2004

The Efficient Design of Option Contracts: Principles and Applications

Avery W. Katz
Columbia Law School, ak472@columbia.edu

Columbia Law School
The Center for Law and Economic Studies
435 West 116th St.
New York, NY 10027-7201

The Efficient Design of Option Contracts:
Principles and Applications

Avery Wiener Katz

March 2004

The Efficient Design of Option Contracts:
Principles and Applications

Avery Wiener Katz[*]

March 2004

Abstract: The law of contracts has often treated options quite differently from other contractual transactions; for example, the characterization of a transaction as an option contract calls forth specially required formalities, but on the other hand often has the effect of releasing parties from doctrinal 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 challenging to explain from a functional viewpoint, in part because all contracts resemble options to the extent they are enforceable in terms of monetary damages, and in part because contracts that are nominally structured as explicit options can be close economic substitutes for contracts that are nominally structured as unconditional.

This essay sets out a theoretical account of the efficient design of option contracts - one that explains how contracting parties should strike the balance among option premium, option life, and exercise price, in order to maximize the expected surplus from their transaction. It shows that the tradeoffs between these various aspects of option contracts can affect the parties incentives to acquire and disclose information, to invest in relation-specific investments, and to take efficient precautions against the event of breach. It then goes on to develop an organizing framework for private parties choosing whether and how to structure their contractual arrangements as options, and for policymakers choosing whether or how to regulate such private choices. In short, the appropriate balance between option premium, option life, and exercise price will depend on the relative importance that the one attaches to these various dimensions of incentives.

The Efficient Design of Option Contracts: Principles and Applications

Many contractual arrangements are structured as options, or include options as important elements. In addition, many aspects of contractual practice, and of the legal doctrine governing contracts, can be interpreted as 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, whichever he prefers. Similarly, the doctrine of promissory estoppel, which attaches liability to pre-contractual statements in cases where they are reasonably relied upon, effectively grants a option to the relying party, who may enforce the promise or not as she finds convenient. A similar consequence follows where contracts are voidable (but not void) for reasons of mistake, infancy, or lack of an agent's authority.

Despite this connection, the law of contracts has often treated options quite differently from other contractual transactions. Options contracts with an explicit zero premium were not enforceable under the traditional common law, and even today are not enforceable without special formalities being undertaken. Conversely, the characterization of a transaction as an option contract often has the effect of releasing parties from doctrinal 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 in part 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 just 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; and a party who is just 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, draw a sharp distinction between zero-premium and positive-premium options?

Similarly, a party who demands a substantial prepaid deposit or ex post 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. Given this relationship, what factors, apart from a desire to avoid the constraints of the common-law penalty doctrine, determine how the parties choose between these possible substitute contracts? And how should the law distinguish, if at all, between front-loaded options 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 often been argued in defense of the practice of resale price maintenance 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 pre-contractual 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 barriers to doing so, but in the retail and wholesale context such arrangements are rare, with the exception of a few superstores that are nominally 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 essay presents such an account, and shows that the tradeoff between these various aspects of option contracts can affect the parties incentives to acquire and disclose information, to invest in relation-specific investments, and to take efficient precautions against the event of breach. The appropriate balance between option premium, option term, and exercise price, accordingly, will depend on the relative importance that the parties attach to these various dimensions of incentives.

The essay 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 screen for or to signal private information in the context of adverse selection; they can also be used for purposes of price discrimination. To this extent, it may be socially desirable to restrict the use of such contracts 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 essay, however, lies not in creating new theory but in surveying and synthesizing theoretical analyses from a variety of literatures and specialized bodies of knowledge, in order to developing a useful taxonomy of the considerations that are relevant in deciding whether and how 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: sales, information licensing, construction, financial instruments, and so on. 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.[1]

---

1. Citation omitted to preserve anonymity of submission process.

The organization of the essay will proceed as follows: section 1 will demonstrate the importance of the analysis by surveying the range of legal problems that reduce to the question of option design.   Section 2 will present a conceptual account of option design that demonstrates the basic relationship among the three fundamental elements of an option: 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. Section 3 relates those tradeoffs to the overall incentive structure created by the contract; it shows 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. Section 4 discusses how the analysis presented in the prior two sections relates to the legal problems that motivated the initial inquiry. Section 5 summarizes the main lessons of the essay and offers conclusions.

A note: the essay does not attempt to present a systematic account of option theory or options pricing; several good introductions to those subject are already available for readers who wish to consult them.[2] The focus of this essay, rather, 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.

---

2. See, e.g., Richard A. Brealey & Stewart C. Myers, Principles of Corporate Finance (6th ed. 2000); John Cox & Mark Rubinstein, Option Markets (1985); Avinash K. Dixit & Robert S. Pindyck, Investment Under Uncertainty (1994).

I. The importance of option contracts

Doctrinal treatments of contract law — casebooks, treatises and the like — typically present option contracts as a specialized and even minor topic, rather than as the widespread transactional device that they constitute in practice. In this section of this essay, accordingly, I survey the range of legal problems for which an option analysis is useful. Subsection A discusses the various black-letter doctrines under which contracts nominally labeled as options are treated differently from contracts generally. Subsection B relates option-contract analysis to the general area of contractual remedies. It discusses and critiques the familiar observation, commonly credited to 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, subsection C shows how taking an option-based perspective raises interesting practical questions about business behavior and about transactional planning, and not just a set of esoteric doctrinal puzzles.

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 counter-promise that flows back to the promisor or his designate.[3] 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: promises should presumptively be enforceable if they are made as part of a deliberate and arms-length economic exchange. On this understanding of the concept, most option contracts should qualify as bargains; they are

---

3. In order to avoid ambiguity when discussing generic option transactions, I adopt the convention in this paper 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.

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 many contracts with option terms as full-fledged exchanges, because the holder of such an option (the optionee) does not give anything to her grantor (the optionor) unless and until the option is exercised. For example, in *Wickham and Burton Coal Co. v. Farmers' Lumber Co.*,[4] 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 did not bind the buyer to any minimum amount. In the court's view, the fact that the buyer had the option to order no coal at all rendered any promise it made illusory, and hence insufficient to provide consideration. Similar results were reached in cases in which one party retained the option to terminate the contract at any time (since that party could if she wished escape all liability by exercising her option), 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).[5]

The evident fact that such promises arose out of arms-length exchanges entered into for purposes of risk and incentive allocation, however, was not lost on courts and commentators; and many strove to find limits on the optionee's discretion that would render his promise other than illusory. In *Wood v. Lucy, Lady Duff Gordon*[6], for instance, Judge Benjamin Cardozo famously found consideration by reading into an exclusive agency contract an implied promise of best efforts on the part of the promisee. Other courts have implied a duty to exercise

---

4. 189 Iowa 1183, 179 N.W. 417 (1920),

5. For a representative at-will termination case, see Miami Coca-Cola Bottling Co. v. Orange Crush Co., 296 F. 693 (5th Cir. 1924); for a representative requirements case, see Oscar Schlegel Mfg. Co. v. Peter Cooper's Glue Factory, 132 N.E. 148 (N.Y. 1921).

6. 118 N.E. 214 (N.Y. 1917)

discretion in good faith, as in *Mattei v. Hopper*[7], where the purchase of a parcel of commercial real estate was made subject to an independent broker's obtaining leases satisfactory to the buyer. The Uniform Commercial Code similarly implies a general duty of good faith in the performance of all contracts within its scope, and specifically uses the duty of good faith to limit the discretion that can be exercised under a requirements or output contract (§2-306), a sales contract which leaves the price or other particulars of performance o be fixed by one of the parties (§§2-305(2), 2-311(1)), and an acceleration clause (§ 1-208). Still others have stretched to find consideration in contractual or statutory notice requirements, even where those requirements did not substantially burden the optionee's exercise of discretion.[8] Nonetheless, the illusory-promise rule remains part of black-letter doctrine to this day; as Restatement (Second) of Contracts § 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."

There remains a channel, however, for parties who want to bind themselves to a contract while leaving their offeree with full discretion whether or not to go forward with the deal. Although the formal device of nominal consideration has been discredited as a general method of creating contractual obligation, Restatement (Second) of Contracts § 87(1) provides: "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." Similarly, in the law of sales, UCC §2-205 allows such "firm offers" to be made by a signed writing, but limits their

---

7.   51 Cal. 2d 119; 330 P.2d 625 (1958).

8.   See, e.g., Gurfein v. Werbelovsky, 97 Conn. 703, 118 A. 32 (1922); Lindner v. Mid-Continent Petroleum Corp., 221 Ark. 241 (1952).

enforceability to a period not to exceed 90 days.[9] 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 historical survey of doctrine is not that the courts held to formal accounts of consideration in the face of the commercial reality of the underlying deals, or 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 the option just slightly, they could have had themselves a legally enforceable deal. For example, 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 up front to hold the option open. Assuming that there is approximately a 50% chance that the option comes "into the money" (i.e., 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.[10]

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

---

9. UCC §2-205 ("An offer by a merchant to buy or sell goods in a signed writing which 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 such period of irrevocability exceed three months; but any such term of assurance on a form supplied by the offeree must be separately signed by the offeror.")

10. Of course, the reduction in exercise 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 up front. 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 2 years, if the time extension increased by an additional 50% the probability that the option comes into the money.

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

2. *Offer and acceptance*

In additional to the formal requirements of consideration referenced above, option contracts also implicate contract formation issues, since as the phrase "firm offer" indicates, options can be and often are viewed as irrevocable offers. The question of revocability, accordingly, reduces to the question of whether an option has been granted. And as with consideration, doctrinal treatment of this issue 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.[11] 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 he completes performance within a reasonable time; while the second Restatement extends this right to the mere beginning of performance, and also explicitly labels the resulting obligation an option.[12]

---

11. See, e.g., I. Maurice Wormser, The True Conception of Unilateral Contracts, 26 Yale L.J. 136 (1916) (presenting notorious hypothetical in which defendant, who has promised to pay plaintiff $100 for crossing the Brooklyn Bridge, withdraws the offer once plaintiff has walked three-quarters of the way across); Petterson v. Pattburg, 161 N.E. 428 (N.Y. 1928) (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);

12. Compare Restatement of Contracts §45("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") with Restatement (Second) of Contracts §45(1) ("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 supplied).

The irrevocability of an option contract, however, is not its only distinctive feature under the law of offer and acceptance. Various rules dealing with the expiration and rejection of offers are applied differently to firm offers than to offers generally. 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 counter-offer by the offeree, on the theory that the optionee is entitled to enjoy the full speculative value that she was promised.[13] On the other hand, contract terms governing the time and manner of the optionee's exercise of acceptance are applied strictly, in contrast to the more liberal treatment accorded to ordinary acceptances. As the comments to Restatement §25 explain: "any relaxation of terms would substantively extend the option contract to subject one party to greater obligations than he bargained for."

Additionally, the abstruse but traditional distinction between unilateral and bilateral contracts turns out in substance to be about options. As all first-year law students are taught, an offer for a unilateral contract can only be accepted by performance of the requested consideration, while an offer for a bilateral contract can only be accepted by a promise to perform. Courts and commentators have vacillated over the years on the difference between the two types of contract and on the question of which type should be presumptively favored. The first Restatement of Contracts, for instance, adopted the position that bilateral contracts were the norm, interpreting ambiguous offers as seeking a counter-promise rather than an act.[14] The second Restatement ostensibly de-emphasizes the distinction by allowing an ambiguous

---

13. See Restatement (Second) of Contracts §37 (1981): "[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 (Winter 1990).

14. See Restatement of Contracts §31 (interpreting ambiguous offers as offers for bilateral contracts).

offer to be accepted by either performance or a counter-promise as the offeree chooses;[15] and UCC Article 2 takes a similar approach in sales cases.[16] But the unilateral/bilateral distinction remains live for some purposes. In cases where an offeror explicitly invites an offeree to choose between acceptance by performance or by promise, a tender or beginning of performance operates as a promise to perform and hence a bilateral contract,[17] but where an offeror clearly requests acceptance by performance, the offeror becomes bound as soon as the offeree tenders or begins the invited performance and the offeree does not.[18]

Whether a given offer is interpreted as unilateral or bilateral, however, and whether the law should impose a presumption in favor of one or another, cannot be answered without having some sense of why parties would use such arrangements and what transactional purposes they serve. In this regard, it is important to recognize (indeed, comments to the second Restatement suggest that it is the "principal value of the distinction"[19]) that a bilateral contract binds the offeree unconditionally to perform while a unilateral contract leaves her the option to choose whether or not to perform. 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 Article 2 of the UCC, true to its standard interpretative philosophy, tells us

---

15. See Restatement (Second) of Contracts, §32. Indeed, the second Restatement even eschews the use of the terms "unilateral contract" and "bilateral contract," on the grounds of "doubt as to the utility of the distinction, often treated as fundamental, between the two types." Restatement (Second) of Contracts, §1, comment f.

16. See UCC §2-206(a), which provides that "an offer to make a contract shall be construed as inviting acceptance in any manner and by any medium reasonable in the circumstances."

17. Restatement (Second) of Contracts, §62. More precisely, subsection (1) provides that the tender or beginning of performance amounts to an acceptance by performance (hence a unilateral contract), but subsection (2) provides that such an acceptance operates as a promise to complete performance (and hence a bilateral contract).

18. See Restatement (Second) of Contracts, §45(1).

19. Restatement (Second) of Contracts, §1.

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.[20] 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 in the meantime.[21] 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.[22]

Viewed from a formal perspective, the logical rationale of the anticipatory repudiation and adequate assurances doctrines does not apply in the option contract setting. An optionee has no duty to go through with the deal; on the contrary she has bargained for the right to speculate,

---

20. See Restatement (Second) of Contracts, § 253 (Effect of a Repudiation as a Breach and on Other Party's Duties). The analogous provision under sales law is UCC §2-610 (Anticipatory Repudiation.)

21. See Restatement (Second) of Contracts, §251 (When a Failure to Give Assurance May Be Treated as a Repudiation); § 2-609 (Right to Adequate Assurance of Performance.)

22. See generally Richard Craswell, Insecurity, Repudiation, and Cure, 19 Journal of Legal Studies 399 (1990).

which remains valuable up till 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. (An option can of course be renounced; and a promise not to exercise it could, if relied on, be enforced on an estoppel theory, but otherwise the option remains valid.[23]) 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. How should courts and transactional planners weigh one consideration against the other?

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: to seek out a substitute arrangement, and to deduct the amounts earned under the substitute arrangement from the damages payable on the original contract. Any losses that could have been avoided by such mitigation 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 aggrieved party and not by the breacher.)

In contrast, in cases where the contract is interpreted as the promisor's having bargained for an option on the time or services of the promisee, the duty to mitigate is excused.[24] For example, in a so-called "pay-or-play" contract, common in the entertainment industry, the

---

23. [Cite to legal authorities on this proposition.]

24. See cases surveyed by Victor Goldberg, Bloomer Girl Revisited. or How to Frame an Unmade Picture, 1998 Wis. L. Rev. 1051 (1998).

promisor promises either to employ the promisee for a given job or to pay an equivalent salary. 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. 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 taken place. No breach means that no duty on the part of either party to mitigate ever arises. Whether the result makes sense in substantive terms, however, 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; 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 their removal from the project[25]; and Masten and Crocker find a similar pattern in the "take-or-pay" contracts used between natural gas suppliers and pipelines.[26] But a full explanation requires an account of why some contracting parties would do away with the duty to mitigate while others would not. The disadvantages of disclaiming the duty to mitigate are plain, in that the optionor's ex post incentives to take efficient steps to reduce losses are eliminated. What then are the countervailing advantages?

   c. Liquidated damages and penalties — Structuring a contract as an option may in many cases help the parties evade the penalty doctrine, which provides that parties may not contract for liquidated damages in an amount in excess of a fair estimate of the aggrieved party's damages for lost expectation. This can be done by providing for an up-front or unconditional

---

25. Ibid.

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

payment that is characterized as the purchase of an option, which can be exercised by tendering an additional sum. (The take-or-pay contract 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; there is no breach, and hence no opportunity for the penalty doctrine to come into play.[27]

If the prepayment is characterized as a deposit or performance bond, in contrast, the law will limit the extent to which it can be made non-refundable. For instance, in sales contracts, UCC 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 or $500, whichever is larger. And in the common-law setting, the breacher may have a good claim for restitution, on the theory that the forfeiture results in unjust enrichment for the aggrieved party to the extent it puts him in a better position than contractual performance would have done.[28]

Recent scholarship on the economics of contractual penalties have called into question whether the penalty doctrine can be justified on efficiency grounds[29]; and on such a view, it is desirable that a device that allows the parties to evade it be available. But if the economic critics are right, why don't contracting parties use this device all the time? And if they are not, why should legal doctrine provide this potential loophole?

B. Remedial rules generally

Even apart from any special treatment given to explicitly denoted option contracts, the law of contract remedies is usefully analyzed in terms of options because of the fact that in the

---

27. [Cite from cases declining to apply the penalty doctrine to take-or-pay contracts.]

28. But see cases involving earnest money deposits in real estate transactions, treated as presumptively non-refundable to the extent they comply with trade usage. E.g., Lawrence v. Miller, 86 N.Y. 131 (1881); Maxton Builders, Inc. v. Lo Galbo, 113 A.D.2d 923, 493 N.Y.S.2d 825 (2d Dep't 1985), aff'd 68 N.Y.S.2d 373, 382, 509 N.Y.S. 507, 512 (1986).

29. For a survey of arguments, see Aaron S. Edlin and Alan Schwartz, Optimal Penalties in Contracts, 78 Chicago-Kent L. Rev. 33 (2002).

ordinary case, breach of contract gives rise only to a claim of monetary damages, not to specific performance. As Holmes famously remarked:

> Nowhere is the confusion between legal and moral ideas more manifest than in the law of contract. Among other things, here again the so called primary rights and duties are invested with a mystic significance beyond what can be as-signed and explained. The duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it, and nothing else. If you commit a tort, you are liable to pay a compensatory sum. If you commit a contract, you are liable to pay a compensatory sum unless the promised event comes to pass, and that is all the difference. 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. . .[30]

Modern option terminology would restate Holmes' point by saying that the promisor holds a call option to buy her way out of the contract by paying a strike price equal to the value of court-awarded damages.

This option theory of contracts has been widely criticized over the years by commentators who reject Holmes's positivist account of legal obligation, and who have argued that breaking a promise is not merely a social inconvenience but a wrong.[31] Such critics have also attacked the modern-day economic interpretation of Holmes's theory, the so-called theory of efficient breach. But much of this criticism misses the mark, since not even the most uncompromising deontologist would object if the parties were explicitly to write Holmes' call option into their contract. If they did so, the party calling the option would not be breaking any promise; on the contrary 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 deontological wrong, accordingly, depends on whether the parties have contracted with the intention of incorporating such a call option into their agreement.

---

30. Oliver Wendell Holmes, The Path of the Law, 10 Harv. L. Rev. 457, x (1897).

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

As Shavell has observed, the legal regime simply sets up a default term for contracting when it established a rule of damages.[32] Under a regime where specific performance is the nominal default, as in some of the civil law systems of Europe, the parties can if they wish contract around this default rule by writing a payment option into their contract. Under the common-law rule that makes expectation damages the default, parties can, in principle if not in actual doctrine, escape the default by writing in a different damages term or even a term calling for specific enforcement. A deontologist or other autonomy-based 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.[33] The aspect of common-law doctrine most vulnerable to attack on autonomy grounds, accordingly, 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.[34] Indeed, an efficiency-based default rule arguably promotes

---

32. Steven Shavell, Damage Measures for Breach of Contract, 11 Bell J. Econ. 466 (1980); Steven Shavell, The Design of Contracts and Remedies for Breach, 99 Q.J. Econ. 121 (1984); Steven Shavell, Contracts, in The New Palgrave Dictionary of Economics and the Law, ed. Peter Newman. (New York: Stockton Press, 1998).

33. For elaboration of this defense of efficiency analysis against deontological critics, see Richard Craswell, Contract Law, Default Rules, and the Philosophy of Promising, 88 Mich. L. Rev. 489, (1989).

34. Recent statutory innovations, incidentally, may have the effect of expanding the parties' contractual freedom in this way. Proposed UCC §2-718 would allow the parties to liquidate damages by setting a formula as well as a fixed amount (a device that is not explicitly authorized in current law), proposed §2-715 allows sellers as well as buyers the option of collecting consequential damages in appropriate cases, and the official comments to proposed section 2-716 suggest that the parties' ex ante consent to specific performance should carry weight in an equity court's subsequent decision whether to issue such an award Cf. Stokes v. Moore [citation].

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.[35]

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 arms' 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.[36] 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 legal rules that set the points at which 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, in cases where there is a repudiation, the time at which the promisee's damages are measured determines whether or not she gets an option to speculate against the contract. If damages are measured at repudiation, as they are under UCC §2-713 when a seller repudiates the obligation to deliver goods, the parties' positions are closed out at that moment

---

35. "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 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, Contracts, supra, note x.

36. See generally Paul Mahoney, Contract Damages and Options Pricing, 24 J. Legal Stud.139 (1995)

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 UCC §2-708 when a buyer repudiates the obligation to purchase, the aggrieved party gets an option to enforce the contract if prices move against her, but to forego enforcement in the event that prices move in her favor. This additional option implies an increase in her expected damages following breach.[37]

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 — since 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, by influencing the effective strike price and the effective option term. For instance, the sales law doctrine of acceptance, by limiting the period of time in which a buyer is entitled to reject non-confirming goods, reduces the value of her option to reject. This limitation affects both the seller's incentive to perform, and the ex ante price of the goods that are traded.

Finally, the discussion so far has proceeded as if only one of the contracting parties holds an exit option but in a bilateral contract where both have promised to perform, both parties hold exit options. 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.[38] This paper abstracts from this symmetry in the succeeding discussion for the sake of analytic and

---

37. See Thomas H. Jackson, Anticipatory Repudiation and the Temporal Element of Contract Law: An Economic Inquiry into Contract Damages in Cases of Prospective Nonperformance, 31 Stan. L. Rev. 69 (1978) (explaining relationship between time of measuring damages and the expected value of damages payable, and criticizing the rule of 2-708 on the grounds that it gives the aggrieved party an unbargained-for and inefficient speculative opportunity.) Note that proposed revisions to Article 2-708, recently promulgated by the ALI and NCCUSL 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 revised §2-708(1)(b), comment 4 (2003).

38. Alexander J. Triantis and George G. Triantis, Timing Problems in Contract Breach Decisions, 41 J. Law & Econ. 163 (1998).

expositional simplicity, but the reader should keep in mind the pervasiveness of the option feature in contracting and in the law of contracts.

C. Some illustrative transactional puzzles

Before turning to the theoretical account of options that is presented in the next section of the essay, it is worth discussing some examples of how 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 it is helpful in motivating the discussion to follow to see some of the perhaps counterintuitive instances in which they are or are not used.

1. *Vacation deposits*

Anecdotal evidence suggests that hotels and motels operating in resort areas vary their deposit policy between high and low seasons.[39] In season, when demand is high and rate 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.

---

39. Such anecdotes are based on my own experience and conversations with others I have had in the course of preparing this essay. Revised versions of this essay may include citations to more systematic empirical surveys.

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 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. 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 (here, out of season) and inefficient performance when damages are too high (here, in season.)[40] In order to compensate for this efficiency loss, this pattern of contracting must have other advantages for the innkeeper. What are those advantages?

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

---

40. See Shavell, Damage Measures for Breach of Contract, and Shavell, The Design of Contracts and Remedies for Breach, supra, note x, and Robert D. Cooter and Melvin A. Eisenberg, Damages for Breach of Contract, 73 Calif. L. Rev. 1432 (1985).

efficient incentives to a customer deciding whether to cancel her reservation.[41] Innkeepers don't typically debase other terms of service during the high season, for example 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 don't. Rental car companies also experience periods of slack and tight demand, for instance, and they also could require a non-refundable deposit. The same is true of restaurants. Such companies do vary their prices in the form of weekend rates and early-bird specials, but they don't 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 non-refundable 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?

2. *Resale price maintenance and recovery of selling costs*

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 (RPM), has long been controversial in policy terms

---

41. Puzzles about why sellers don't 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, Jack L. Knetsch, and Richard Thaler, Fairness as a Constraint on Profit Seeking: Entitlements in the Market, 76 Amer. Econ. Rev. 728 (1986); Robert J. Shiller, Maxim Boycko, and Vladimir Korobov, Popular Attitudes Toward Free Markets: The Soviet Union and the United States Compared, 81 Amer. Econ. Rev. 385 (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.

and is currently illegal under U.S. antitrust law.[42] 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.[43]

   Defenders of a more laissez-faire antitrust policy, however, beginning with Lester Telser, have argued that resale price maintenance may be necessary in order to induce their wholesalers and retailers to invest adequately in selling expenses such as well-stocked showrooms and knowledgeable salespersons.[44] 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.[45] On this view, it is more cost-effective to restrict price competition, thus redirecting competitive pressure into non-price directions. Resale price maintenance should

---

42. Relevant case law in this regard includes Dr Miles Medical Co v John D Park & Sons, 220 U.S. 373 (1911) and U.S. v. Parke, Davis & Co. (1960). 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; and at the high-water mark of the practice, 43 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.

43. [Cites to antitrust authorities.]

44. Lester Telser, Why Should Manufacturers Want Fair Trade, 3 J.L. & Econ. 86 (1960). See also Robert H. Bork, The Antitrust Paradox: A Policy at War with Itself __ (1st ed. 1978) (elaborating and endorsing Telser's argument); Richard A. Posner, Antitrust Law: An Economic Perspective __ (1976) (same).

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

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, resale price maintenance 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 may or may not be enforceable as a matter of contract law, depending on the offer and acceptance rules followed by the local jurisdiction, but it is still an option in business terms.[46]

An analogous issue arises in the field of contract law in connection with the assessment of sellers' damages for lost profits following a buyer's breach. Under UCC §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; conventional wisdom holds that this is the case whenever the seller is unable to resell the goods that the buyer refused to take. But in response to the conventional wisdom, Victor Goldberg has argued that in assessing lost profit it is critical to take into account the up-front expenses the seller incurred in making the sale initially.[47] If the cost of such expenses is adequately covered by the expected price in the event that the contract is completed, on 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 implicit arrangement, including particularly the amount of any agreed deposit. The analysis is similar to the analysis of RPM because if Goldberg is right, and the retail price is set high enough to cover sales costs expended on customers who do not buy, then the

---

46. Some states have countermanded the common-law rule that advertisements are not offers by passing so-called bait-and-switch statutes, which require retailer to stock and to stand ready to sell a reasonable quantity of any goods being advertised. [Cites.]

47. Victor Goldberg, An Economic Analysis of the Lost-Volume Retail Seller, 57 S. Cal. L. Rev. 283 (1984)

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 Telser's defense of RPM and 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 don't automobile or electronics retailers 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 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 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 (since shopping imposes costs on the store even when one doesn't buy), and the subsequent decision whether or not to purchase (since 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, don't more retailers offer it?

II. A theoretical account of option design

The preceding discussion may persuade the reader that the topic of option contracts is important, but may raise suspicions that no single account of this transactional device is possible. In order to begin to address the analytical and transactional questions raised in the previous sections, accordingly, this section of the essay sets out a simple model of the option

contract and shows how the basic elements of such contracts relate to and can be traded off against each other.

A. Three essential terms: option premium, strike price, and option life

The model is a simple one; 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 up front 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, and the option life (denoted as T), which is the period of time for which the option remains open. A firm offer extended under UCC §2-205, for instance, would have a P equal to zero, a T equal to 90 days, and an S equal to whatever price at which the offeror stood ready to sell.

Similarly, a prepaid airline reservation which requires the passenger to charge a $400 ticket to her credit card at the time the reservation is issued, but which 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 he 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. (Actually, 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=$150 and the strike price S=$250.)[48]

As the airline example illustrates, that 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 above, 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]

---

48. Alternatively, we could characterize the ticket not as granting a call option on the right to keep this particular reservation, but an outright sale of the trip, bundled together with a put option in which the passenger held 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 flows 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 financial economists label as "put-call parity." See, e.g., Michael S. Knoll, Put-Call Parity and the Law, 24 Cardozo L. Rev. 61 (2002). Accordingly, we are free to choose one characterization or the other at our convenience.

49. In addition, the framework is general enough to encompass contracts that we would in common parlance describe as unconditional. A idealized unconditional sale of a widget for $1 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 can be escaped for period of time T that is very short but perhaps not zero.

 B. The relationship among option premium, strike price, and option life

In a competitive market or indeed in any arms-length exchange,, these three terms of the option contract will stand in a regular relationship with one another. other. 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 we denote as E(V), depends on the possible distribution of values V that the underlying good or service might take on, and 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, on the other hand, 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 one extreme example, 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 could be

---

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

the same as the optionee's value V or different from it, although in general the parties' values will be positively correlated for most ordinary commodities. In all cases where the optionee elects to exercise the option (i.e., in all cases where V>S), the optionor will lose 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 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 arms-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 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

---

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 (i.e., 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 chance 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 now and t_1 equals $p_0 + p_1(1-p_0)$.,

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 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 an 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

---

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 effect (call it the "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 (call this 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.

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.

that is extended; in compensation, 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 arms-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; as a general matter, however, the reader should understand that increases in interest rates or market volatility will generally have the same effect as increases in the time horizon T.[54]

In what follows, I also abstract from the complication that in bilateral contracts, there are two options at work simultaneously, in that the buyer can escape the contract by paying the seller's expectation damages, and the seller can symmetrically escape the contract by paying the buyer's expectation damages. To consider symmetric options would needlessly complicate the exposition of my argument, and once its basic intuition is understood, the reader should be able to extend its logic to the bilateral case in straightforward fashion.

C. Efficient option design

Given this functional relationship among P, S, and T, there are a continuum 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 (i.e, those with T=0 and S=∞, and some value of P) and a continuum of more or less equivalent option contracts

---

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.

(i.e, those with T>0 and S<∞, and a somewhat higher value of P). How, then, can and should the parties choose among these possible contracts?

Phrased at 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 tradeoff 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 arms-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 price might differ in actual transactions. Accordingly, the following section of this paper 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 section of the paper, it is useful to divide the various reasons why parties might wish to choose one option over another into three categories: efficiency considerations, non-efficiency 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 non-efficiency considerations such as a seller's desire to gain market power or to foreclose competition, non-enforcement and perhaps affirmative regulation of contracting behavior is in order. We start with non-efficiency considerations: not because they are necessarily more likely to motivate such transactions, but because they are somewhat more straightforward to explain and analyze.

A. Non-efficiency 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.[55] As a result, choices between 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 attitudes toward those variables disproportionately. For example, one fairly robust finding of the behavior 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 optionees, the up-front payment P may be more salient than the subsequent 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 backload the total

---

55. See generally Amos Tversky & Daniel Kahneman, Judgment Under Uncertainty: Heuristics and Biases, in Judgment Under Uncertainty 3 (Daniel Kahneman, Paul Slovic & Amos Tversky eds., 1982); Christine Jolls, Cass R. Sunstein & Richard Thaler, Behavioral Responses to Law and Economics, 50 Stan. L. Rev. 1471, 1479 (1998).

payments under the option, lowering P and raising S past the point that would otherwise be transactionally efficient. A similar effect would be produced 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 "hyperbolic discounting."[56]

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 unequally subject to it, because just as it is especially salient for the optionee to have to make an unconditional payment of P and especially tempting to put off the pain by backloading 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 till some future time at which he might or might not receive S. In cases where one party is more sophisticated than the other or more able to invoke procedures that rationalize its decisionmaking, however, we would expect the sophisticated actor to 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

---

56. See David Laibson, Golden Eggs and Hyperbolic Discounting, 112 Q.J. Econ. 443 (1997); George Loewenstein & Richard H. Thaler, Intertemporal Choice, in Thaler, The Winner's Curse: Paradoxes and Anomalies of Economic Life 182, 185-86 (1992). A less formal account of the phenomenon can be found in Jon Elster, Ulysses and the Sirens: Studies in Rationality and Irrationality 71 (1979).

possible.[57] 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 the direction of 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 subsection 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, 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 competitive entrant who enters the market after the option premium has been paid. Setting up exchanges in the form of prepaid options, or more generally, front-loading payment to a greater extent than would be otherwise profitable, may be a way of excluding competitors from the market. Several economic commentators have show how liquidated damages clauses can be used in this way[58]; and the traditional non-Chicago antitrust account of resale price maintenance is also based in part on this concern.[59] Recent work on consumer

---

57. See Avery Katz, Your Terms or Mine: The Duty to Read the Fine Print in Contracts, 21 Rand J. Econ. 518 (1990); Avery Katz, The Strategic Structure of Offer and Acceptance: Game Theory and the Law of Contract Formation, 89 Mich. L. Rev. 215, x (1990).

58. See Philippe Aghion & Patrick Bolton, Contracts as a Barrier to Entry, 77 Amer. 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, 284-89 (1992); Kathryn E. Spier & Michael D. Whinston, On the Efficiency of Privately Stipulated Damages for Breach of Contract: Entry Barriers, Reliance and Renegotiation, 46 Rand J. Econ. 180, 183-86, 195-96 (1995).

59. [Cites.]

"lock-in" in imperfectly competitive markets has also shown that this could be a more general phenomenon.[60]

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 be in an informational position to judge whether anticompetitive effects on third parties outweigh the benefits of option arrangements to the contracting parties. More practically, 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.[61] 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, accordingly, the consumer will have little incentive to to challenge the anticompetitive clause.

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. Much creative work in the tax area in recent years, both by practitioners and by scholars, has focused on the arbitrage possibilities

---

60. 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 (1995).

61. Ibid.

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. And a third example is provided by the privileged treatment of leases under UCC Article 9 and the federal bankruptcy code, where 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 a sale, subject to UCC 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. This paper 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 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 such potential reasons; here I classify them into 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, although it is worth taking account of recent normative critiques of the standard efficient markets story.[62]

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 asset; for instance, the asset may have better diversification value for one of the parties given the other assets in its portfolio. It may also be that one party is relatively immune to risk because of its size or the frequency with which it trades, so that it is profitable for the large party to insure the smaller one. Such a rationale applies in purely financial settings and context of real contractual exchange alike, and it may explain why large sellers are willing to offer requirements contracts (and relatively large buyers such as grain elevators are willing to offer output contracts) at a relatively favorable price.

To the extent that risk aversion and the value of insurance vary over time and over particular ranges of potential asset values, furthermore, 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, given other orders she has on file, 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 (i.e., to the point where the purpose of the contract would be frustrated), but she would not be willing to insure the risk that the buyer's valuation falls only slightly (so that the exchange becomes marginally unprofitable). Optimal risk allocation would then require the buyer to bear the former portion of risk, but not the latter.

---

62. [Cites.]

2. *Ex post incentives*

Risk and speculative motives help explain the existence of options trading in both financial and non-financial 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, in which event the underlying asset and all the costs and benefits associated with it will belong to him. It is on this precise logic that executive stock options are claimed to improve the incentives of corporate officers to act in ways that increase the market value of the firm.[63]

The difference between an option and an outright sale in this regard is that at the same time, the seller of the option retains the incentives to invest in actions that increase or are valuable in the event of downside risk.[64] Thus, options will be economically advantageous 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 the

---

63. [Cite to executive compensation literature, but also to articles questioning whether this is the main motivation for such arrangements.]

64. This division of incentives is a feature of both put and call options, by the way, with the only difference being the identity of the residual owner in the event that the option is not exercised.

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

The incentive effects of options have been has recognized and applied in specific contexts by a number of contributors to the law-and-economics literature, but its general implications for contract law have not been drawn out. For instance, in his well-known article on the problem of providing multiple parties to a transaction with simultaneous incentives to take appropriate precautions against loss (what he calls the problem "double responsibility on the margin"), Cooter identifies the call option as one of the standard legal devices that can be used to give such double responsibility. 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 does not exercise, and the optionee will want to take optimal precautions because she anticipates that once the optionor has taken precautions, she will then want to exercise.[65] In a similar spirit, Nöldëke and Schmidt have shown how options can be used to solve the problem of efficient promisee reliance[66], Edlin and Hermalin have shown how

---

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

66. Georg Nöldeke and 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).

options can be used to promote efficient relational investment[67], and 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.[68]

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 its terms 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. 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 chances of exercise and thus 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 or another is exactly balanced by the marginal value of enhancing the optionee's incentives.

---

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

68. See Masten and Crocker, supra, note 23.

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, in interpreting whether ambiguous contracts are options or not (e.g., in distinguishing between an offer for a bilateral contract and an offer for a unilateral one), courts should keep in mind the substantive reasons why parties would or would not want to use options. Conversely, to the extent that the use of options is driven by inefficient motivation 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. But what makes these problems harder (and more interesting) is that there is an additional set of reasons for using option contracts that do not clearly argue in favor of or against an efficiency interpretation. In many cases, the contracting parties may wish to use options for reasons that enhance their private gains from exchange, but that have an ambiguous effect on social efficiency generally. 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 subsection, 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 credibly to convey 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 one, 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.[69]

Granting an option to purchase is another way for such a seller to signal product quality; and 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=∞.)

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 non-refundable 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 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.

---

69. See, e.g., Sanford Grossman, The Informational Role of Warranties and Private Disclosure about Product Quality, 21 J. Law & Econ. 461 (1981); George L. Priest, A Theory of the Consumer Product Warranty, 90 Yale Law Journal 1297 (1981).

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 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 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 his reliability, for instance, part of this gain comes at the expense of his competitors. After the signal, this seller will be perceived as more reliable than average, but the others in his 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 thru signaling is not necessarily efficient, in contrast to ordinary price competition, which is generally efficient. Sellers who signal thru 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.[70]

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.[71] Because we do not otherwise seem 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

---

70. The classic formulation of this argument can be found in A. Michael Spence, Job Market Signaling, 87 Q. J. Econ. 355 (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).

71. See, for example, Severin Borenstein, The Economics of Costly Risk Sorting in Competitive Insurance Markets, 9 International Rev L & Econ 25, 32-33 (1989) (demonstrating ambiguity of optimal policy response in insurance setting).

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 price 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 low option premium and high strike price, and another with high option premium and 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 non-refundable 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 till the flight leaves the gate; the latter are cheaper but can only be purchase some time in advance of travel.

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

unconditional payment) and high S (i.e.,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.[72]

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 their actual marginal cost. Thus they manage to cover their overhead and increase 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.[73] In other cases, however, price discrimination can reduce market efficiency.[74] 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

---

72.   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 (1992).

73. See Walter Oi, A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly, 85 Q.J. Econ. 77 (1971); Richard Schmalensee, Monopolistic Two-part Pricing Arrangements, 12 Bell J. Econ. 445 (1981).

74. See generally Hal R. Varian, Price Discrimination, in Handbook of Industrial Organization, Richard L. Schmalensee and Robert D. Willig, eds. (Elsevier, 1989).

opposed to norms of distributional fairness or equal treatment, I would conclude that 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 general. Given that, there is 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' interest 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 non-refundable 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. But conversely, if other hotels allow cancellation without penalty, there will be more rooms available at the last minute, thus leading more customers to be willing 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 that there is a network externality that results from the fact that some minimum number of transactions are 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 path of past

behavior.[75] Private attempts to coordinate pricing policies could improve efficiency, but they could also lead to cartel-like behavior and price-fixing. Thus this motivation is a mixed one, possibly justifying regulatory intervention to discourage cartelization or to help parties coordinate on an equilibrium that they cannot reach privately.

IV.   Applications to legal and transactional problems

The previous sections of this paper 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.[76] Nonetheless, the foregoing discussion does suggests some basic heuristic principles that might be used to assist both private planning and public regulation. As an illustration of these principles, accordingly, in this section I return to a number of the doctrinal and transactional puzzles that motivated the original inquiry. 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, since I have undertaken no detailed empirical surveys in this regard, but rather as a guide to and recapitulation of the basic lessons of the paper.

---

75. See Michael Katz and Carl Shapiro, Network Externalities, Competition and Compatibility, 75 Amer. Econ. Rev. 424 (1985); W. Brian Arthur, Self-Reinforcing Mechanisms in Economics, in Philip W. Anderson, Kenneth J. Arrow and David Pines (eds.) The Economy as an Evolving Complex System, Addison Wesley (1988).

76. Citation omitted to preserve anonymity of submission process.

A. Distinguishing between option contracts and contracts generally

As we saw in section 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, accordingly, it is necessary to identify some substantive reasons for treating options differently and some criteria for determining which contracts qualify for treatment as options and which do not. Since the substantive reasons for differential treatment are presumably as varied as the specific doctrines that provide for it, this paper does not attempt to provide an exhaustive list of such explanations. Instead, I focus for illustrative purposes on two specific legal applications: contractual penalties and requirements contracts.

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, since penalties themselves can be used for various purposes: to motivate specific investment, to signal information, to facilitate price discrimination, or to exclude subsequent competitors.[77] 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 not clear why one would want to distinguish between options and penalties from a public regulatory viewpoint. However, if the main explanation for restricting contractual penalties is bounded rationality, 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.[78] Casting a penalty clause in option clothing, however, probably makes

---

77. See Edlin & Schwartz, supra, note x, for a survey of these explanations.

78. 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, 725 (2000).

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 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 have the effect of penalties but that also are 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. Such contracts typically provide for non-refundable down payments in the amount of 25% of the overall purchase price, even though 25% might well be an overestimate in some cases of the developer's lost expectation should the purchaser back out of the deal. Of course, it might also be an underestimate, as this market is a particularly volatile one; in New York City in particular, the luxury condominium market has experienced short-run price swings of 25% or greater several times in the last twenty years.

In this context, the non-refundable 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 and 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 cancelling buyer, its other economic benefits may justify exempting

it from the usual common-law rule prohibiting contractual penalties; and some important commercial jurisdictions have provided just such an exemption.[79] 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*

   As indicated in section I, 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 the symmetric output contract 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 under UCC 2-306 makes such contracts enforceable, with the complication that the optionee's discretion how much to buy must be exercised in good faith and in reasonable proportion to estimated or previously traded quantities. 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; while 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

---

79. See, e.g., Lawrence v. Miller, supra, note _; Maxton Builders, Inc. v. Lo Galbo, supra, note __.

hence determining whether a given exchange should be structures as a requirements or a fixed-quantity contract) requires attention to the various considerations outlined in the framework of the previous sections.

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 requirement 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 value of which depends on the seller's performance and tender of the underlying goods.

On the other hand, requirements contracts have the disadvantage that the seller, in order 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 variable cost and strike price 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 the variable cost but falls short of the exercise price. Absent some other device for eliminating such inefficiencies such as ex post renegotiation, accordingly, 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*[80], which involved a dispute over express quantity terms in a contract between two chemical companies. Columbia, the buyer and defendant, had for several years sold

---

80. 451 F.2d 3 (4th Cir. 1971).

significant amounts of nitrogen to Royster, which used the nitrogen as an input to the manufacture of chemical fertilizer. 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 31,000 tons of phosphate each year for three years to Columbia, with an option to extend the term. The contract also contained a standard merger clause that excluded any verbal or prior understandings relating to the contract.

In 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. 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 she turned out not to need the goods. Essentially, Columbia claimed that an apparently explicit fixed-quantity contracts was really a requirements contract. It offered to prove this claim using 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.

The *Columbia Nitrogen* case is generally regarded as the leading authority on UCC §2-202(a), 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. Its outcome turned on the application of §1-205, which provides that "[t]he express terms of an agreement and an applicable course of dealing or usage of trade shall be construed wherever reasonable as consistent with each other; but when such construction is unreasonable express terms control both course of dealing and usage of trade and course of dealing controls usage of trade." 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.

The court's decision in the case has been criticized by various commentators 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.[81] Whether this criticism is merited in the particular instance cannot be definitively determined on basis of the information available in the court papers, but the framework of the previous sections 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, although 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 parties' 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 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.

---

81. See e.g., Lisa Bernstein, ____; Victor Goldberg, Columbia Nitrogen v. Royster: Do as They Say, Not as They Do, unpublished manuscript, 2003; Richard Craswell, Do Trade Customs Exist?, in The Jurisprudential Foundations of Corporate and Commercial Law (Kraus & Waltz eds., 1999); Southern Concrete Servs., Inc. v. Mabelton Contractors, Inc., 407 F. Supp. 581 (N.D. Ga. 1975).

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 was 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. Both 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.

B. Determinants of two-part pricing

In the first section of this paper, we also identified a number of transactional contexts in which parties front- or back-load payment in a way that appears to create ex post inefficient incentives for performance. Given the framework we have now developed, we are now in a position to identify some explanations for these arrangements as well, which include, inter alia, firm offers, non-refundable deposits, and resale price maintenance.

1. *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.[82] 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, and in transactions involving new or untested products, where other signaling or commitment mechanisms are unavailable.

But conversely, 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 record-keeping 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

---

82. For a formal model illustrating these incentives, see Avery Katz, Your Terms or Mine: The Duty to Read the Fine Print in Contracts, 21 Rand J. Econ. 518 (1990).

parties together and thus reduce the chances of opportunistic behavior.[83] It may be this type of consideration that explained the pattern of behavior shown in the *Columbia Nitrogen* case discussed in subsection A.2 above — even putting aside the questions of whether that pattern should rise to the level of a legally binding trade usage, or whether such a usage should have been interpreted to extend to the specific contract at issue in the case.

   2. *Variations in deposit policies.*

As we saw earlier, the apparent seasonal pattern of deposit policies is difficult to explain in terms of ex post incentives, because 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, since 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 three explanations for this pattern that do make sense, the first of which is bounded rationality. If vacation customers don't pay sufficient attention to the risk that they might need to cancel and hence lose their deposit, but do pay a lot of attention to 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, and will 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

---

83. [Cite Ernst Fehr and others on how repeated gift exchange can provide superior incentives in environments
     where some fraction of the population is altruistic; also standard references on gift exchange.]

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.[84]

A second explanation, already mentioned above, might be coordination. 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. Such an explanation could 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, non-refundable deposits could be a form of price discrimination, on the model of non-refundable 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 activities. 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 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

---

84. See generally Samuel Issacharoff et al., Regulation for Conservatives: Behavioral Economics and the Case for 'Asymmetric Paternalism', 151 U. Pa. L. Rev. 1211 (2003).

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 in 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 grounds that the latter have more elastic demand.)

   3. *Resale price maintenance and free provision of pre-sale services.*

   Finally, consider the example of resale price maintenance As we have seen, this device resembles firm offers in that retailers charge a low or zero up-front price for the sales and informational services that are necessary for a potential buyer to get in a position to decide whether to buy. The main difference is that in the RPM context there is no explicit option offered top the customer (except to the extent that contract law or an applicable bait-and-switch statute requires the retailer to stand ready to sell.) But given that the up-front services have positive 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. Why would firms want to give away sales services for free, especially in a competitive environment where some customers will consume the sales services and then purchase elsewhere from a discounter?

   As we just observed in our discussion of firm offers, up-front cover charges won't work for new or unfamiliar products because buyers are uncertain about whether they will want to purchase. 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 only works if those buyers willing to pay a higher price for customer service and sales-related overhead are also

those who would only be willing to pay a lower price for the underlying product. (This is a straightforward application of the classic product tying model, here applied to a tie between the underlying commodity and its associated sales quality.)[85]   But such a correlation seems implausible if both customer service and the underlying product are normal goods. Absent a plausible alternative explanation, accordingly, we may tentatively conclude that the traditional antitrust critics of RPM are right, and the main motivation for the practice is to exclude competition or to facilitate a retailer's cartel.

V. 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, accordingly, or in interpreting them ex post, it is essential explicitly to consider 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 in order 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 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.

---

85. See W. James Adams and Janet Yellen, Commodity Bundling and the Burden of Monopoly, 90 Q.J. Econ, 475 (1976).