel.com/reports (on file with the Virginia Law Review Association).

2230   *Virginia Law Review*   [Vol. 90:2187

it from the usual common law rule prohibiting contractual penalties. Some important commercial jurisdictions have provided just such an exemption.[86] More generally, the penalty doctrine should not be applied to transactions involving commercially sophisticated parties with significant speculative purposes who are contracting in a volatile market, on similar logic.

2. Requirements Contracts

A requirements contract allows a buyer to obtain a call option on a guaranteed source of supply, without herself committing to purchase any particular quantity. For this reason, such contracts—as well as symmetric output contracts in which a seller obtains a put option to sell her entire production output—were not enforceable at common law, on the grounds that the party with the option had given no consideration. The modern rule makes such contracts enforceable, with the complication that the optionee's discretion in choosing how much to buy must be exercised in good faith and in reasonable proportion to estimated or previously traded quantities.[87] These remaining doctrinal limitations are considered by most commentators to be necessary safeguards against the risk that the optionee will exercise her discretion in an opportunistic way—for instance, by excessively increasing her purchases following a sharp increase in market price in order to resell to other buyers when the parties had originally anticipated that the buyer would purchase the underlying good only as a production input.

Requirements contracts have very different risk and incentive properties than traditional fixed-quantity contracts. A fixed-quantity contract enforceable via money damages is also an option, but with a quite different exercise price. Under the fixed-price contract, the strike price, S, equals the contract price minus the expected cost of damages payable upon breach. Under the requirements contract, the strike price equals the contract price in full. Accordingly, determining which configuration of strike price, option premium, and term is appropriate (and hence determining whether a given exchange should be structured as a requirements

---

[86] See, e.g., Lawrence v. Miller, 86 N.Y. 131, 139–41 (1881); Maxton Builders, Inc. v. lo Galbo, 493 N.Y.S.2d 825 (N.Y. App. Div. 1985), aff'd, 502 N.E.2d 184 (N.Y. 1986).
[87] U.C.C. § 2-306 (2004).

contract or a fixed-quantity contract) requires attention to the various considerations outlined in the framework developed above.

One obvious transactional consideration in this regard is risk allocation, in that requirements contracts allow the seller to insure the buyer against the risk of events that affect her ex post demand. Indeed, it is on risk allocation grounds that requirements contracts are most frequently justified. But such contracts also serve other functions appearing on our list. They give the seller a stronger incentive to produce goods that the buyer will find useful, and signal the seller's confidence that the goods will be of high quality and available in sufficient supply. They also encourage the buyer to enter into relational investments such as storage facilities and purchases of complementary inputs, the values of which depend on the seller's performance and tender of the underlying goods.

Requirements contracts also have the disadvantage that the seller, to cover the expected costs of risk-bearing and buyer opportunism, must set a strike price above his ex post variable cost in order to break even, just as under resale price maintenance the seller must set a strike price that covers fully allocated overhead cost. The gap between the strike price and the optionor's marginal cost of performance implies that there will be some amount of allocative efficiency, which arises whenever the buyer's ex post willingness to pay for the goods exceeds marginal cost but falls short of the strike price. Absent some other device for eliminating such inefficiencies, such as ex post renegotiation, requirements contracts are only justified when this inefficiency is outweighed by countervailing advantages of risk allocation and signaling.

As a more specific illustration of these factors, consider the case of *Columbia Nitrogen Corp. v. Royster Co.*, which involved a dispute over express quantity terms in a contract between two chemical companies.[88] Columbia, the buyer and defendant, had for several years sold significant amounts of nitrogen to Royster, which used the nitrogen as an input to the manufacture of chemical fertilizer.[89] Royster then constructed a new facility that produced more phosphate (another fertilizer input) than it could itself use, and the companies executed a contract for Royster's sale of a minimum of

---

[88] 451 F.2d 3 (4th Cir. 1971).
[89] Id. at 6.

31,000 tons of phosphate each year for three years to Columbia, with an option to extend the term.[90] The contract also contained a standard merger clause that excluded any verbal or prior understandings relating to the contract.[91]

Between the execution of the contract and the time of performance, however, the market price of phosphate plunged, and as a result Columbia ordered less than a tenth of the phosphate that Royster was scheduled to ship in the first contract year.[92] When the case came to trial, Columbia claimed that notwithstanding the apparently explicit quantity clause, the contract was subject to a purported trade usage under which minimum quantities could be reduced at the buyer's option if the buyer turned out not to need the goods.[93] Essentially, Columbia claimed that an apparently explicit fixed-quantity contract was really a requirements contract. It offered to prove this claim with the testimony of other participants in the trade and by pointing to its pattern of dealings with Royster on previous contracts, in which Columbia as seller had allegedly allowed Royster as buyer to treat minimum quantities as optional requirements.[94]

Columbia Nitrogen is generally regarded as the leading authority on U.C.C. Section 2-202(1), which provides that contract terms may be explained or supplemented by course of dealing, usage of trade, or course of performance, even when the terms appear to be unambiguous or when the writing is found to be a complete integration of the parties' agreement.[95] Its outcome turned on the application of U.C.C. Section 1-302(e), which in present form provides:

> [T]he express terms of an agreement and an applicable course of performance, course of dealing, or usage of trade must be construed wherever reasonable as consistent with each other. If such construction is unreasonable:

---

[90] Id.

[91] Id. at 6 n.2.

[92] Id. at 7.

[93] Id. at 7–8.

[94] Id.

[95] U.C.C. § 2-202(1) (2004).

2004]   *The Option Element*   2233

(1) express terms prevail over course of performance, course of dealing, and usage of trade;

(2) course of performance prevails over course of dealing and usage of trade; and

(3) course of dealing prevails over usage of trade.[96]

In the end, the court held that it was possible to construe the alleged trade usage as consistent with the express quantity terms and so allowed the evidence to go to the jury.[97]

The court's decision in the case has been criticized on the dual grounds that it failed to give adequate weight to the express language of the contract and that it mistakenly labeled as a binding trade usage a pattern of behavior that more accurately reflected the parties' practical decisions not to insist on enforcing their legal rights.[98] Whether this criticism is merited in the particular instance cannot be definitively determined on the basis of the information available in the court papers, but the framework of the previous parts of this Article can certainly be used to shed light on whether a requirement contract makes sense in this context, either as a general trade usage or in the particular contract in dispute. Given a pattern of dealing between the parties, a thick market on which to resell or cover, and the expectation of future business, structuring the agreement as a zero premium requirements contract could well be efficient as a generalized trade usage. The facts that the underlying good was a homogenous commodity subject to price fluctuation in world markets and that parties were operating in part as brokers for resale to others—both of which would subject an optionor to significant speculative risk—might call this conclusion into question. But given the differences from the typical business context of the standard exchange in the industry and from the par-

---

[96] Id. § 1-303(e).

[97] 451 F.2d at 15.

[98] See, e.g., S. Concrete Servs., Inc. v. Mabelton Contractors, Inc., 407 F. Supp. 581, 584 (N.D. Ga. 1975); Lisa Bernstein, The Questionable Empirical Basis of Article 2's Incorporation Strategy: A Preliminary Study, 66 U. Chi. L. Rev. 710 (1999); Richard Craswell, Do Trade Customs Exist?, *in* The Jurisprudential Foundations of Corporate and Commercial Law 118 (Jody S. Kraus & Steven D. Walt eds., 2000); Victor Goldberg, *Columbia Nitrogen v. Royster*: Do as They Say, Not as They Do (Sept. 24, 2004) (unpublished manuscript, on file with the Virginia Law Review Association).

ties' past course of dealing, it would likely not have been a good idea to apply this usage, if it existed, to the phosphate contract that underlay the parties' dispute in this case. Specifically, Royster had just entered a new line of business and had constructed a facility with significant excess capacity, so it was unclear how this new arrangement would work out or whether the parties would continue to deal on the same basis in the future. Similarly, Royster had likely used up significant capital and borrowing reserves in constructing its new facility, and it thus faced an increased risk of finding itself in a substantially less liquid and perhaps overextended position in the event of an adverse price shift.

For these reasons, an arrangement that may have been functional in the usual case was probably rather less so in this new economic environment. If the trial court in *Columbia Nitrogen* had been alert to these economic considerations, it might have concluded that even if there were a trade usage in the industry making all contracts into requirements contracts, that usage did not properly extend to this new situation. Or alternatively, if the parties and their attorneys had been alert to these considerations at the time they entered into their contract, they could have avoided the dispute and consequent litigation by explicitly referencing the alleged trade usage in their contract, and stating clearly that it was not to apply to this new exchange. These possibilities illustrate how attention to the economic underpinnings of the option contract can improve legal decisionmaking, both from the perspective of ex post litigation and of ex ante planning.

3. Firm Offers

Recall that zero premium offers were unenforceable at common law due to lack of consideration, and still require special formalities to be enforceable today absent an estoppel claim. Reaching a policy conclusion over whether these rules should be further liberalized, or determining whether an offeree's reliance is reasonable in a particular case, requires understanding why zero premium offers are made in the first place. The above framework suggests at least two scenarios in which such offers make economic sense, one applicable in new relationships, and one applicable in established relationships.

The best explanation of such arrangements in new relationships is that they provide a channel for informational signaling. Specifically, in situations where a buyer is uninformed about the value of the underlying transaction, she will be unwilling to pay a significant up-front premium for an option that she does not yet know whether she will want to exercise. Indeed, under these circumstances it is a rational strategy for an informed seller to offer terms that are just barely acceptable to the buyer ex post, thus expropriating the full value of the up-front premium.[99] For the same reason, buyers will be reluctant to make any significant informational investment in a proposed exchange, even if such an investment is necessary to evaluate whether the transaction is worthwhile.

The seller thus offers a zero premium (or even a negative one) as a way of inducing the buyer to enter into the option, and also as a way of credibly signaling that the underlying transaction is of high value. Given this dynamic, we should expect to observe zero premium offers in contexts where the parties do not know each other very well, as well as in transactions involving new or untested products where other signaling or commitment mechanisms are unavailable.

Zero premium price options may also make sense in relational settings as a way of saving on transaction costs. If the parties expect to engage in multiple repeated transactions, they can omit the up-front premium, saving the extra cost of recordkeeping and payment processing, and make up the lost revenues on future contracts. Additionally, to the extent that options serve an insurance function for the optionee, parties who anticipate a long-term relationship can use firm offers to spread risk over their various individual transactions. And a practice of offering free options with positive value may also operate as a type of repeated gift exchange that, for either sociological or reputational reasons, helps bond the parties together and thus reduce the chances of opportunistic behavior.[100] It may be this type of consid-

---

[99] For a formal model illustrating these incentives, see Katz, supra note 58, at 518–28.

[100] For example, Professors Fehr and Gächter have shown how repeated gift exchange can provide superior incentives in environments where some fraction of the population is altruistic. Ernst Fehr & Simon Gächter, Fairness and Retaliation: The Economics of Reciprocity, 14 J. Econ. Persp. 159 (2000).

eration that explains the pattern of behavior shown in *Columbia Nitrogen*.

B. Some Illustrative Transactional Puzzles

As we have just seen, focusing on the option element in contracting can shed light on a number of otherwise puzzling features of legal doctrine. But questions of option design affect not just lawyers concerned with the analysis of doctrine, but also lawyers engaged in the planning of transactions. Options make sense in some contexts and not in others, and we can use the framework developed above to help us better understand some of the situations in which options are and are not used.

1. Variations in Deposit Policies

Anecdotal evidence suggests that hotels and motels operating in resort areas often vary their deposit policy between high and low seasons. In season, when demand is high and rates are marked up, the innkeeper charges a stricter deposit—and not just a larger amount to correspond to the higher in-season room rate. Rather, the innkeeper's contract provides for a shorter cancellation period in which the guest can change plans and still obtain a refund. Some innkeepers refuse to allow any refunds at all after a deposit has been paid. Out of season, however, deposit policies are more liberal, and innkeepers may allow the guest to cancel without any penalty at all up to the expected time of arrival. This variation in deposit policy can be interpreted as a variation in the terms of the guest's exit option from the rental contract. In season, the overall price of the room is high, and the period of time in which the guest is free to exit without penalty is short. Out of season, the room price is relatively low, and the period of time in which the guest is free to exit is relatively long.

Such a pattern of contracting behavior may seem unexceptional, except for the fact that it runs precisely counter to the usual explanation of deposits in terms of compensation for the innkeeper's expected losses from cancellation. In season, the innkeeper's expected losses from cancellation are relatively low, because there are many tourists in the area and so it is easier to fill the empty room. Thus a deposit that truly reflected lost expectation would

2004]   *The Option Element*   2237

equal the room rate multiplied by the relatively low chance that the room would go unrented. Conversely, out of season, when hotels operate at less than capacity, it is much more likely that a cancelled room will go unrented, so the innkeeper's losses from cancellation are relatively high. Here, a deposit that approximates lost expectation would equal the room rate times the chance of the room going empty, which out of season is probably one. Why would the innkeeper choose a deposit policy that results in overcompensation in season and undercompensation out of season? What advantages are there to doing this?

A superficially appealing but ultimately unconvincing explanation is that the in-season deposit is higher because in-season customers have a higher willingness to pay. The answer is appealing because in-season customers are indeed willing to pay more, and there is a higher consumer surplus for the innkeeper to extract. The answer is incorrect, however, because there are other ways for the innkeeper to extract the customer's surplus by adjusting other terms of the contract without suffering the efficiency loss from an inaccurately priced deposit. The innkeeper could, for instance, charge an even higher price for the room while liberalizing his policy regarding deposits. Such a strategy could raise the same revenue while providing more efficient incentives to a customer deciding whether to cancel her reservation.[101] Innkeepers do not typically debase other terms of service during the high season, such as cutting back on housekeeping or restaurant facilities. If anything, they do the opposite and build the costs of those services into the room rate. Why then would they choose to worsen the terms of the deposit?

A variation on this question is why hotels and motels adopt such pricing policies, but some other service providers do not. Rental car companies also experience periods of slack and tight demand, for instance, and they also could require a nonrefundable deposit.

---

[101] Puzzles about why sellers do not sufficiently use price to ration quantity in periods of high demand are often answered with reference to social norms against price gouging. See, e.g., Daniel Kahneman et al., Fairness as a Constraint on Profit Seeking: Entitlements in the Market, 76 Am. Econ. Rev. 728 (1986); Robert J. Shiller et al., Popular Attitudes Toward Free Markets: The Soviet Union and the United States Compared, 81 Am. Econ. Rev. 385, 398 (1991). But there is already a significant gap between in-season and off-season rates, and it is unclear why social norms would object to a further increase in the gap.

The same is true of restaurants. Such companies do vary their prices in the form of weekend rates and early bird specials, but they do not generally require deposits when they take reservations, even though they easily could (indeed, car rental companies already require a credit card number as a condition of placing a reservation). Airlines, in contrast, follow an intermediate policy. Unlike innkeepers, they collect nonrefundable deposits all year round, but assess them only for some types of tickets (typically advance discount fares) and not for others, and the prices of these two types of tickets can vary by a factor of two or more. Does the optimally designed option vary between these markets, and if so, why? The analytical framework developed above helps answer these questions. First, it is clear that this seasonal pattern of deposit policies cannot be explained in terms of ex post incentives. We know from the standard economic analysis of contract damages that setting liquidated damages greater or less than the promisee's expectation is inefficient, because it leads to inefficient breach when the damages are too low and inefficient performance when damages are too high.[102] Here, the structure of the deposits excessively deters the traveler from canceling a reservation in season, when the deposit is high relative to the resort's likely expectation losses, and inadequately deters cancellation out of season, when the deposit is low relative to expectation losses.

Risk bearing seems a similarly unlikely explanation for in-season policies because the resort has many customers and is in business for the long haul, and is thus in a better position to spread the risk of most events that lead to cancellation. (On the other hand, risk bearing might well explain the lack of deposits out of season.) And an information signaling story seems implausible because the value of the exchange to the innkeeper is unlikely to depend on the traveler's private valuation. Indeed, it should be the other way around.

There are, however, three explanations for this pattern that do make sense. The first is bounded rationality. If vacation customers do not pay sufficient attention to the risk that they might need to cancel and thus lose their deposit, but do pay a lot of attention to

---

[102] See Robert Cooter & Melvin Aron Eisenberg, Damages for Breach of Contract, 73 Cal. L. Rev. 1432 (1985); Steven Shavell, Damage Measures for Breach of Contract, 11 Bell J. Econ. 466, 468 (1980); Steven Shavell, The Design of Contracts and Remedies for Breach, 99 Q.J. Econ. 121, 125 (1984).

the posted price, then a resort that chooses an ex post efficient deposit and adjusts its room rates accordingly will find that the customers respond disproportionately to the latter adjustment. The resort will thus lose revenues. Under such circumstances, it is possible that there would be a welfare improvement from regulation that required resorts to disclose or post their deposit policies on the same basis as price, although given the likely enforcement cost of such regulation and the relatively small efficiency loss from mispricing, the case for such a policy does not seem overwhelming.[103]

A second partial explanation might be coordination or network externalities. Specifically, an individual resort might choose a restrictive deposit policy in season because given that all other resorts in the area do the same, there are few last minute walk-ins and the expected loss from cancellation is accordingly large. This explanation does not account, however, for the failure to require deposits out of season. Since excess capacity out of season is presumably common knowledge, the decision of travelers to chance a last minute trip should not depend on resorts' deposit policies; and the chances of replacing a cancelled reservation is low in any event. Efficient ex post pricing therefore requires a deposit to protect the resort's lost expectation. Similarly, coordination would not account for the failure of car rental agencies or restaurants to permit free cancellation (although a comparative advantage in risk bearing might).

Finally, nonrefundable deposits could be a tool for price discrimination on the model of nonrefundable airline tickets. This explanation could make sense if it were the case that last minute travelers had more elastic demand than those who make reservations in advance, which seems at least plausible. Vacation travelers who reserve in advance presumably have a high preference for the specific location being reserved, while last minute travelers are more likely to be choosing among a number of resort destinations. If so, it would be rational for sellers to want to charge the last-minute customers a lower price and the advance customers a higher price. One way to do this would be to cut prices at the last

---

[103] But see Colin Camerer et al., Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism," 151 U. Pa. L. Rev. 1211 (2003) (arguing for policies that prevent boundedly rational parties from making mistakes).

minute, but this policy would not be sustainable if it became generally known, because then advance customers could cancel their reservations at the last minute and then immediately rebook at the last-minute price. The deposit policy achieves a similar result by collecting some of the price charged to advance travelers in the form of forfeited deposits. Last-minute customers, who never put down a deposit and hence never lose one, thus pay a lower price on average. The price discrimination is sustainable, furthermore, because in order to obtain the lower price, it is necessary to run the risk of being unable to get a reservation at the particular resort, which advance customers would be less willing to chance. (Note that on this explanation, price discrimination works exactly opposite to the way it works in the airline context, where airlines charge a premium to last-minute business travelers and a discount to those who book in advance, on the ground that the latter have more elastic demand.)

2. Resale Price Maintenance and Recovery of Selling Costs

A second example comes from the borderline of contract and antitrust law. Manufacturers of goods have at various times tried to require their distributors to charge a minimum markup when reselling the goods to their own wholesale or retail customers. This practice, called resale price maintenance, has long been controversial and is currently illegal under U.S. antitrust law.[104] Opponents of the practice have made both the formal argument that it is a form of price-fixing, otherwise per se illegal under the Sherman Act, and the functional argument that it provides a method for retailers to cartelize and to exclude competition from more efficient discounters.[105]

---

[104] Relevant case law in this regard includes *United States v. Parke, Davis & Co.*, 362 U.S. 29 (1960) and *Dr. Miles Medical Co. v. John D. Park & Sons Co.*, 220 U.S. 373 (1911). In 1937, Congress passed the Miller-Tydings Act, which legalized RPM in states that chose to authorize it in local so-called "fair trade" laws. At the high-water mark of the practice, four states had adopted such laws. In 1975, however, this act was repealed and so the practice again became per se illegal, pursuant to prior law. See Richard A. Posner, Antitrust Law 176–89 (2d ed. 2001).

[105] See, e.g., Herbert Hovenkamp, Vertical Restrictions and Monopoly Power, 64 B.U. L. Rev. 521, 529–35 (1984).

2004]   *The Option Element*   2241

Defenders of a more laissez-faire antitrust policy, however, beginning with Professor Lester Telser, argue in response that RPM may be necessary in order to induce the wholesalers and retailers to invest adequately in selling expenses such as well-stocked showrooms and knowledgeable salespersons.[106] Otherwise, wholesalers and retailers who do provide such services will find themselves undercut by cost-cutting competitors who offer low sales and service quality, and who rely on their customers to do their initial shopping for free at the high-price outlet. Manufacturers could directly require their distributors to provide a given service level as a condition of being allowed to carry the manufacturer's brand, but specifying that service level and monitoring the retailer to ensure that it is actually provided is expensive, and may be impossible in light of the subjective nature of the standards by which such inputs must be measured.[107] On this view, it is more cost-effective to restrict price competition, thus redirecting competitive pressure into nonprice directions. Resale price maintenance should accordingly be permitted as a reasonable business practice, rather than treated as a per se violation of the Sherman Act as it is under current precedent.

In our conceptual framework, however, RPM can be seen as a specialized kind of zero premium option, under which the distributor first provides sales and informational services at no cost, and then stands ready to sell the underlying product at a strike price that is fixed by the manufacturer-distributor contract. The main difference between RPM and the standard firm offer is that under RPM, the option offered to the customer is implicit rather than explicit. The distributor's obligation to sell may or may not be enforceable as a matter of contract law, depending on the contract formation rules of the local jurisdiction, but it is still an option

---

[106] Lester G. Telser, Why Should Manufacturers Want Fair Trade?, 3 J.L. & Econ. 86, 89–96 (1960); see also Robert H. Bork, The Antitrust Paradox: A Policy at War with Itself 290–91 (1978) (elaborating and endorsing Professor Telser's argument); Richard A. Posner, Antitrust Law: An Economic Perspective 147–48 (1976) (same).

[107] Cf. Benjamin Klein, Transaction Cost Determinants of "Unfair" Contractual Arrangements, 70 Am. Econ. Rev. (Papers & Proc.) 356, 359 (1980) (explaining and defending at-will termination clauses on similar grounds); Alan Schwartz, Relational Contracts in the Courts: An Analysis of Incomplete Agreements and Judicial Strategies, 21 J. Legal Stud. 271, 315–18 (1992) (arguing generally that courts cannot effectively enforce subjectively measured contractual obligations).

when viewed from an economic perspective, and likely to be carried out for credible business motives.

An analogous issue arises in the field of contract law in connection with the assessment of a seller's damages for lost profits following a buyer's breach. Under U.C.C. Section 2-708(2), sellers are entitled to such damages whenever the usual contract-minus-market damage formula is inadequate to put the seller in as good a position as performance would have.[108] Conventional wisdom holds that this is the case whenever the seller is unable to resell the goods that the buyer refused to take.[109] But in response to the conventional wisdom, Professor Victor Goldberg argues that in assessing lost profit it is critical to take into account the up-front expenses the seller incurred in making the sale initially.[110] If the cost of such expenses is adequately covered by the expected price in the event that the contract is completed, on Professor Goldberg's view, there is no need to assess lost-profit damages because the parties will have implicitly dealt with the problem in setting the terms of their arrangement, including particularly the amount of any agreed deposit. The analysis is similar to the analysis of RPM because if Professor Goldberg is right, and the retail price is set high enough to cover sales costs expended on customers who do not buy, then the customer's obligation to repay the associated portion of lost profits ought to be interpreted, as with RPM pricing, as optional.

In order to evaluate Professor Telser's defense of RPM and Professor Goldberg's argument against lost-profit damages, however, we need to explain why retailers might prefer to recover their selling costs as part of the retail price of the goods, rather than charging for them directly. Why do automobile or electronics retailers not charge admission to their showrooms, for instance, so that the overhead cost of maintaining showrooms would be paid by all who are investigating a purchase, and so that the units that were actually purchased could be sold at variable cost? In our framework, this arrangement would also be a type of option, albeit one with a

---

[108] U.C.C. § 2-708(2) (2004).

[109] For a survey of commentators taking this position, see John M. Breen, The Lost Volume Seller and Lost Profits under U.C.C. § 2-708(2): A Conceptual and Linguistic Critique, 50 U. Miami L. Rev. 779, 781 n.4 (1996).

[110] Victor P. Goldberg, An Economic Analysis of the Lost-Volume Retail Seller, 57 S. Cal. L. Rev. 283, 290–97 (1984).

different mix of option premium and strike price. Under this alternative arrangement, the customer would, in exchange for the showroom admission price, purchase an option to buy the good at variable cost if, after inspecting the floor model and conversing with the sales staff, she still wished to do so. This alternative would provide customers with more efficient incentives with regard to both the initial decision to go shopping (because shopping imposes costs on the store even when one does not buy), and the subsequent decision whether to purchase (because bundling sales costs into the retail purchase price makes it more likely that customers who have already acquired the information necessary to decide whether to purchase will go to a discount chain, even if the discounter's variable cost is greater than that of the high-quality merchant). Why, then, do more retailers not offer it?

To address these questions, recall the possible explanations why sellers might prefer to offer a back loaded option. As before, ex post efficiency is unlikely to be the answer. Because the up-front informational services do have positive marginal cost, it would be more efficient, other things being equal, to charge for them separately and accordingly to lower the price of the final product. Bundling the cost of pre-sale services into the final product cost means depressing demand below the efficient level, as those who buy cross-subsidize those who shop but do not buy.

As we just observed in our discussion of firm offers, however, up-front cover charges will not work for new or unfamiliar products because buyers are uncertain whether they will want to purchase. Thus, RPM would make sense for firms selling in new but competitive markets, because otherwise it will be in no single seller's interest to provide pre-sale services. This is not a plausible explanation, however, for established firms with a commercial reputation, and historically it has been branded manufacturers that have been most likely to engage in RPM. Price discrimination is a logically possible explanation, but works only if those buyers willing to pay a higher price for customer service and sales-related overhead are also those who would be willing to pay only a lower price for the underlying product. (This is a straightforward application of the classic product-tying model, here applied to a tie be-

tween the underlying commodity and its associated sales quality.[111]) But such a correlation seems implausible if both customer service and the underlying product are normal goods. Absent a plausible alternative explanation, we thus might tentatively conclude that the traditional antitrust critics of RPM are right—the main motivation for the practice is to exclude competition or to facilitate a cartel among retailers.

CONCLUSION

All contracts contain features that resemble or that are economically equivalent to options, and many contracts are deliberately designed to take account of such options. In designing such contracts ex ante or in interpreting them ex post, it is essential to consider explicitly these option features and the costs and benefits that flow from them. The tradeoff among the basic elements of an option—the up-front premium paid for the option, the strike price that the option holder must pay to exercise the option, and the time period in which the option can be exercised—can affect the parties' incentives to perform or to invest in the contract, the allocation of risk achieved by the contract, the information exchanged between the parties before entering into the exchange, and even their incentive to contract in the first place. Thus, while option contracts may differ substantially across different fields of legal practice and different economic markets, it is still possible and helpful to identify and generalize from the basic features of option design.

---

[111] See William James Adams & Janet L. Yellen, Commodity Bundling and the Burden of Monopoly, 90 Q.J. Econ. 475, 477–90 (1976).