

NBER WORKING PAPER SERIES

ECONOMIC ANALYSIS OF
CONTRACT LAW

Steven Shavell

Working Paper 9696
<http://www.nber.org/papers/w9696>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2003

Research support from the John M. Olin Center for Law, Economics, and Business is gratefully acknowledged. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

©2003 by Steven Shavell. All rights reserved. Short sections of text not to exceed two paragraphs, may be quoted without explicit permission provided that full credit including ©notice, is given to the source.

Economic Analysis of Contract Law
Steven Shavell
NBER Working Paper No. 9696
May 2003
JEL No. D00, D8, K00, K12

ABSTRACT

Contract law governs agreements between parties. This paper contains the chapters on contract law from a general, forthcoming book, *Foundations of Economic Analysis of Law* (Harvard University Press, 2003). Chapter 13 presents an overview of the subject. Chapter 14 is concerned with contract formation, that is, with the process through which parties find contracting partners, with aspects of contract negotiation, and with the rules governing when an arrangement between parties becomes legally recognized as a contract. Chapter 15 considers at length an important type of contract: the contract to produce something. Chapter 16 is concerned with two other types of contract: the contract for transfer of possession of something that already exists (such as land or a painting), and donative contracts.

Steven Shavell
Harvard Law School
1575 Massachusetts Avenue
Hauser Hall 508
Cambridge, MA 02138
and NBER
shavell@law.harvard.edu

Table of Contents

Economic Analysis of Contract Law

Chapter 13. Overview of Contracts

1. Definitions and Framework of Analysis
2. Contract Formation
3. General Justifications for Contracts and for Their Enforcement
4. Incompleteness of Contracts
5. Interpretation of Contracts
6. Damage Measures for Breach of Contract
7. Specific Performance as the Remedy for Breach of Contract
8. Renegotiation of Contracts
9. Legal Overriding of Contracts
10. Extra-Legal Means of Contract Enforcement

Chapter 14. Contract Formation

1. Search Effort
2. Fundamental Rule of Recognition of Contracts: Mutual Assent
3. Offer and Acceptance
4. Fraud
5. Mistake
6. Information Disclosure
7. Duress

Chapter 15. Production Contracts

1. Completely Specified Contracts
2. Remedies for Breach and Incomplete Contracts
3. Reliance
4. Renegotiation

Chapter 16. Other Types of Contract

1. Contracts for Transfer of Possession
2. Donative Contracts

Summary Table of Contents of
Foundations of Economic Analysis of Law

(forthcoming 2003, Harvard University Press)

Chapter 1. Introduction

Part One. Accident Law

Chapter 2. Liability and Deterrence: Basic Theory

Chapter 3. Liability of Firms

Chapter 4. Extensions of the Basic Theory

Chapter 5. Liability, Risk-bearing, and Insurance

Chapter 6. Liability and Administrative Costs

Part Two. Property Law

Chapter 7. Definition, Justification, and Emergence of Property Rights

Chapter 8. Division of Property Rights

Chapter 9. Acquisition and Transfer of Property

Chapter 10. Conflict and Cooperation in the Use of Property: The Problem of Externalities

Chapter 11. Public Property

Chapter 12. Property Rights in Information

Part Three. Contract Law

Chapter 13. Overview of Contracts

Chapter 14. Contract Formation

Chapter 15. Production Contracts

Chapter 16. Other Types of Contract

Part Four. Litigation and the Legal Process

Chapter 17. Basic Theory of Litigation

Chapter 18. Extensions of the Basic Theory of Litigation

Chapter 19. General Topics on the Legal Process

Part Five. Public Law Enforcement and Criminal Law

Chapter 20. Deterrence with Monetary Sanctions

Chapter 21. Deterrence with Nonmonetary Sanctions

Chapter 22. Extensions of the Theory of Deterrence

Chapter 23. Other Functions of Sanctions: Incapacitation, Rehabilitation, Retribution

Chapter 24. Criminal Law

Part Six. General Structure of the Law

Chapter 25. The General Structure of the Law and Its Optimality

Part Seven. Welfare Economics, Morality, and the Law

Chapter 26. Welfare Economics and Morality

Chapter 27. Implications for the Analysis of Law

Chapter 28. Income Distributional Equity and the Law

Chapter 29. Commonly-Raised Questions about Economic Analysis of Law

Economic Analysis of Contract Law

(part of Foundations of Economics Analysis of Law)

Steven Shavell

©2003. Steven Shavell. All Rights Reserved.

Chapter 13

OVERVIEW OF CONTRACTS

This chapter presents an overview of contracts; it is concerned with the definition of contracts, the basic justifications for their existence, and with important aspects of contractual practice and of the law of contracts. Subsequent chapters will deal in greater detail with certain aspects of contract law and with particular types of contracts.¹

1. Definitions and Framework of Analysis

1.1 Basic definitions. By a *contract* I mean a specification of the *actions* that named parties are supposed to take at various times, generally as a function of the *conditions* that hold. The actions typically pertain to delivery of goods, performance of services, and payments of money, and the conditions include uncertain contingencies, past actions of parties, and messages sent by them. For example, a contract might state that a photographer should take pictures at a wedding on February 1st, that the buyer should pay the photographer \$1,000 within a week of the wedding, that the buyer may cancel if he notifies the photographer by January 1st, and that the photographer may cancel if he becomes ill. It is apparent that because the notions of actions and of conditions are broad, the conception of a contract is very broad.

A contract will be said to be *completely specified* (or simply *complete*) if the list of conditions on which the actions are based is *explicitly exhaustive*, that is, if the contract provides literally for each and every possible condition in some relevant universe of conditions. In a contract for a photographer to take wedding photographs, suppose that the universe of conditions is everything that could happen to the photographer (becoming ill, receiving an offer to take photographs at another wedding the same day) and everything that could happen to the wedding couple (becoming ill themselves, breaking off their engagement). A completely specified contract would then have to include an explicit provision for each of these possible conditions pertaining to the photographer and to the wedding couple. Although, as we will discuss, contracts are far from completely specified in reality, the concept of a complete contract will be helpful for clarifying our thinking about contracts. Moreover, we will sometimes want to simplify by assuming that the universe of relevant conditions is small (we might suppose that it

¹For general introductions to economic analysis of contract law, see for example Posner 1998, chapter 4, and Shavell 1998.

is only: the wedding photographer either becomes ill, or he does not), in which case we can well imagine a completely specified contract.

A contract will be said to be *incomplete* if it is not completely specified, which is to say, if the contract does not list explicitly all the possible conditions in the universe under consideration. For example, a contract that reads “Photographer shall take wedding pictures on March 14,” would obviously be incomplete because it does not list any conditions; so would a contract that says “Photographer shall take wedding pictures on March 14, unless the photographer develops appendicitis,” as this contract mentions only the single condition of appendicitis in the universe of possible conditions. Note that although these two contracts are incomplete, they do *implicitly* provide complete instructions for what the parties are to do under all conditions. The contract that says simply that the photographer shall take wedding pictures on March 14 implies that he should take the pictures under all conditions, even though it does not state that explicitly. Thus, according to the definition we are employing, an incomplete contract may well provide a complete set of instructions by implication.²

An incomplete contract that does not provide a complete set of instructions explicitly *or* by implication is said to have *gaps*. For example, suppose that the wedding photography contract states that if the weather is sunny, the ceremony will be held in the backyard and a video camera should be used, and that if there is rain and the ceremony is held inside the house, only still photographs need to be taken. This contract does not state explicitly or imply what is to be done if the weather is cloudy; thus, it has a gap.

1.2 Mutually beneficial contracts. A contract is said to be *mutually beneficial* or, in the language of economics, *Pareto efficient*, if the contract cannot be modified so as to raise the well-being -- the expected utility -- of each of the parties to it. We would suppose that contracts would tend to be mutually beneficial: If a contract can be altered in a way that would raise the expected utility of each party, we would think that this would be done. For example, suppose that the wedding contract states that the photographer should appear at 10 AM in the morning, but that an alternative contract under which he would arrive at 9 AM and would be paid an additional \$100, is preferred both by the wedding couple and by the photographer. Then the first contract would not be mutually beneficial, and we would expect the modification of the contract for earlier arrival of the photographer and higher payment to be made.

1.3 Enforcement of contracts. Contracts are assumed to be enforced by a court, which generally will be interpreted to be a state-authorized court. However, in many respects, an entity other than a state-authorized court -- a decisionmaking body within a firm, a trade association or a religious group, or an arbitration organization -- could serve as a tribunal and enforce contracts. Moreover, reputation and related factors may also serve to some degree to enforce

²The use of the term “incomplete contract” in the economics literature is consistent with the definition I have given in this paragraph. In the economics literature, a contract is called incomplete if some variable on which the contract could depend (and typically would be valuable to include in the contract) is not included. For example, in the contract with the photographer, the contract could be imagined to depend on the photographer’s effort (how well he circulated among guests), but if it did not depend on the photographer’s effort, the contract would be considered incomplete. This contract would also be an incomplete contract according to my definition, for the photographer’s effort level is a condition (a past action) on which the contract could in principle depend. Note too that in this example, as in many examples of incomplete contracts studied in the economics literature, an incomplete contract does provide, by implication, a complete set of instructions for the parties to the contract.

contracts. These extra-state means of enforcement will be discussed in section 10 below.

Contract enforcement involves the functions and actions of courts. Typically, courts act only when parties to contracts decide to come before them. Several general functions of courts should be mentioned.

A basic function of courts is to decide about *contract formation*, that is, when a valid contract has been made.

Given that a contract has been properly made and is deemed valid, courts must often engage in contract *interpretation*, notably, fill gaps in contracts and resolve ambiguities.

Another function of courts concerns *breach* of contract. Courts must decide when breach has occurred and impose sanctions or “remedies” for breach. Courts may impose two different types of sanctions for breach of a contract by a party to it: They may force a party in breach to pay money *damages* to the other, or they insist that the contract be performed in a literal sense (for example, require land to be conveyed, as stipulated in the contract), which is to say, insist on *specific performance* of the contract.

Finally, courts may also decide to *override* a contract. That is, even though a contract was properly formed and is not invalid on that count, and has not been breached, the court may refuse to enforce it.³

1.4 Social welfare and the welfare of contracting parties. It will generally be assumed that the goal of courts is to maximize social welfare. This will usually mean that courts act to further the welfare of the parties to the contract, for they will ordinarily be the only parties affected by the contract. If, however, other parties are affected by a contract, then the well-being of these parties outside the contract will also be assumed to be taken into account by the court.

2. Contract Formation

As mentioned, one of the basic functions of courts in relation to contracts is to decide when contracts are recognized as having been formed -- that is, when they are deemed valid and will be enforced. Several aspects of the law of contract formation will illustrate its significance.

One dimension of the law of formation concerns the ease with which parties can determine whether contracts will be legally recognized. If legal recognition of contract formation is based on a clear sign of agreement from each party -- such that each is easily able to know when there has been mutual assent and a contract has been formed -- then two essential benefits will follow. First, because parties have the ability to make contracts and to know immediately that their contracts will be enforced, the parties will be able to benefit without any delay from undertaking value-enhancing activities (such as hiring workers and purchasing materials for construction). Second, because parties can avoid making contracts by not making the sign of agreement that would lead to recognition of a contract, they will not be afraid to engage in search for partners, to seek information about possible contracts, and to negotiate about them, for parties will not fear being said to be in a contract that they do not want.

Another aspect of rules regarding contract formation concerns whether one of the parties

³From a formal point of view, all of these judicial tasks may be regarded as involving application of legal rules in a broad sense: Suppose that a legal rule is any function whose domain is the pair constituted of the contract of the parties (or, more exactly, various initial, observable statements they have made to each other) and certain subsequent events (observable actions of parties, messages, contingencies), and whose range is a set of actions of the court (such as declaring a contract to be formed, naming payments the parties must make, other actions of the parties, revision of contract terms).

to a possible contract was under duress or in an emergency situation. If, despite otherwise proper signs of agreement to a contract, a contract is not recognized because of duress or emergency, two socially beneficial and one detrimental consequence follow. First, socially undesirable effort will not be spent in order to place certain parties in problematic situations in which, due to duress or emergency, they would be led to make contracts on terms very favorable to those that put them there. Second, parties will obtain a kind of implicit insurance against having to pay high prices if they find themselves in an emergency situation. But third, the incentives of potential contracting parties to help those in bad straits will be dulled.

An additional aspect of the law of contract formation involves the information that is divulged by potential parties to contracts. The more information that parties are required to reveal, the better the matching of partners to each other and the more efficient the actions of parties once they make contracts. However, legal obligations to reveal information generally dilute incentives to acquire information, which can be socially undesirable.

These and other aspects of the law of contract formation will be considered in chapter 14.

3. General Justifications for Contracts and for Their Enforcement

3.1 Why contracts are made. A basic question about contracts is why parties should want to make them, that is, why they should want to make plans with each other. Several important reasons may be offered.

An obvious warrant for contracts involves the future *provision of goods and services*. It is often the case that one party will want a good or service in the future, and that another party can supply the good or service, giving rise to the mutual desirability of a contract. It should be noted, however, that contracts will not be necessary for the future supply of goods or services if a well-organized market in them exists, for then a future need can be met on the spot. (If I will want food for dinner a month from now, I do not need to make a contract to get it, for I will be able to purchase the food at that time.) Thus, it is mainly for custom or specialized goods and services, not readily available on markets, that production contracts may be beneficial.

Another reason for contracting is the mutually beneficial *reallocation* or *sharing of risks*. Insurance contracts, whereby risk-averse insureds pay premiums and are covered against risk by a risk-neutral insurer, are a primary example of agreements made for this reason, and other examples in which risk allocation is a primary feature, such as partnership agreements to divide total profits, abound.

A third reason for contracts concerns *differences of opinion* about subsequent events. When transactions in securities or in durable assets occur, the explanation is often, at least in part, that the buyer and seller have different beliefs about their future prices; when bets occur, the explanation is typically that the two sides hold different beliefs about the likelihood of the bet event.

A fourth general reason for contracting involves altering the *timing of consumption*. When individuals borrow or lend, they are making mutually beneficial arrangements in which their temporal patterns of consumption are altered.

Of course, more than one of these reasons may apply in a given case. For instance, a cost-plus contract for the provision of a good may achieve not only provision of the good, but also beneficial risk-sharing (suppose the producer is averse to risk).⁴

⁴Although the reasons for contracts discussed in this section are, I believe, the primary ones, others exist. For Chapter 13 – Page 4

3.2 Why enforcement is desired. Given that there are reasons for parties to make contracts, to make plans for future actions with each other, why do parties want their contracts enforced by courts? That is, why might contracts be broken in the absence of enforcement by an outside party, and why, exactly, would such contract breach be undesirable for the parties?

There are three general answers to these questions. The first is that without enforcement, a party would be able to *appropriate funds* that had been paid before contract performance, generally rendering the contract unworkable. For example, because borrowers would be able to keep what they had been lent and would not be forced to repay loans, loans would become impossible without contract enforcement, and because insurers would be able to keep premiums and would not be made to cover losses, insurance would become impossible without enforcement. Thus, most financial contracts, bets, and risk-sharing arrangements would become unworkable. And any contract other than one in which there is a *simultaneous* exchange of money for goods or services would also become unworkable in the absence of contract enforcement.⁵

The second general reason for parties' desiring contract enforcement is that otherwise a party *might not deliver the promised good or perform the promised service*.⁶ A party who has promised a good or a service may find that another, better opportunity has arisen, that costs of performance have increased more than expected, and the like. If so, and if negotiation with the other party to the contract would be inconvenient or unlikely to succeed smoothly, the promisor might decide not to perform. If there is failure to perform even though performance would be best because its value exceeds its true cost, then the value of the contractual arrangement is diminished for the parties. Such reductions in the value of contracts can be avoided if contracts are enforced.

The third reason for enforcement is that without enforcement, the *price cannot be fixed in advance*, which is to say, *price holdup* might occur -- a party might *bargain opportunistically* about the price of a transaction -- reducing the value of the contract or discouraging the making of it altogether. To illustrate, consider a buyer who wants a custom desk that would be worth \$1,000 to him and would cost \$700 for a seller to produce. In the absence of contract enforcement, the buyer and the seller cannot fix in advance the price that the buyer will pay for the desk at the time of delivery. (Note, the buyer will not pay the seller in advance, for the seller could then walk away with what he receives.) The buyer and the seller will agree on a price only *after* the seller makes the desk, and the buyer will at that time pay the seller in a simultaneous exchange of money for the desk. But at that point, the seller's production cost will be sunk and he will be vulnerable to holdup; the situation will be that he has a desk that, being custom-made,

example, parties may want to warrant the quality of goods sold on spot markets and thus want contracts for that purpose, or parties may want to induce certain behavior on the part of the recipients of gifts and thus want to make donative contracts (on which, see section 2 of chapter 16).

⁵ The enforcement of a contractual obligation to pay money, such as to repay a loan, might come, at least in part, under the head of tort law (to prevent "conversion" of assets) or criminal law, rather than contract law; but this is not of significance for our purposes, for what matters is that some form of legal enforcement of contractual obligations is needed.

⁶ Such failure might occur even though simultaneous exchange of money for the good or service is possible, so the present reason for enforcement is different from that of the last paragraph.

will have little or no alternative value.⁷ The outcome of bargaining between him and the buyer might thus be a price lower than the seller's cost of \$700; say the price is \$500. If so, and the seller anticipates receiving only the \$500 price, he will not produce the desk. This is true even though production and sale at a price between \$700 and \$1,000, such as \$800, would be mutually beneficial for the seller and the buyer. More broadly, the problem of holdup at the stage of negotiation for performance and for payment will result in all manner of underinvestment in the contractual enterprise.⁸

4. Incompleteness of Contracts

Having defined contracts and given general reasons why they are made and enforced, let me now examine the nature of contracts themselves. An aspect of contractual practice that will be seen to be of considerable importance is that contracts are significantly incomplete. Contracts typically omit all manner of variables and contingencies that are of potential relevance to contracting parties. A contract to take pictures at a wedding would be likely to fail to include many contingencies that might make it difficult or impossible for the photographer to perform, as well as many circumstances that would alter the couples' desire for photographs or for other types of records that they want to be made of their wedding.

There are several types of reasons for the incompleteness of contracts, that is, for why parties find it in their mutual interest to leave contracts incomplete. One category of reasons concerns the effort and cost of anticipating possible contingencies, bargaining about their resolution (given that they are anticipated), and then describing them adequately in contracts. In particular, parties will tend not to specify terms for low probability events, because the expected loss from this type of exclusion will be minimal, whereas the cost of including the terms would be borne with certainty. For example, it might take 15 minutes to discuss and include a term about what to do if the photographer is involved in a car accident on the way to the wedding, but if the likelihood of such an event is quite low, it will not be worth the parties' while to include a provision for such an outcome in the contract.⁹

A second general reason for incompleteness has nothing to do with the cost of including a term in a contract, but rather involves the subsequent cost of enforcing a contractual term. Notably, if the cost of providing evidence to the courts that a relevant contingency or condition has occurred is sufficiently large, then the term will not be worthwhile including. A third important reason for incompleteness is that some contingencies (such as whether the seller has a

⁷Similar forms of holdup would arise in the absence of contract enforcement where parties want to convey property that already exists, such as land; for instance, a seller might worry about being held up by the buyer if the seller waits and forgoes a present opportunity to sell his land to a new party who makes a bid for it.

⁸The idea of contract enforcement as a cure for holdup-related inadequacy of investment and effort in the contractual enterprise was initially stressed in the economics literature by Grout 1984, Klein, Crawford, and Alchian 1978, and Williamson 1975. However, the general idea that contract enforcement is privately and socially desirable because it fosters production and trade is made (usually, with little articulation) by most writers on contract law and one supposes that it has always been appreciated. See for example E. Allen Farnsworth 1999, 6-7, and Pound 1959, 133-34.

⁹More precisely, suppose that the cost of including a term for an (anticipated) contingency is c , that the likelihood of the contingency is p , and that the loss the parties would jointly suffer from failing to include a term for the contingency is l . Then the parties will tend to exclude the contingency if the associated expected loss of pl is less than the cost c of inclusion, that is, if p is less than c/l .

stomach ache) or some variables (such as the effort level of the seller, or technical production difficulties) cannot be verified by courts. If a contingency or the value of a variable cannot be verified by courts – if there is an asymmetry of information between the parties and the courts -- then were the parties to include the contingency or variable in the contract, one of the parties would generally find it in his interest to make a claim about the contingency or the variable, causing problems. (For example, if the contract specifies that the seller need not perform if he has a stomach ache, he would claim he had a stomach ache if he later did not want to perform; or if the contract specifies that the buyer does not have to pay for a service if it is performed poorly and the quality of performance cannot be verified by the court, the buyer would always find it in his interest to claim that performance was subpar in order to escape having to pay.) It should be noted that even if the parties can themselves verify contingencies and named variables, contracts that include them will still be unworkable if the courts cannot verify them (even if the buyer knows the seller's effort level, a contract depending on effort level will be unworkable if the court cannot verify it, for both the buyer and the seller can make false claims about it). Of course, many variables that seem unverifiable can be made verifiable (perhaps the quality of service performance can be made verifiable through videotaping), but that would involve expense.

A fourth factor explaining incompleteness of contracts is that the expected consequences of incompleteness may not be very harmful to contracting parties.¹⁰ To amplify, a court might interpret an incomplete contract in a desirable manner, as we are about to discuss. In addition, as we shall see, the prospect of having to pay damages for breach of contract may serve as an implicit substitute for more detailed terms because it may lead parties to act as they would have under more detailed terms. Furthermore, the opportunity to renegotiate a contract often furnishes a way for parties to alter terms in the light of circumstances for which contractual provisions had not been made, and will lead them to do what they would have provided for had they written a more detailed contract in the first place. Finally, incompleteness may not matter at all because it may concern contractually irrelevant events. There are a multitude of such irrelevant events -- for example, whether it rains in China will be irrelevant to the parties to the wedding photography contract -- and parties obviously will not specify terms for irrelevant events because of the positive cost of so doing.

5. Interpretation of Contracts

Given that parties leave contracts incomplete, questions naturally arise about the interpretation of contracts by courts. As a general matter, parties will want incomplete contracts to be interpreted in the way that they would have written had they spent the time and effort to specify more detailed terms.¹¹ For example, suppose that a builder and a buyer do not include a term in their

¹⁰In strict logic, this is not an independent reason for incompleteness but rather one that complements the previous reasons: The lower are the losses from incompleteness, the more likely it is that parties will find the costs of writing or enforcing terms not worthwhile bearing, and the more likely they will not find it worthwhile incurring the costs of rendering a variable verifiable to the court.

¹¹It should be noted that such interpretation can be carried out by courts only when the reason for incompleteness was the effort to anticipate or expense for parties of specifying more terms. When the reason for incompleteness is that the court cannot verify a term (such as the wedding photographer's level of effort), then the courts by assumption cannot attempt to complete the contract (by taking into account the photographer's effort level).

contract stating whether the builder is to perform if material prices rise steeply, but had they included the term, it would have relieved the builder of having to perform in that circumstance. The parties would want the courts to interpret the incomplete contract in that way should prices rise steeply.¹²

The advantage to parties of correct interpretation of their intentions by courts is not only direct -- that terms that they leave incomplete will be filled in as they would have wanted. The advantage of correct interpretation is also indirect -- that the parties can then omit more explicit terms and thereby save drafting and negotiating costs. Indeed, the formal statement of how to evaluate a method of contract interpretation makes it clear that interpretation has both direct and indirect effects. The formal evaluation is as follows: Given a method *M* of contract interpretation, first determine what terms the parties to a contract would rationally choose to include, presuming that the parties know that *M* will be used to interpret their contracts (in other words, take into account the indirect effect of use of *M*);¹³ and second, calculate social welfare having ascertained what terms the parties will include and how the courts will interpret incompleteness.¹⁴

Several comments should be made about the courts' task of interpretation. Consider first

¹²To amplify why the parties would be made better off, suppose that the seller would insist on raising the contract price by \$100 if he were obligated to perform when prices turned out to be very high -- in order to cover his increase in expected costs -- but that the buyer would only place an expected value of \$50 on receiving performance in that circumstance. Then were the two parties to include an explicit term regarding high prices, the term would state that the seller not perform, for the buyer would not want to pay the extra \$100 in the price for a \$50 benefit. If the court interprets the contract in this way, the parties tend to be made better off. If the court, however, misinterprets the parties' intentions and they know this, then the cost of the contract to the buyer would rise by \$100 for only a \$50 benefit, making the contract less valuable for him. As a consequence of this and similar misinterpretation, the buyer's willingness to pay for the contract would fall, harming the seller as well. Conversely, correct interpretation tends to benefit both parties.

¹³A sketch of how this could be done is as follows. Suppose that the parties contemplate including a term *t* in their contract. Each party can calculate his expected utility if *t* is not included, given that method *M* will be used to interpret the contract if a contingency relevant to *t* arises. (To calculate expected utility, some assumption about the determination of contract price must be made, such as that the price is such that one party obtains a fixed percentage of the surplus.) Each party can also calculate his expected utility if *t* is included; this calculation will take into account the cost of including *t*. It would be natural to assume that the contract excludes *t* if both parties are better off by so doing, and that it includes *t* if either party wants that.

¹⁴The social welfare maximizing method of interpretation can be described loosely as that which minimizes the sum of writing costs (the cost of including terms) plus error costs of interpretation (the social welfare goal would be precisely this if social welfare equals the sum of parties' utilities and they are risk neutral). Therefore, a method *M* that tends to be an accurate reflection of parties' desires lowers error costs, and it also leads (as explained in the last footnote) to exclusion of terms and thus lower writing costs; thus, both the direct and indirect advantages of *M* enter into the formal calculus, as stated in the text. It should be observed that the optimal method of interpretation may involve some subtleties. For example, according to the optimal method, a term might not be interpreted in the way that is best in the majority of transactions. Suppose that term *t*₁ is best in the majority of transactions and that the parties to these transactions can include *t*₁ explicitly, at little cost on a per-contract basis, because they are repeat players. Suppose that term *t*₂ is best only in the minority of transactions, but that for the parties to these transactions to include *t*₂ explicitly will not be cheap on a per-contract basis because they are not repeat players. Then the optimal method of interpretation would make *t*₂ the default term in an incomplete contract even though it is best only in a minority of transactions.

the situation in which there is a literal gap in a contract, like the wedding photography contract that mentions sunny days, when videos should be taken of the ceremony, and rainy days, when only still pictures are to be taken of the ceremony, but fails to say what should be done on cloudy days. In such a case, courts know that they must fill the gap -- if a cloudy day arose, the courts must say what should have been done, and the job of the courts is to determine from evidence what the parties would have wanted. Consider next an incomplete contract that does not have gaps, such as a contract that says the wedding photographer shall take videos of the ceremony unless there is rain, in which case stills are to be taken. By implication, this contract covers cloudy days -- video pictures are to be taken of the ceremony on such days -- but there are still possible ambiguities about the contract: Did the parties really mean for videos of the ceremony to be taken on cloudy days, or did they not mean that and leave out explicit mention of cloudy days because of the cost of so doing and, perhaps, knowledge of how the courts would interpret their contract if the weather were cloudy?¹⁵ Incomplete contracts that do not contain literal gaps always involve such ambiguities about the parties' real desires, so that the need for judicial interpretation is not clear, and the possibility that courts would err in determining what the parties want may increase as a consequence.¹⁶

6. Damage Measures for Breach of Contract

6.1 Damage measures defined. As noted in section 1.3, when parties breach a contract, they often have to pay damages. The damage measure, that is, the rule or formula governing what a party in breach should pay, can be applied by the court or it can be stipulated in advance by the parties to the contract (in which case damages are sometimes referred to as *liquidated* damages because they are intended to liquidate, to terminate, the legal obligations of the party in breach).¹⁷ One would expect parties to specify their own damage measure when it would better serve their purposes than the measure the court would employ, and otherwise to allow the court to select the damage measure. In either case, I now examine the functioning and utility of damage measures to contracting parties.

6.2 Damage measures and incentives to perform. It is clear that damage measures provide parties incentives to perform, by threatening them with having to pay damages if they do

¹⁵The ambiguity that I am referring to here arises, in strict logic, when parties do not have a clear understanding of how the courts will interpret incompleteness; thus the situation is more complicated than that described in the previous two notes, in which I assumed the parties know for sure the method *M* of interpretation.

¹⁶On various aspects of contract interpretation, see, for example, Ayres and Gertner 1989, Cohen 2000, Hadfield 1994, Katz 1990c, 1998, Alan Schwartz 1992, and Triantis 2000.

¹⁷There is a possible ambiguity in the meaning of the word "breach." What is meant by breach in ordinary language is that a party does not do what it is said that he will do in the contract. For example, if the contract states that the seller shall deliver goods and the seller does not do that, he would usually be said to be in breach. Suppose, however, that the contract contains a liquidated damages provision stating that if the seller does not deliver goods, he should pay \$100 in damages. Then if he does not deliver the goods and pays the \$100, it might be said that he did not breach the contract because he did what was required in the contract, namely, pay the \$100. However, in such cases where the contract names damage measures, we will still refer to an event in which a person does not do something specified in the contract and instead pays damages as being a breach. In any event, the issue under discussion is really semantic, because however we choose to describe parties' behavior, our analysis of behavior under different contracts, and under different legal rules, will be the same.

not. Suppose that the buyer wants a custom desk built, and that the measure of damages for seller breach is \$800. Then the seller would be induced to build the desk if he would profit from so doing or if his losses would be less than \$800, but the seller would commit breach if his losses from performing would be higher than \$800. Thus, a particular damage measure provides a particular degree of incentive to perform, and in general, it is evident that the higher is the damage measure, the greater the incentive to perform.

Best measure of damages when contracts are completely specified. The question naturally arises, what measure of damages provides the best incentive for the parties to perform? That is, what damage measure would most raise their expected utilities from contracting? It might seem that a high damage measure, even a punitive measure, would be best, for that would give a strong motivation to obey a contract. This idea is correct if a contract is truly completely specified. In that case, a high damage measure -- high enough that no party would ever breach the contract -- would be in the parties' mutual interests because they would then be assured that exactly the contract they want would be obeyed.¹⁸

Let us illustrate with a contract for the building of the desk, and let us assume that the buyer places a value of \$1,000 on having the desk. If such a contract were mutually beneficial and completely specified, then it can be shown to have the following simple character: The seller is to make the desk if the production cost would be less than \$1,000; and the seller is excused from performance if the production cost would exceed \$1,000. (In essence, the explanation is that the buyer would be willing to pay enough to the seller to induce him to accept terms specifying that the desk should be built if production cost is less than \$1,000, but the buyer would not be willing to pay enough to the seller to induce him to include terms calling for performance when the production cost would be higher than \$1,000.) Now note two points about the outcome if the damage measure for breach were high enough to guarantee performance of the terms in this contract: First, the seller would be led to construct the desk when the production cost would be less than \$1,000, but second, the seller would *not* be led to construct the desk when the production cost would exceed \$1,000, for the completely specified contract does not call for that, and thus no damages would be paid by the seller when he fails to construct the desk in such circumstances. Observe, moreover, that this statement is true no matter how high the damages for breach are. By contrast, under a damage measure that is not high enough always to induce performance of the contract, there will be, by hypothesis, some situations in which the construction cost is lower than \$1,000 and the seller will decide to commit breach and pay damages. Thus, the actual outcomes under this contract will be different from what is intended by the parties, and it can be shown that the parties will generally be worse off under this lower measure than under a higher damage measure that always induces performance.

The general points illustrated by this example are that, under a damage measure that is sufficiently high so as necessarily to induce performance of a mutually beneficial completely specified contract, (a) performance is always guaranteed, yet (b) there is no risk of a party's having to perform when that would be onerous, and there is no risk of having to bear high damages for breach. The latter points are true because, whenever performance would be onerous,

¹⁸A slightly different way to express the point that the parties do not want breach is this: The hypothesis that the parties would want breach in some circumstance contradicts the assumption that the completely specified contract is mutually desirable in that circumstance (and could have been altered to allow for nonperformance in that circumstance, but was not).

the contract, being completely specified and mutually beneficial, will not call for performance.

Best measure of damages when contracts are incomplete. When contracts are not completely specified, then damage measures that are high enough always to lead to performance of the incomplete contract, or to lead to too frequent performance of that contract, are often undesirable for the parties. Instead, only moderate damages are desirable, because they will result in breach when performance of the incomplete contract would be difficult.

Before amplifying on this point, let us reconsider the example from above. Suppose that the contract states simply that the seller shall make a desk for the buyer and the buyer shall pay for it at the outset. The contract does not have specific terms because, say, of the cost of taking the time to include them. Given this incomplete contract calling for performance under all circumstances, a high measure of damages would be needed to guarantee performance, to make it certain. For instance, suppose that production costs could range up to \$5,000. Then the damage measure for breach would have to exceed \$5,000 in order to guarantee performance; a lower measure, such as \$3,000, would result in breach whenever production cost would exceed \$3,000.

Now a damage measure that is so high as to result in performance of the incomplete contract all the time would result in outcomes very different from that under the mutually beneficial completely specified contract: Under the complete contract, the desk would be constructed *only* when its production cost is less than the buyer's valuation of \$1,000, whereas under the incomplete contract with a high damage measure, the desk would be built even when the production cost exceeds \$1,000. This suggests what will later be shown to be true, that the parties will be worse off with the high damage measure, due to the excessive performance it brings about. (The kernel of the explanation is that the seller will charge the buyer a higher price because of the costly performance he might be led to undertake, and the buyer by assumption would prefer not to have performance in these costly circumstances, in order to benefit from a lower contract price.)

A moderate damage measure, however, will not lead to the problem of excessive performance, for if damages are less than the high production cost levels, the seller will commit breach when production cost is high. Indeed, if the damage measure equals \$1,000, the value of the desk to the buyer, the seller will be led to perform precisely when he would have performed in the mutually beneficial completely specified contract: For the seller will perform when production cost is less than \$1,000, and he will breach the contract and pay damages when production cost would exceed \$1,000. This damage measure, equal to the value of performance, is the *expectation measure*, the most commonly employed measure of damages, and as will be seen subsequently in chapter 15, it leads under fairly general circumstances to performance when parties would wish that.

Because moderate damage measures allow breach of incomplete contracts when performance would be expensive and induce performance when it would not be expensive, moderate damage measures lead to performance in circumstances resembling those (and in the example, identical to those) under mutually beneficial completely specified contracts. This suggests what will be later shown, that moderate damage measures are preferred by *both* parties to other damage measures.¹⁹

¹⁹In particular, it will be shown in chapter 15 that both parties would often elect to choose the expectation measure over other damage measures when writing the contract .

Moderate damage measures serve as implicit substitutes for more complete contracts.

One implication of the preference of both parties to a contract for moderate damage measures is that such damage measures function as substitutes for detailed contracts. It has been seen that if a contract leaves out terms stating when contracts should be performed and when not, use of a properly chosen moderate damage measure will lead to performance in approximately the circumstances that the parties would have named in a more completely specified contract. That is because performance will be induced when it is not too burdensome to perform, but performance will not be induced when it would be too difficult to perform. Therefore, the opportunity of the parties to employ moderate damage measures enables them to write contracts that lack great detail while still knowing that performance will occur roughly when they want.

The value of damage measures to parties as a substitute for more complete contracts depends on the transaction costs of the use of damage measures versus the costs of specifying contracts more fully in advance. It should also be observed that damage measures can serve implicitly to complete contracts when it would be impossible for parties to write them, due to the inability of courts to verify the occurrence of contingencies. Suppose that the production cost of making the desk in our example is something that is inherently unobservable by courts, because the production cost depends on idiosyncratic factors having to do with specialized carpentry, making it impractical for the builder to convince the court that the production cost would be high. Then a contract specifying that there should be no production when production cost is high would not be workable. But a contract with a moderate damage measure of \$1,000 for breach would be workable and would lead, as has been emphasized, to the result that the parties want, of no performance when production cost is high, and this would be so *without* the court's having to verify that the level of production cost would be high.

Qualification: when contractual duties are financial, damage measures cannot serve as substitutes for more completely specified contracts. If a contract is incomplete and a party's contractual duty is to pay an amount of money, then a damage measure cannot serve to induce performance when it would have occurred in a more complete contract. Consider a building construction contract and the obligation of the buyer to pay the seller \$1,000,000 for the building when it is finished. Suppose that a complete contract would say that the buyer would only have to pay \$700,000 if he suffers a significant financial reverse. If the contract is incomplete and does not have a provision for the financial reverse, the use of a damage measure cannot relieve the buyer's obligation in that circumstance. In particular, the damage measure must be \$1,000,000 to induce the buyer to do what he promises (namely, pay \$1,000,000) in ordinary circumstances; yet if that is so, the buyer has to do the same when he suffers a financial reverse.²⁰

Thus, in general, when contractual obligations are financial, damage measures cannot serve to fill out incomplete contracts; they can only induce performance of the incomplete contract that is written. Hence, for parties to avoid the problems due to incompleteness, they must either rely on courts' interpretation of their contracts (perhaps the courts would lower the buyer's obligation to pay if he suffered a reverse) or they must simply take the trouble to write a

²⁰Note the difference when a party's obligation is to take an action other than pay money. If the action, such as building something, becomes difficult or expensive for the person, he can relieve his burden -- benefit himself -- by paying damages of a set amount. When his obligation is *itself* monetary, then paying damages equal to that amount cannot possibly help him.

more complete contract in the first place. This point, then, applies for the parties' obligations to pay in contracts to provide goods and services, and generally for purely financial contracts (loans, insurance, and the like).

6.3 Is breach and payment of damages immoral? The discussion in section 6.2 sheds light on the often-debated question of whether a breach of contract is an immoral act, similar to the breaking of a promise.²¹

To understand and evaluate this assertion, let us assume in this section that the type of promise that ought to be kept is a completely specified promise that the parties could be imagined to make. This assumption is natural, for, by definition, it is only the completely specified promise that is explicit about the desires of the parties in each of the circumstances of possible relevance to them. It would not be natural to interpret an incompletely specified promise as embodying the desires of the parties in a particular circumstance if the parties would have stipulated something different from what the incomplete promise says for that circumstance, and if the reason that the parties did not provide for the circumstance was that it would have been inconvenient for them to take the time to do so.

Given, then, the assumption that the completely specified contract represents the promise of the parties that ought to be kept, and that incomplete contracts are not what ought to be kept, we can see that the view that it is immoral to breach contracts and pay damages is confused and may well represent the opposite of the truth. Consider the incomplete contract for the making of the desk that is worth \$1,000 to the buyer, that names no contingencies concerning production cost, and for which the expectation measure would determine damages for breach. Under this measure, breach will occur whenever production cost exceeds \$1,000. In such instances, nonperformance is exactly what would have been allowed in the completely specified contract that represents the real wishes of the parties and the promise that they would want met. Thus, the breach induced by the damage measure is seen to satisfy the true promise agreement of the parties, not to abrogate it; in this sense, the truth about breach and damage measures is the opposite of what the view that breach is immoral suggests.²² That view reflects a failure to recognize the possibility -- and the practical reality and reasons for -- incomplete contracts.²³

6.4 Incentives to rely. Another function of damage measures for breach is that, because they encourage contract performance, they provide contracting parties with incentives to take actions relying on performance. These actions can raise the value of contracts for parties, inuring to their benefit. For example, a restaurateur expecting construction of a restaurant by his builder, could hire and train staff and advertise the opening of the restaurant, and thereby enhance the profitability of the contract. Such reliance actions are a byproduct of the confidence in

²¹In the philosophical literature, see, for example, Kant [1785] 1997, 15,32,38, Ross 1930, chapter 2, and Searle 1964; and in the legal literature, see, for example, Fried 1981, chapters 1-2. These views are reviewed in Kaplow and Shavell 2002c, 157-65.

²²This point is initially made in Shavell 1980b.

²³Lest I be misunderstood, what I have said implies that breach would be immoral if the damage measure is not sufficiently high to induce performance in a circumstance in which the completely specified contract would have called for performance. Thus, if the completely specified contract would have stipulated performance when the desk costs \$500 to produce, and for some reason the seller is able to commit breach and pay only \$100 in damages, breach would properly be considered immoral.

performance that the use of damage measures produces. Although one might be inclined to think that the inducement of reliance is necessarily a good thing -- there being too little reliance if there is no contract enforcement -- it turns out that the use of damage measures also may actually lead to excessive reliance in a sense that will be made precise later, in chapter 15.

6.5 Risk-bearing. A third important function of damage measures concerns the allocation of risk. Notably, because the payment of damages compensates to one or another degree the victim of a breach, the measure might be mutually desirable as an implicit form of insurance if the victim is risk averse. For this reason, damage measures may gain additional appeal for the parties on risk-bearing grounds.

However, the prospect of having to pay damages also constitutes a risk for a party who might be led to commit breach (such as a seller whose costs suddenly rise), and he might be risk averse. This consideration may lead parties to want to lower damages, or to avoid use of damages as an incentive device, by writing more detailed contracts. (For instance, the parties could go to the expense of specifying in the contract that a seller can be excused from performance when his costs are high.)

A full consideration of damage measures and efficient risk allocation would also take into account whether the risk that a party bears is detrimental or beneficial. For example, if a party wants to breach, not because he has run into costly production difficulties, but rather because another party has bid more for what he has made, then risk-bearing considerations would not lead to lower damages for the seller. Another relevant consideration is whether a risk is monetary or nonmonetary. If, for instance, the victim's loss is nonmonetary, such as the loss due to the failure of a photographer to appear at a wedding, financial compensation in the form of damages may not constitute an optimal form of insurance.²⁴ An additional consideration is the availability of commercial insurance to the parties for the losses due to breach; if such insurance is available, then the need for damages to compensate the victim is negated, and damages have a role mainly as an incentive device.

Note on the literature. The point that a moderate damage measure, and in particular the expectation measure, is socially desirable because it induces performance if and only if the cost of performance is relatively low was originally stated, informally, in Posner (1972a),²⁵ but he did not observe that the expectation measure is mutually desirable for the parties. In Shavell (1980b) I first stress the mutual desirability of moderate damage measures for the contracting parties themselves and the role of damage measures as implicit substitutes for more complete contracts, and I also first analyze damage measures and incentives to rely.²⁶

7. Specific Performance as the Remedy for Breach of Contract

²⁴Recall the discussion of insurance for nonmonetary losses in section 6, Chapter 5.

²⁵Two other writers, Birmingham 1970 and Barton 1972, adumbrate this point, although the meaning of their articles is at times obscure.

²⁶On remedies for breach, see the surveys Edlin 1998, Shavell 1998, and the references cited in section 2 of chapter 15.

7.1 Specific performance defined. As observed at the outset, an alternative to use of a damage measure for breach of contract is specific performance: requiring a party to satisfy his contractual obligation. The interpretation of specific performance depends on the nature of the contractual obligation. Usually, specific performance refers to obligations to deliver a good or to perform a service, in which case it means that exactly that must be done.²⁷ If the contractual obligation is to pay a given amount, such as for an insurance company to pay coverage to an insured, then the meaning of specific performance is the payment of money. Specific performance can be accomplished with a sufficiently high threat, or by exercise of the state's police powers, such as by a sheriff forcibly removing a person from the land that he promised to convey to another. Note too that if a monetary penalty can be employed to induce performance, then specific performance is equivalent to a damage measure with a high level of damages.

7.2 Incentives to perform and specific performance. What I said above about damage measures bears on the desirability of specific performance.

Specific performance is desirable for completely specified contracts. If contracts are completely specified, then parties want them adhered to, so that specific performance is a desirable remedy for breach, because it means that there will be no breach. In the example concerning the desk, the seller would be required, under specific performance, to make the desk whenever the production cost would be less than \$1,000 -- but only then, as I emphasized. Specific performance would never constitute a burden for the performing party because any difficult contingency (such as production cost exceeding \$1,000) would have been included in the contract and the provision for it would have allowed the party not to perform.

Specific performance is usually undesirable for incomplete contracts. If contracts are incomplete, then, for the reasons given favoring moderate damage measures, specific performance would not be desired by the parties. In the example of the desk, specific performance would lead the maker of the desk to construct it when production cost would be high, exceeding \$1,000, regardless of the magnitude of production cost. This would lower the value of the contract to the two parties. Thus, like very high damage measures, specific performance would not be what the parties would want as a remedy for breach.

Qualifications. Nevertheless, there are circumstances in which specific performance will be desirable for the parties, as will be explained in chapters 15 and 16. One important situation concerns cases where courts would have difficulty in estimating the value of performance, meaning, among other things, that if a damage measure were employed, there would be a danger that breach would occur when performance would be in the mutual interests of the parties; use of specific performance would avoid this danger. Yet specific performance might result in the problem of excessive performance stressed above. That problem, though, will be mitigated in some circumstances, especially because of the possibility of renegotiation of contracts (see below). Thus, and as will be seen later, the details of the argument suggesting that specific performance may be desirable as a means of inducing mutually desirable performance are somewhat subtle (see section 1 of chapter 16).

7.3 Incentives to rely; risk-bearing. It is obvious that specific performance supplies

²⁷However, some economists have employed "specific performance" in an unconventional sense: They would say that a contract to make something is specifically performed even if the seller breaches, provided that the contract names a liquidated damage amount that the seller pays (because then the seller is doing exactly what the contract requires).

parties strong incentives to rely on performance, so that this is generally a positive aspect of specific performance (again, though, see the discussion below for details).

With regard to risk-bearing, specific performance imposes large risks on sellers if, as is often the case, they might face very large costs of performance. When so, specific performance would frequently be mutually undesirable on grounds of risk-bearing; the parties would not choose it as the remedy for breach just on that ground, as the seller would charge more for bearing that risk than the buyer would be willing to pay. However, there are circumstances in which the risk-bearing implications of specific performance are not onerous for the seller. Notably, and as mentioned in section 6.5, suppose that the seller might want to commit breach not because he faces high costs of performance, but rather because he encounters another party who would pay much more for his product or possession or for his service; then specific performance would not impose a cost on him, merely deny him a positive opportunity.

7.4 Ability to enforce. The ability of courts to enforce specific performance depends on the type of contractual obligation. If the obligation is to perform a service or make something, then enforcement means requiring a person to undertake particular actions, and thus may entail special difficulties, especially if the person is recalcitrant. If the obligation is to convey a material thing, such as a painting or land, then specific performance does not involve that difficulty, but does require that the thing to be located (unless it is land) and for it to be taken from the holder and given to the buyer. Another point of note is that enforcement of specific performance does not require, as damage measures do, that the assets of the party in breach be found and that he be forced to pay, as the police powers of the state are used directly to force performance.²⁸

8. Renegotiation of Contracts

8.1 Reasons why renegotiation may or may not occur. Heretofore, the possibility that contracts might be renegotiated when difficulties arise has not been explicitly considered, but the possibility often arises. For example, if construction cost is high relative to the value of performance, and the damage measure would induce seller performance, might not the seller renegotiate with the buyer and pay him to be allowed not to perform? There are appealing reasons to consider such renegotiation to be likely, the main ones being that, having made an initial contract, the parties know of each other's existence, will usually be aware of each other's locations, and will be cognizant of many particulars of the contractual situation that would make renegotiation mutually beneficial.²⁹

However, before discussing the implications of renegotiation, let me briefly note why renegotiation may not occur. One reason is simply that, at the time that difficulties are experienced, one party might benefit from acting quickly but not be in contact with the other, and arranging immediate renegotiation might be costly. A producer might benefit from acting

²⁸For economic analysis of specific performance and its comparison to damage remedies, see Bishop 1985, Kronman 1978b, A. Schwartz 1979, Shavell 1984a, and Ulen 1984. (Specific performance is also examined in many of the articles cited in chapter 15.)

²⁹For this reason, much of the economics literature on contracts assumes that renegotiation always occurs when outcomes that are not mutually beneficial would otherwise result; see, for example, Hart 1987, Hart and Holmstrom 1987, and Rogerson 1984.

quickly because, for instance, a problem may occur during the course of production and the producer may have to decide on the spot whether to abort the process or proceed at greater cost. Or a new bid may be heard and have to be immediately answered.

A second reason why renegotiation may not transpire, or more exactly, may not succeed, is that even if the parties are in contact with one another, asymmetric information between them may lead to breakdowns in bargaining.

Another reason why renegotiation may not occur is that it may be impossible to alter the outcome: Rather than a breach being the result of a party's decision that can be modified -- such as whether the seller conveys a painting in his possession to the buyer -- breach may be the result of a *past* decision of the party that cannot be undone -- such as whether the seller took precautions to prevent one of his employees from selling the painting to another person.³⁰ If the breach event cannot be undone, then the issue of renegotiation of the contract is obviously rendered moot.

Despite these reasons why renegotiation may not result in a new contract, let us presume in the remainder of this section that when difficulties arise and a mutually beneficial renegotiated contract exists in principle, it will be made.

8.2 Renegotiation and contractual performance. If contracts will be renegotiated when difficulties arise, then performance of contracts will occur whenever that would be mutually beneficial, despite the incompleteness of contracts.

Let me illustrate with the example of the production contract for the desk worth \$1,000, and recall the statement that in a completely specified contract, the parties would have specified performance when production cost is less than \$1,000 but not when it is more. Suppose, though, that the contract does not mention any contingencies and, initially, assume that the remedy for breach is specific performance. Then, in the absence of renegotiation, the seller would be led to make the desk when production cost exceeds \$1,000 as well as when production cost is less than \$1,000. However, the contract would be renegotiated whenever the production cost exceeds \$1,000. For instance, if the cost would be \$1,500, the seller could pay the buyer \$1,250 for an agreement to allow him not to perform; this would be mutually beneficial because \$1,250 exceeds the \$1,000 value of performance to the buyer, and \$1,250 is less than the production cost of performance for the seller.³¹ A similar argument shows that when a contract with damages for breach would result in outcomes that differ from those in the completely specified contract, renegotiation would occur. Specifically, if the seller would be induced by the threat of high damages to perform when the production cost exceeds the value of performance, an agreement will be made in which the seller pays the buyer an amount less than damages and does not perform. And if the seller would be led to commit breach, because damages are low, when production cost is less than the value of performance, then an agreement will be reached in which the buyer pays the seller to induce him to perform.³²

³⁰Suppose the seller is a dealer and he failed to issue clear instructions to his personnel not to sell the painting. Thus, the breach may be said to occur as a probabilistic result of an action (issuing of inadequate instructions by the dealer). Such situations in which breach is a probabilistic result of actions are an important category of case.

³¹Any amount between \$1,000 and \$1,500 would be mutually agreeable.

³²Examples are spelled out in section 4 of chapter 15.

In general, whatever the degree of incompleteness of the contract, and whatever the remedy for breach, renegotiation will lead to performance exactly when that would have been stipulated in a mutually desirable completely specified contract. Therefore, renegotiation reduces the need for complete contracts and serves as an implicit substitute for them.³³ (In this sense, renegotiation serves a purpose that is similar to damage measures.)

Qualification for financial contracts. The above argument that renegotiation may reduce the need for complete contracts is premised on the assumption that when a mutually desirable provision for a contingency is left out of a contract and the contingency at issue arises, there will be, at that time, a mutually desirable alteration in the contract. Thus, when the production cost exceeds \$1,000 and the contract calls for production, there is a mutually desirable change in which the seller does not produce. However, if purely financial contracts are incomplete, there generally is *not* a mutually desirable change that exists when unprovided-for contingencies arise. For example, suppose that a contract between two companies stipulates that costs in a joint venture should always be shared equally, but the contract is incomplete in that it does not state that less than half of the costs should be paid by a company if finds itself in very bad financial straits, a term that might be mutually desirable. If an event of financial stress for a company occurs, there is no mutually beneficial alteration in the contract terms that can then be arranged; for any reduction in the amount that one company pays hurts the other company. Thus, renegotiation cannot ameliorate the problem of contractual incompleteness for financial contracts.

8.3 Renegotiation and reliance. The prospect of renegotiation affects the incentives of parties to invest in the contractual relationship, and quite possibly it will result in inadequate investment because of the ability of one party to hold up another in renegotiating the contract. Suppose that a buyer who wants a building completed for his business invests in training workers and advertising and then finds that he must renegotiate with the seller for performance. At that point, the price that the seller would obtain would reflect the value that the buyer would receive from performance, having made his reliance investment. The knowledge that his profits from reliance investments will be partially extracted from him through renegotiation might lead the buyer not to rely to a desirable extent in the first place, lowering the value of the contractual enterprise for the parties.

This general point, however, needs to be qualified because renegotiation is influenced by, among other elements, the damage measure that applies for breach. As will be seen in chapter 15, the damage measure, together with renegotiation, sometimes may ameliorate or correct the problem of inadequate reliance, and also could lead to excessive reliance.

8.4 Renegotiation and risk-bearing. Renegotiation of contracts has implications for risk-bearing. First, renegotiation tends to reduce risk for parties who make payments, for they would generally be worse off if they did not make these payments; if a seller who would otherwise face very high production costs pays to escape his production obligation, his risk is thereby reduced. However, the uncertainty in the amounts that will be paid when contracts are

³³Indeed, if the only issue of importance between the parties was when performance would occur, one might think that there would be no need for any specifications in contracts if renegotiation were costless and perfect. It should be noted, in reflecting on this point, that any losses faced by a party due to renegotiation can be adjusted for in the contract price, so that renegotiation does not lead to any disadvantage or advantage, all things considered, in the renegotiated contract, at least for risk-neutral parties (I will discuss risk aversion below).

renegotiated, because these amounts are not set in advance, implies that risk remains when contracts are renegotiated. To reduce the risk associated with contracts that will have to be renegotiated, parties can employ damage measures that lead to mutually beneficial behavior and/or more fully specify contractual terms.

8.5 Costliness of renegotiation. Another point about renegotiation obviously is its cost, and several remarks about that are worth making. First, if one views writing an explicit term about a difficulty as an alternative to renegotiation over it, one is comparing the ex ante, sure cost of negotiating about the term to the ex post cost of renegotiating about it. Thus, if the negotiation and renegotiation costs are equal, the fact that the renegotiation cost is incurred only with a probability would make renegotiation preferable on grounds of expected cost (although the cost of renegotiation might not be equal -- it could, for instance, be higher if the parties have to locate one another). Second, if one views a properly chosen damage measure as an alternative to renegotiation as a means of ensuring desirable performance, then the cost comparison between damage measures and renegotiation is one that arises only when a problematic contingency in question arises. It might be thought that damage measures involve less cost for parties than renegotiation, for if the damage measure is properly selected, it will automatically result in correct performance (either inducing performance, or leading to breach and payment of damages, usually after a settlement rather than litigation). However, the damage measure that leads to proper performance (the expectation measure, as we mentioned above, and as will be discussed in chapter 15) may not be employed for one reason or another, in which case, renegotiation will sometimes occur under the damage measure.

8.6 Desirability of enforcement of renegotiated contracts. One suspects that in most situations it is desirable for renegotiated contracts to be enforced. That is, prospectively, parties who make a contract will be made better off if they know that, should they renegotiate their contract, the modified contract will then be enforced. This is the case with our example of the contract for production of a desk worth \$1,000 to the buyer and calling for specific performance; for if the parties know that the contract will be renegotiated when production cost exceeds \$1,000 and that the renegotiated contract will be enforced, then in effect performance will be exactly what the parties would want and arrange for in a completely specified contract. Yet they save the time and trouble of writing such a contract and can rely on renegotiation to cure problems if they arise. It is for such reasons that the enforcement of renegotiated arrangements is usually a good thing ex ante for contracting parties, and the law tends to enforce renegotiated contracts.³⁴

However, for somewhat subtle reasons, it is not always true that the enforcement of renegotiated contracts will help the parties prospectively.³⁵ This suggests that a desirable policy

³⁴Indeed, it is easily shown that enforcing renegotiated contracts must raise the expected utility of each contracting party to the original contract if the only effect of renegotiation in a contingency is to raise the expected utility of each party in *that* contingency: Formally, if $U(2)$ is the expected utility of a party under the initial contract given contingency 2, and allowing renegotiated contracts in a particular contingency 2 only has the effect of raising $U(2)$ but does not affect $U(2)$ for other 2, then allowing renegotiated contracts must raise $\int U(2)f(2)d2$, the expected utility of the party.

³⁵A case in point is provided by Fudenberg and Tirole 1990. They investigate a contract between a risk-averse agent and a risk-neutral principal, where the agent's effort is unobservable (so cannot be included in the contract) and output is a probabilistic function of agent effort. For example, the crop yield may be a function of the

for courts to adopt is to enforce renegotiated contracts unless parties originally stated that the contract was not renegotiable. This policy would lead parties to specify as irrevocable contracts that they do not want renegotiated, but courts do not ordinarily allow that to be done.³⁶

Note on the literature. The subject of renegotiation of contracts was analyzed initially by Rogerson (1984) and Shavell (1984a); and then, beginning with Hart (1987), Hart and Moore (1988), and Hart and Holmstrom (1987), it has been investigated in a more general setting that has led to substantial theoretical development.³⁷

9. Legal Overriding of Contracts

9.1 Harmful externalities. A basic rationale for legislative or judicial overriding of contracts is the existence of harmful externalities. Contracts that are likely to harm third parties are often not enforced, for example, agreements to commit crimes, price-fixing compacts, liability insurance policies against fines, and sales contracts for certain goods (such as for machine guns).

Of course, the harm to third parties must tend to exceed the benefits of a contract to the parties themselves for it to be socially desirable not to enforce a contract. Thus, a contract between musical performers and a person who wants to have a party might cause some disturbance to neighbors who would prefer to enjoy a quiet evening, but if the disturbance is not great, the contract would on net be socially beneficial.³⁸

9.2 Losses in welfare to the contracting parties. Another general rationale for nonenforcement of contracts is to prevent a loss in welfare to one or both of the parties to contracts themselves (as opposed to averting harm to third parties). This concern may motivate nonenforcement when a party lacks relevant information, such as when a person buys food that is mislabeled or a security that is not correctly described and as a result is made worse off by the transaction. Similarly, an incompetent person or a child might agree to a contract that makes him or her worse off, and transactions by such individuals are generally not enforceable.

The rationale for nonenforcement that parties would sometimes be made worse off by enforcement also may apply in the context of contract interpretation. As discussed in section 5, interpretation may amount to overriding terms of contracts (such as a term that simply says the

effort of a farmer (the agent) in planting and fertilizing and of weather (which is probabilistic). The original contract might specify that the farmer's pay depend substantially on the yield, in order to give him an incentive to devote effort to the crop. If the contract can be renegotiated, however, the two parties will have an incentive to do that *after* the farmer has sown the crops, for at that juncture he can be protected against risk and his effort is determined. But anticipation of this renegotiation will undermine the farmer's initial incentives to devote effort to the crop. Hence, preventing renegotiation will be beneficial to these parties. Note that in this case, the crucial assumption of the previous footnote does not hold: Renegotiation affects the well-being of parties in *many* contingencies (by affecting effort), not just in a particular contingency.

³⁶For an economic and legal analysis of the possible undesirability of enforcement of renegotiated agreements, see Jolls 1997 and the literature cited therein.

³⁷See in particular the references and discussion in section 4 of chapter 15.

³⁸Moreover, virtually any contract may cause some external harm (denying other potential contracting parties the opportunity to contract with the parties to the contract in question), yet because the harm is generally less than the benefits to the parties, the contracts are desirable to enforce.

contract should be performed), and still increase the welfare of both parties to the contract by making the contract more like a mutually beneficial completely specified contract.³⁹

9.3 Inalienability; paternalism. Two other rationales that are offered for not enforcing contracts may be noted. One is that contracts sometimes are not enforced because they involve the sale of things that are said to be inalienable, such as human organs, babies, and voting rights. However, it seems that wherever the justification of inalienability is adduced, various of the previous two rationales, externality and losses in welfare to the parties themselves, exist (perhaps in subtle form), so that one might doubt whether inalienability stands for a distinct rationale for failing to enforce a contract.⁴⁰ For example, the sale of human organs might be thought undesirable because some individuals will sell their own organs (such as kidneys) without realizing the detrimental consequences to themselves (that is, the contracting parties will be made worse off due to a problem of lack of information); because some individuals will not receive the care they otherwise would and will die earlier than necessary in order for their organs to be harvested (the contracting parties will be made worse off on account of the contract-induced behavior of others); and because the very existence of the market will be understood by individuals as eroding norms of respect for human life (a species of harmful psychological externality), where these norms are themselves welfare-enhancing because they reduce violence and encourage beneficial behavior. When the inalienability justification is used, in other words, it is really a stand-in for a set of such factors as these, either factors that lower the well-being of parties to the contracts or harmful externalities.

Likewise, contracts are sometimes not enforced because of paternalism, such as when a person is not allowed to purchase certain drugs or a child is not allowed to buy pornographic material. This rationale, like that of inalienability, seems to be reducible to the two previous rationales concerning externalities and harm to the contracting parties themselves. If a person is not allowed to purchase drugs, the justification may lie in the possibility that he or she does not understand the true properties of the drugs, or that using them (suppose they are addictive) may result in problems for third parties.

9.4 State's ability to prevent undesirable contracts. The state often will not experience any difficulty in preventing the making of undesirable contracts, for all it need do is refuse to enforce them. Consider, for example, a socially undesirable contract between an insurer and an insured that covers fines of some type. For the insured to collect against an insurer who refuses to pay, he must bring a suit, and if the courts will not enforce the contract, then the insured will not be able to collect. Hence, the insured will never make the contract in the first place. Thus, as long as the courts are needed to enforce contracts, the contracts will not be made, and the state does not need to police the actual making of contracts and root out the undesirable ones.

10. Extra-Legal Means of Contract Enforcement

Although I have assumed generally above, as stated at the outset, that state-authorized courts enforce contracts, other means of contract enforcement should be mentioned.

³⁹Additionally, at least in theory, nonenforcement of contracts might also be beneficial to parties where they would be led to include terms constituting wasteful signals of unobservable characteristics. See Aghion and Hermalin 1990.

⁴⁰See generally Rose-Ackerman 1985 and Trebilcock 1993.

10.1 Private adjudication.⁴¹ Another avenue for enforcement of contracts is use of private adjudicators, such as are provided by arbitration organizations and some trade organizations. Private adjudication can be envisioned to benefit from as much information about the contractual situation as that in state-authorized courts. Indeed, private adjudication can be superior for the parties in that they can select adjudicators who have specific knowledge of the contractual context (often, industry-specific knowledge), and they can also choose the procedures to be followed (notably, they can simplify adjudication and save expense). Such opportunities are not open to those who go to the courts. For these reasons, it seems to be socially desirable for the courts generally to enforce the findings of private adjudicators, and this is in fact usually done.⁴² (However, if private adjudication is employed where contracts would have harmful external effects, it would obviously not be desirable for courts to enforce the private adjudicative findings.)

10.2 Reputation. It is a commonplace that the fear of harm to reputation can serve to induce parties to adhere to contracts. In principle, this reputational factor could lead to the same enforcement of contracts that we imagine to occur in courts. However, that is unlikely for two reasons.

First, courts resolve disputes taking into account much information about contractual situations, and probably more information than would tend to be reflected in parties' reputations. For example, if a person commits a breach, the court awarded damages would often take into account all manner of factors relating to the victim's loss, whereas the effect of the breach on the person's reputation would be unlikely to be calibrated so well to the loss. Or, if a party wants his contract interpreted by the court in accordance with what would be provided in a completely specified contract, say such as to excuse him due to problems he is facing, the courts might do that knowing his true situation, but would his excuse be recognized and would he escape a reputational penalty?

Second, information aside, the reputational incentives of parties to adhere to their contracts may not be sufficient to induce that. Imagine a party who is not going to be transacting very often and whose transactions usually will be relatively modest in scope, but who is presently a party to a very large contract that it would benefit him greatly to breach. He may well rationally do so, paying little or nothing in damages, despite any loss to his reputation.

Thus, although reputation can help to enforce contracts, it will generally be an imperfect substitute for courts, both because it is usually a crude substitute and a less effective one. Nevertheless, reputation is often a cheaper means of enforcement of contracts. Indeed, in many contexts, litigation costs are high enough to make court proceedings not worthwhile, so that it is only reputation that can function to enforce contracts.⁴³

⁴¹The subject of private adjudication is discussed more generally in section 1 of chapter 19.

⁴²See Goldberg, Sander, and Rogers 1999, 235-36, 244-48.

⁴³See, for example, Bernstein 1992, 1998, Charny 1990, Greif 1998, and Klein and Leffler 1981.

Chapter 14

CONTRACT FORMATION

In this chapter I consider various issues concerning the formation of contracts.

1. Search Effort

An important aspect of contract formation is the effort individuals devote to it -- the time and resources they expend searching for and investigating contractual opportunities -- and I will begin our investigation of contract formation with a comparison of the socially ideal and the privately-desired degree of search effort.

With regard to the former, it is socially optimal for a person to search as long as the social benefit of search exceeds the cost to him. The social benefit of search over some time period by the person is the expected value of any contract made by that person and his contracting partner *minus* the expected value of any contract that was prevented because the person's contracting partner became unavailable for contracting with someone else (if A who is searching contracts with B, then B may be prevented from contracting with C).

The private incentive to search for a contract may diverge from the socially optimal incentive for two different, and countervailing, reasons. On one hand, a person who searches will be led to do so only by the return that he himself makes from concluding contracts; he will not count as a negative any benefit that society forgoes because he keeps certain others from making contracts. This suggests that too much effort is devoted by individuals to search for contracting partners. (The present argument, note, is very much like that suggesting that too much effort is devoted to fishing; here finding a contracting partner is similar to catching a fish, which denies another fisherman the opportunity to catch the fish.)

On the other hand (and putting to the side the possibility that when one contract is made another contract may be prevented), when a person makes a contract with someone, he usually obtains only a part of the surplus thereby created -- it is divided between the two parties through the contract price. This means that the private return from making a contract is less than the social return. For instance, a contract that creates \$100 of surplus would create only \$50 of profit for each party if they split the surplus. As a consequence of this division of contractual surplus, the privately-motivated degree of search could fall short of the socially optimal.

The conclusion, therefore, is that there may be, on net, either too much or too little search for contracts, depending on circumstances. If the first factor, that contract formation by one party tends to deny others from making contracts, is dominant, then search effort will tend to be socially excessive; if the second factor, that the return from making a contract is less than the surplus that that contract itself creates, is dominant, then search activity will tend to be socially inadequate.⁴⁴

⁴⁴To illustrate, consider a simple situation. There is a person deciding on the amount of effort e to devote to search for a contract partner. There is one suitable partner, who the person will discover with probability $p(e)$, where p is increasing in e . If the person discovers the partner, a contract will be made, resulting in surplus s , and the person

The nature of the comparison between the private and the socially desirable degree of search effort raises questions about whether social authorities could obtain the information needed to formulate corrective policy. In any case, overt policies to alter search effort do not seem to exist.

2. Fundamental Rule of Recognition of Contracts: Mutual Assent

The basic rule of legal recognition of contracts is that a contract is deemed valid if and only if both parties give a clear indication of assent, such as signing their names on a document.⁴⁵ This rule involving mutual assent has the two fundamental private and social virtues mentioned in section 2 of chapter 13.

First, the fact that mutual assent is *sufficient* for a contract to be recognized, and that the two parties naturally have the power each to give assent, allows them to make enforceable contracts when they so desire. Further, their knowledge that they have in fact consummated a legally recognized contract allows parties to stop search and immediately to engage in actions that will raise the value of the contract, such as hiring workers and buying materials for production.

Second, the fact that mutual assent is *required* for a contract to be recognized means that no party will become obligated unless he wishes that. This in turn is desirable for parties, and socially desirable, because it fosters search and negotiation for contracts. If a party were somehow to become legally obligated against his will, as a result of search and/or negotiation alone, then these activities would be curtailed, and the degree and quality of contracting would suffer.

It may be noted, however, that despite the usual requirement that mutual assent is necessary for contract formation, certain legal doctrines sometimes result in parties becoming contractually bound without having given their assent. This may occur when a party is led to make significant efforts or investments in anticipation of contract formation.⁴⁶ This legal policy not only may result in a chilling of search effort, it may also induce wasteful early investment as a strategy to achieve contract formation.⁴⁷

will obtain a fraction $\frac{1}{2}$ of this, so his net payoff from search will be $p(e) \frac{s}{2} - e$. If the person does not find the partner, the partner will make a contract with another party, and that will produce surplus t . The socially optimal amount of search is determined by maximizing the social payoff, $p(e)(\frac{s}{2} + t) - e$, since the social payoff to the person's finding the partner is $\frac{s}{2} + t$. Clearly, the person will search too much if $\frac{s}{2} > s - t$, and will search too little if $\frac{s}{2} < s - t$. Both of these possibilities arise in Diamond and Maskin 1979, who examine a specific model of search and contracting. Their model takes into account a complexity not mentioned in the text: that the contract price and net payoff from search will reflect to some degree the possibility of a party concluding a contract with other partners. For instance, if a seller knows that he might sell to a buyer other than the one with whom he decides to contract, he will tend to charge a higher price to that buyer. This effect lowers the buyer's anticipated return from search, and thus implicitly makes the buyer take into partial account the fact that when he makes a contract, he denies others the opportunity to do so.

⁴⁵See, for example, Calamari and Perillo 1998, 25, and E. Farnsworth 1999, 10-11.

⁴⁶See Bebchuk and Ben-Shahar 2001, 424, on the United States, and see Wils 1993, 122-30, on civil law countries.

⁴⁷It is true that early investment is sometimes efficient, and a number of articles find that some form of liability may be desirable to induce precontractual investment; see Bebchuk and Ben-Shahar 2001, Craswell 1996, and Wils 1993.

3. Offer and Acceptance

Mutual assent sometimes is not simultaneous; one party will make an offer and time will pass before the other agrees. This is of interest for several reasons, and I will discuss two prominent ones.

The first is that delay may occur between offer and acceptance because the offeree (the party to whom the offer is made) wants to investigate the offer. The information an offeree seeks might concern, for instance, the quality of the neighborhood in which a house offered for sale is located. The offeree's desire for information raises the issue of how long, and the circumstances under which, the offeror will want to be held to his offer, and whether he should be held to it. Suppose that an offeror is held to his terms, say that he offers to sell a house at the price of \$100,000. Then an offeree will often be led to invest effort in investigating contractual opportunities, for he will know he can purchase the house at the named price. If, however, the price can be changed, the offeree might fear that if he comes back and expresses serious interest after investigation, the offeror will then raise the price, say to \$125,000. The anticipation of such offeror advantage-taking would reduce offerees' incentives to engage in investigation and thus might undesirably diminish contract formation. Hence, it may often be both in offerors' and society's interests for offered terms to be enforced for some period of time. Yet offerors' circumstances may change, making it privately and socially advantageous for them to alter contract terms and thus, among other things, to hold open an offer only for a limited amount of time. In the light of these observations, it is not surprising that the law generally enforces the terms of offers that parties make, including the time during which an offer is intended to be effective.

A second reason for a delay between offer and acceptance is simply that the two parties are not physically proximate and therefore that it takes time for the offeror's message to be sent and for the offeree's response to be received. In this regard, two alternative legal rules generally govern contract formation: the so-called *mailbox rule*, under which an acceptance is legally recognized at the time that the offeree sends a message of acceptance (such as by putting it in the mailbox), and the *receipt rule*, under which an acceptance is recognized only at the time that the acceptance is received by the offeror, provided that he has not made another contract in the interim. Under the mailbox rule, early reliance by the offeree is promoted relative to reliance under the receipt rule, for the offeree is assured that there is a contract from the time that he sends an acceptance message. Further, the offeror is not encouraged to make alternative contracts until he hears from the offeree or sends him a message revoking his offer. Which rule will be the superior depends substantially on whether it is more valuable to promote early reliance by the offeree or freedom to make alternative contracts for the offeror. (For example, if the offeror is unlikely to be in contact with other potential contracting partners, but the offeree is likely to hear of many opportunities that he might have to act upon quickly, the mailbox rule would seem superior.) In any event, the adoption of a definite rule will prevent the taking of wasteful actions in a mistaken belief by either of the parties that a contract has or has not been

However, a party who wants to make such early investment could attempt to advance the time of contract formation or could make a preliminary contract that compensates him for his investment whether or not a final contract is made.

made.⁴⁸

4. Fraud

A contract that is regarded as fraudulent generally will not be legally recognized even though it meets the usual requirements for validity.⁴⁹ By fraud is meant actions taken to deceive a party to the contract about information relevant to its value, notably, about the quality and character of the contracted-for good or service or about the price. For example, the seller of a car may hide rust spots, the seller of a restaurant may doctor its sales records, or the buyer of a parcel of land may pay with securities that are really worthless.

Fraud is socially undesirable for several obvious reasons.⁵⁰ First, efforts taken to carry it out are economically sterile. If the seller of a restaurant falsifies the records to show that it had more business than it actually did have, the resources devoted to this task are a waste, as they do not produce anything of direct value to anyone. Second, efforts made to detect fraud, such as by suspicious buyers of restaurants, also constitute a waste, and such efforts are made in a world with fraud, so are an indirect social cost of fraud. Third, to the extent that fraud is successful, it may result in inefficient actions (hiring more staff for the restaurant that will not in fact have many customers) and in poor matches between contracting partners.

By refusing to enforce contracts that were formed using fraudulent methods, the law discourages fraud. Thus, this aspect of the law of contract formation is socially desirable.

5. Mistake

A contract may involve a mistake in the sense that one side knows that the other side does not understand some relevant point about the contract. For example, a seller of a building that is not zoned for business might learn that the buyer believes it to be so zoned, or the buyer of furniture might realize that its price has been mismarked to a tenth of its normal level. (It is assumed here that the lack of information of the mistaken party comes about through his lack of precautions, not through the deceptive effort of the other party. That is what distinguishes fraud from mistake.)

In such situations, allowing the contracts that are made to be enforced has two social disadvantages compared to negating the contracts. First, the making of such mistaken contracts may result in socially inefficient use of resources, like the purchase of a building by someone who wants to use it for business purposes but who cannot really use it for that. (Note too that if the person purchases the building, learns that it cannot be used by him and resells it, transaction costs will have been needlessly incurred.) Second, the fear that one might lose by failing to notice one's mistake may lead to the taking of excessive precautions to reduce the likelihood of mistake. Rather than so inducing parties to reduce the likelihood of mistake, it may be preferable for society to harness the information that turns out to exist about the occurrence of mistake, by

⁴⁸For a summary of the law of offer and acceptance, see, for example, Calamari and Perillo 1998, 25-117, and E. Farnsworth 1999, 109-222. For economic analysis of offer and acceptance, emphasizing aspects of reliance, see Craswell 1996 and Katz 1990b, 1996; see also Katz 1993.

⁴⁹See, for example, Calamari and Perillo 1998, 325-26, and E. Farnsworth 1999, 260-64.

⁵⁰The social welfare criterion that is implicitly under consideration here is the value of things to those who use them minus costs involved in transactions.

doing what the law does, refusing to enforce contracts that are mistakenly made.⁵¹

6. Information Disclosure

When contracts are formed, parties often possess private information that is relevant to the contract. For example, the seller of a house may know that the basement leaks when there is a heavy rain, the seller of a commodity may have information about the future course of commodities, and the buyer of a parcel of land may have information about its oil bearing potential. The existence of such private information raises the question of whether parties should be obligated to disclose what they know when they make contracts.⁵² In fact the law sometimes requires disclosure and sometimes does not.⁵³

6.1 Effects of disclosure obligations. Let us first discuss the effects of disclosure requirements. The direct effects of disclosure obligations are the transmission of information that otherwise would not have been forthcoming, and accompanying changes in prices. In the absence of a disclosure requirement, sellers will tend not to disclose unfavorable information -- such as that the basement of a home leaks -- because this information would tend to lower the price; thus a disclosure requirement results in revelation of unfavorable information by sellers and lower prices, such as for homes with leaky basements.⁵⁴ Similarly, buyers will not disclose favorable information in the absence of disclosure requirements, for that would tend to raise the price that they would have to pay. For example, the buyer of a commodity that he knows is likely to go up in price would have to pay a higher price if he revealed that information, and a disclosure obligation would cause that to occur.

There is also an indirect effect of disclosure obligations: Such obligations dull incentives to acquire information before the making of contracts. If the buyer of the commodity has to disclose his information, he will have less incentive to invest in determining the likely future commodity price because he will have to pay a higher price if his information is positive; if a seller has to divulge the result of an appraisal, the value to him of an appraisal will fall, for he will have to accept a lower than usual price if the appraisal is below expectations.

It should be observed that this second effect is moot when information is naturally in the

⁵¹This section discusses what is known as unilateral mistake, because only one party has made a mistake; on the relevant law, see, for example, Calamari and Perillo 1998, 354-56, and E. Farnsworth 1999, 631-37. Not discussed is the subject of mutual mistake, such as where both parties believe a cow that is sold to be barren but it is in fact not. This subject raises different issues, including incentives to obtain information, and is beyond our scope. On economic analysis of mistake, see Rasmusen and Ayres 1993, and Smith and Smith 1990.

⁵²It may be noticed that in cases of fraud and (unilateral) mistake, just discussed, the situation is that one side possesses private information, so the present case would seem to have a relationship to the previous cases. I will comment on that in section 6.3.

⁵³See, for example, Calamari and Perillo 1998, 325-47.

⁵⁴A complexity in determining the effect of a disclosure requirement is that, in the absence of a requirement, silence about something (such as whether there is a leaky basement) might lead to a rational negative inference (that there is a leaky basement). However, rational inference from silence will generally leave parties with some uncertainty about the truth for a variety of reasons (for example, silence about whether there is a termite problem might mean only that the homeowner never tested for termites, not that he has something to hide). On inferences from silence, see, originally, Grossman 1981 and Milgrom 1981, and see also Gertner 1998, Fishman and Hagerty 1990, Okuno-Fujiwara, Postlewaite, and Suzumara 1990, and Shavell 1994.

possession of a party -- for example, we would expect a homeowner to know automatically whether his basement leaks; he will know this by virtue of living in the home. Likewise, if information comes to a party adventitiously, for free, there cannot be an effect of a disclosure obligation on the party's possession of the information.

6.2 Social desirability of disclosure obligations. The desirability, or lack thereof, of disclosure obligations depends in a somewhat complex way on a number of factors, and three important factors are now sketched.⁵⁵

Whether the buyer or the seller possesses information. The appeal of requiring disclosure is stronger when sellers possess information than when buyers do, for the simple reason that it is buyers who typically can make socially valuable use of information. Thus, it is desirable that the seller of a house tell the buyer about the leaky basement so that the buyer will not store valuables there or can fix the problem. By contrast, it is not necessary for the buyer of land who has information about its mineral bearing potential to divulge this to the seller for the information to be put to use, for the buyer of the land will be the party who extracts the minerals.⁵⁶

Whether incentives to acquire information would be undesirably reduced. Because the incentives to acquire information are diluted by the obligation to disclose, disclosure obligations may sometimes be socially undesirable. If incentives of parties to appraise the mineral bearing potential of land were negated by the obligation to disclose findings, such obligations would be undesirable. Indeed, this point was emphasized in relation to a case in which a company undertook an aerial survey to determine the mineral bearing potential of land and bought mineral rights to the land without divulging that information to the seller. It was suggested that the company would not have spent money on the aerial survey -- and the minerals that were found would not have been discovered -- had the company anticipated that it would have to report on its findings to sellers.⁵⁷

However, several important qualifying points should be made about this argument against disclosure obligations. One is that the incentives to invest in acquisition of information may be socially excessive if there is no obligation to disclose,⁵⁸ and certainly that will be the case if the information is not socially valuable in the first place (see below). A second factor is that the dulling of incentives tends to be less serious for sellers than for buyers, because sellers can obtain in bargaining some of the enhanced value from information. A seller who learns that his land has valuable minerals can charge for the extra profit that can be made from the minerals; whereas if a potential buyer has to reveal this information, he cannot benefit from it assuming that, once the buyer has divulged it, the seller could refuse a deal and find another buyer who could extract the minerals.

⁵⁵The social welfare criterion employed here is the value of things to people, minus any costs involved in transactions, minus costs of acquiring information.

⁵⁶This presumes that an agreement is made. If the buyer fails to conclude an agreement, and never divulges what he knows, then another buyer may be ignorant of the minerals on the land, whereas he would be informed if the seller were told, for the seller would then inform any other buyer in order to obtain a higher price for the land.

⁵⁷See Kronman 1978a, 20-21, and his discussion of a case involving the Texas Gulf Sulfur company.

⁵⁸It can be shown in a natural model of information acquisition that, in the absence of the requirement to divulge information, private incentives to obtain information will be excessive. However, requiring disgorgement of information may, as stressed in the text, result in insufficient incentives to invest in information.

Of course, there may be no effect on incentives to acquire information, as noted above. In the case of the leaky basement that would be true, so that a disclosure requirement about that information for the seller of the home would not detrimentally affect the acquisition of that information; he would have it regardless.

Whether information is socially valuable or merely has private value. Some information has social value in the sense that it can be used by someone to enhance the value of something. For instance, knowledge that the basement of a house is leaky has social value because, as stated, the person who comes to live in the house can take precautions not to store things in the basement or can have it repaired. Likewise, information about the mineral bearing potential of land can be used to extract the minerals. Some information, though, has low or, in principle, might have no social value, even though it has definite private value. For example, advance information that the commodity price is going to rise due to a fungus that will reduce the supply of the commodity might have little social value if nothing can be done with that information, yet the information might have significant private value because of the profits from price changes that can be made from it. Where information has low social value, its costly acquisition should be discouraged, suggesting the desirability of requiring its disclosure (and note, paradoxically, that this is so even though there may be no intrinsic value in disclosure itself).⁵⁹

6.3 Comment on the relationship to fraud and mistake. Situations of fraud and mistake are similar to those considered here in that they also involve asymmetry of information. But here, unlike with fraud and mistake, it is sometimes desirable to allow a party to the contract not to disclose information. The main reason for the difference in conclusion is that in the context of this section, inducing acquisition of information may be socially desirable. However, in the context of fraud and mistake, allowing nondisclosure (that is, allowing fraud and, allowing mistakes to go uncorrected) can only lead to socially undesirable effects, namely, to efforts to perpetrate fraud and to defensive efforts against fraud and mistake.

7. Duress

It may happen in contractual situations that a party is in duress or finds himself in an emergency of some type. For example, a ship may be sinking and need to be saved and pulled to harbor by another vessel, or a person may be at the airport during a snow storm and need a taxi to attend a very important business meeting. In such cases, the party under duress would be willing to make a contract under particularly disadvantageous terms. Let us evaluate the desirability of a legal rule that refuses to recognize certain contracts made under duress. In fact, the law generally refuses to enforce contracts made under duress.⁶⁰

7.1 Induced duress. Suppose first that situations of duress are engineered by other parties. For example, someone might direct an inexperienced sailor toward a dangerous area, and then, when he gets into trouble, come to his rescue, but only for a high price. It is clear that in

⁵⁹The general subject of the economic analysis of legal disclosure requirements relating to contracts is addressed in Kronman 1978a and Shavell 1994; for an empirical study, see Mathios 2000. Kronman originally emphasized the possibility that requirements might dull incentives to acquire information, but did not take into account the distinctions made above and in Shavell 1994 between buyers and sellers and between socially valuable and only privately valuable information.

⁶⁰See, for example, Calamari and Perillo 1998, 308-21, and E. Farnsworth 1999, 264-76.

such situations of induced duress, it would be best not to enforce the contracts. If the contracts are not enforced, then the motive of parties to create situations of duress will be removed. Otherwise, social losses will arise: the effort made to engender the dangerous situations is a social waste, rescue effort is a social waste (since it would not be needed but for the induced occurrence of the dangerous situations), and, when rescue does not come about, serious harm can occur.

7.2 Naturally occurring duress. Now consider situations of duress that occur naturally, such as when ships get into trouble themselves and need rescue. Here there are still reasons for disallowing contracts in which rescuers, or more generally, contracting partners, obtain very high prices.

First, allowing exorbitant prices to be charged imposes risk on individuals, for they realize that should they find themselves in bad straits, they will have to pay a large amount to extricate themselves. Presuming that they are risk averse, they will be better off if the law is such as to provide them with an implicit insurance policy in the form of refusal to enforce contracts with very high prices.

Second, risk aversion aside, allowing high prices for rescue might lead individuals to take excessive precautions in order to prevent the occurrence of situations in which they would need to be rescued. If it is in fact socially inexpensive to provide rescue to individuals in emergencies, then we would not want them taking extraordinary steps to prevent rescue situations from ever arising. Yet individuals would do that if they would have to pay extremely large amounts for rescue.⁶¹

An important qualification to these arguments is suggested by the example of rescue at sea, namely, that the likelihood of rescue may be affected by the contract amount; the higher the amount, the more vessels will be willing to engage in rescue attempts and the greater the effort they will devote to an attempt.⁶² This implies that in some cases it may be best for the law not to insist on low prices for enforcement of contracts, but to allow for premiums in prices to reflect

⁶¹ To amplify, suppose that individuals who might be in duress and potential rescuers are risk neutral. Let x be an individual's effort to prevent a dangerous situation of duress from arising and $p(x)$ be the probability of such a situation, where $p'(x) < 0$ and $p(0) > 0$. Suppose also that if duress arises, a potential rescuer's effort is y and $q(y)$ is the conditional probability of rescue, where $q'(y) > 0$ and $q(0) < 0$. Let v be the individual's losses if rescue fails. The social object is to minimize expected effort plus losses: $x + p(x)[y + (1 - q(y))v]$. Let the optimal values of x and y be denoted x^* and y^* , and assume that they are positive. The optimal rescue effort clearly minimizes $[y + (1 - q(y))v]$, so satisfies $q'(y)v = 1$. The optimal effort to prevent duress thus satisfies $-p'(x)[y^* + (1 - q(y^*))v] = 1$. That is, the value of raising x is that it reduces the likelihood of *minimized* expected losses in the event of duress, which are less than v (as y^* minimizes $y + (1 - q(y))v$ over y , we know that $y^* + (1 - q(y^*))v < 0 + (1 - q(0))v \neq v$), and much less than v if the rescue likelihood $q(y^*)$ is high. Now suppose that an individual anticipates that if he is in duress, he will be led to agree to pay a high price to a potential rescuer, due to enforcement of contracts made at such times. In particular, suppose the rescue contract price is k , where $k > y^* + (1 - q(y^*))v$ (in fact, the individual in duress would be willing to pay as much as v). Since the individual will choose x to minimize $x + p(x)k$, where $k > y^* + (1 - q(y^*))v$, he will choose $x > x^*$.

⁶² In the model of the previous note, it is assumed that rescue likelihood depends on rescue effort y . If the rescuer and the individual in duress cannot specify rescue effort in their contract, they will want payment to depend on the success of rescue effort, in which case, the higher the contract payment, the greater will be y . If the rescuer and the individual can contract on y , they will specify y^* . (This is consistent with the previous note, as the price k there can be interpreted as an expected contract price or a certain contract price.)

the desirability of inducing rescue.⁶³

⁶³Two further points may be mentioned. One is that in many contexts, a modest payment will induce most of what can be done, and is optimal to do, to effect rescue. Thus, disallowing very high prices may have little effect on the rescue probability $q(y)$. Second, allowing high prices for rescue may lead to a socially excessive number of parties engaging in rescue activity (by the logic leading to the conclusion that there is too much fishing activity). Both of these points suggest that the tradeoff between incentives to rescue and the disadvantages of high prices favors giving greater weight to the latter. In any event, under admiralty and maritime law, contracts for salvage are enforced, but if the vessel was in an *in extremis* situation, courts often set aside the contracts if compensation is excessive; see, for example, Schoenbaum 2001, 846-47. On economic analysis of rescue and duress, see Landes and Posner 1978.

Chapter 15

PRODUCTION CONTRACTS

In this chapter, I will discuss a significant type of contract -- that for production of a good (or for performance of a service⁶⁴). The aim will be to develop a more detailed understanding of important themes emphasized in the overview of contracts in chapter 13. In particular, I will consider here the nature of a completely specified contract, how damage measures serve as implicit substitutes for completely specified contracts, reliance activities, and renegotiation of contracts.⁶⁵

1. Completely Specified Contracts

1.1 Assumptions about the contractual situation. It will be supposed that there are two parties, a buyer and a seller, that the buyer places a value on having a good, and that the seller will need to produce the good.

In general, it will also be assumed that there are many sources of uncertainty that the two parties might face when they meet. The buyer might be uncertain about the value of the good to him (suppose it is a machine for production in his business, but the future demand he will face is uncertain). The seller might be uncertain about his production costs (suppose material costs may rise unexpectedly, or that unanticipated difficulties in the production process may arise), and the seller might be uncertain about whether another buyer will appear and bid more than the contract buyer for what he is producing. Mainly for expositional convenience, I will generally restrict attention to a situation with just one source of uncertainty -- about the seller's production cost. But I will occasionally remark about conclusions under other sources of uncertainty.

The two parties will generally be assumed to be risk-neutral, and I will discuss the implications of their being risk averse in separate sections.

The contract price will be presumed to be paid at the time of performance. This assumption too is mainly one of convenience, and I will make occasional comments about the situation if payment is made at the outset or at different points of time.

Other assumptions will be made in the course of the discussion.

1.2 Mutually beneficial completely specified contracts. A completely specified contract is a contract that contains a provision for each and every contingency. Under the simplifying assumption that there is only one source of uncertainty, a complete contract is a contract that provides explicitly for each possible level of production cost that the seller might encounter. Therefore, the contract states, for each possible production cost, whether or not the seller will perform.

A mutually beneficial contract, recall, is a contract that cannot be modified in a way that would make both parties better off. As was earlier noted, we would expect contracts to be

⁶⁴For concreteness, I will usually speak of contracts for production of goods rather than for performance of services even though the analysis of the two is the same.

⁶⁵In many respects, I will be using as a basis my articles Shavell 1980a, 1984a, but I will also be depending on numerous other articles, as I will note. See also the surveys Edlin 1998, Kaplow and Shavell 2002a, Mahoney 2000, Schwartz 1998, and Shavell 1998.

mutually beneficial, because we would predict that if they could be modified so as to please both parties, that would happen. I now consider an example illustrating the fundamental point that *mutually beneficial completely specified contracts call for performance if and only if the value of performance exceeds its cost.*

Example 1. Suppose that the value of a machine to the buyer is 100 and that there are three possible costs of production, as shown below.

| Production Cost | Probability |
|-----------------|-------------|
| 20 (low) | 30% |
| 60 (moderate) | 50% |
| 200 (high) | 20% |

Assume that the parties contemplate a completely specified contract, naming whether there shall be performance in each of the three contingencies as to production cost. Let us show that if the contract does not specify production when the cost is 20 and 60 and no production when the cost is 200, then *both* parties would want to change the contract.

Suppose initially that the parties discuss a contract under which there is to be performance under all three contingencies and a price of, say, 80, to be paid at performance.⁶⁶ The value of the contract to the buyer would be $100 - 80 = 20$ since he is assured performance. The expected value of the contract to the seller would be $30\% H(80 - 20) + 50\% H(80 - 60) + 20\% H(80 - 200) = 4$. Now suppose that the contract is altered, so that the seller has to perform only when the cost is 20 or 60. This is an advantage to the seller, for he will then not have to perform when it costs 200 and he would suffer a loss of 120 (which causes him an expected loss of $20\% H 120$ or 24). Thus, the seller should be willing to accept a lower price in exchange for altering the terms of the contract. In particular, suppose that the price is lowered from 80 to 65. Then the seller's expected value will be $30\% H(65 - 20) + 50\% H(65 - 60) = 16$, so he will be better off under the altered contract (16 exceeds 4). Likewise, since the buyer will receive performance with probability 80 percent, his expected value will be $80\% H(100 - 65) = 28$, so he too will be better off (28 exceeds 20) under the altered contract.

Next suppose that the parties initially discuss a contract calling for performance only when the production cost is 20 and a price of, say, 50. Under this contract, the buyer's expected value would be $30\% H(100 - 50) = 15$ and the seller's expected value would be $30\% H(50 - 20) = 9$. Now consider again a modified contract, calling for performance whenever production cost is 20 or 60. This contract would be better for the buyer since he will receive performance more often, so he should be willing to pay a higher price. Let us suppose that he pays a price of 70. Then the buyer's expected value will be $80\% H(100 - 70) = 24$, so he will be better off. The seller's expected value will be $30\% H(70 - 20) + 50\% H(70 - 60) = 20$, so he will also be better off.

This illustrates that the mutually optimal complete contract calls for production when the

⁶⁶Recall that I assume for concreteness that payment is to be made at the time of performance. The main conclusions would not be altered were payment to be made at the outset or partially at the outset and partially at performance, as I will be commenting upon later.

cost is 20 or 60; that contract is preferred by *both* parties either to a contract that calls for production all the time or for a contract that calls for production less often.//

The point of the example, that the mutually optimal completely specified contract is such that there is production when and only when the production cost is lower than the buyer's valuation, holds in general. The underlying reasons are twofold. On one hand, if a contract has a term resulting in performance when production cost exceeds the buyer's value, the seller will want the term changed and be willing to reduce the price for that by enough to make the buyer agree to it. On the other hand, if the contract has a term allowing the seller not to perform when production cost is lower than the buyer's value, the buyer will want that term changed and will be willing to alter the price by enough to make the seller agree.

1.3 Comments. (a) A mutually optimal completely specified contract is fully determined as to the conditions of performance, so that we can speak of "the" mutually optimal conditions under which there is performance. What is not determined is the contract price. In general, there will be a range of prices that one can imagine the two parties agreeing to, with the property that, given the price, each of the parties will be better off than he would be if he made no contract. "Bargaining power," sophistication, or other factors, not discussed at present, will determine the particular price that is agreed upon.

(b) That the conditions under which there would be performance are entirely determined by mutual optimality is a powerful and strong theoretical conclusion that merits reflection. It says that the two parties themselves would be expected to draw up these contingencies as the ones calling for performance; for they would agree to alter any contract that called for performance under different conditions.

(c) One way of explaining why, under the mutually optimal contract, production occurs precisely when its value exceeds cost is that this means that the joint value of the contract is maximized -- the figurative pie that the parties have to share is maximized. Both parties can always be made better off if the pie is made larger, for then each can be given a larger slice of it. The way that division of the pie is accomplished is through variation of the contract price (raising the price gives the seller a larger slice, lowering the price gives the buyer a larger slice).

1.4 Risk aversion. If one or the other party to the contract is risk averse, how does that affect the mutually desirable completely specified contract that they would make? It would not alter the conditions of performance that they would specify: They would still decide that there should be performance if and only if the cost of performance is lower than the value of performance. The reasoning is essentially that given above (for instance, the variations in the contract terms and the price in Example 1 could be carried through, in modified form). Thus, the conditions of performance would be exactly the same as when parties were assumed to be risk neutral.

However, there is a significant difference in the contracts the parties would want to make when one or both are risk averse: The parties would want to reallocate the risk in a contingency in accord with its size and their willingness and capacity to bear risk. Thus, for instance, if the buyer is risk neutral and the seller is risk averse, the buyer would act implicitly as an insurer for the seller, so that the buyer would pay the seller a fixed amount and absorb any variation in the seller's costs (as in a "cost-plus" contract). In general, the risk in the production cost would be allocated so as to reflect the degrees of risk aversion of the two parties, as described by the theory of optimal risk-sharing.

2. Remedies for Breach and Incomplete Contracts

2.1 Damage measures given completely specified contracts. As was emphasized in chapter 13, the parties to a contract would want the terms of a mutually optimal completely specified contract enforced for sure; they would want *no* deviation from such contract terms. Thus, the parties would want a severe sanction -- a very high damage measure (or specific performance) -- to apply for any violation of terms. However, because the terms in fact would not be violated, the severe sanctions would never be applied. The sanctions would serve only to obtain adherence to contract terms. Any problematic circumstances for the seller, such as high production costs, causing him not to want to perform, will already have been included in the terms of a contract – allowing him not to perform -- if the contract is completely specified.

2.2 Assumption of incompleteness. As was also stressed in chapter 13, contracts are in fact incomplete, due to the cost of including provisions and to the difficulty courts would have in verifying contingencies, here production cost. In order to study the implications of incompleteness, it will be helpful to consider the assumption that the contract contains no specific terms at all; the contract merely states, “Seller shall make a good and deliver it to the buyer, who shall pay him price P.”

2.3 Assumption about breach. Let us also assume that there will be no renegotiation between the buyer and the seller when breach is contemplated. As we discussed in chapter 13, this assumption is often appropriate (and renegotiation will be considered later, in section 4). Thus, a party will be assumed to commit breach if his position after breach and payment of damages would be better than if he performed.

2.4 Behavior under damage measures. A basic measure of damages for breach is the *expectation measure*, which is defined to be the amount that, if paid, will put the buyer (or, more generally, the party that is the victim of a breach) in the position he would have enjoyed had the contract been carried out.⁶⁷ For instance, suppose that there is a contract under which the buyer is to receive a machine worth 100 to him and for which he is to pay a price of 75 at the time of delivery. Then if the seller commits breach and does not produce and deliver the machine, the seller must pay 25 under the expectation measure: If there had been performance, the buyer’s gain net of price would have been $100 - 75$, so the seller’s paying him 25 in the event of breach puts him in the position he would have enjoyed had there been performance. Let us illustrate how the seller will behave under the expectation measure.

Example 2. Suppose that the situation is as in Example 1, and that the contract sets a price to be paid at performance of 75. Then because the seller will have to pay 25 under the expectation measure if he breaches, he will clearly decide to perform if his production cost is 20 or 60, for in both cases he will make a profit. If, however, production cost is 200, the seller will breach, for were he to perform he would lose $200 - 75$ or 125, which is more expensive than paying only 25 in damages.//

In this example, and in general, the expectation measure leads to performance if and only if the (gross) value of performance exceeds production cost -- exactly when performance would

⁶⁷The term “expectation” is used because the buyer obtains an amount equivalent to what he expected; and this amount is sometimes called the *expectancy*. This is the usual measure of damages for breach of contract in major legal systems; see, for example, E. Farnsworth 1999, 784-91, on the United States, and Treitel 1988, 75-92, on other countries.

occur under the terms of a mutually desirable completely specified contract. That the expectation measure leads to this precise result can, perhaps, be better understood by considering a modification of the example just considered in which production cost is imagined to be continuously variable. In this case, it can be seen that if the production cost is any amount below 100, the seller will perform; for then he will either make profits or, if his costs exceed the price of 75, his losses will be less than 25, so that he will be better off performing than breaching. However, if his production cost exceeds 100 by any amount, he will commit breach and pay 25, because his losses will then be greater than 25 if he produces the machine.⁶⁸

Similar reasoning shows that damage measures exceeding the expectation measure, as well as specific performance, may lead to performance when its value exceeds production cost -- more often than would occur under the terms of a completely specified contract. Suppose that damages are higher than expectation damages of 25, such as 50, and consider the situation in which production costs are continuously variable. Then the seller will perform as long as his losses would be less than 50, which is to say, as long as production costs are less than 125 (for at 125 his losses would be $125 - 75 = 50$); thus he would be led to perform not only when production costs are less than 100, but also when production costs are in the range between 100 and 125, exceeding the value of performance of 100.⁶⁹

Conversely, damage measures that are less than the expectation measure may lead to breach even though the value of performance exceeds production cost -- performance occurs less often than under a completely specified contract. We will consider particular measures of damages that are less than the expectation measure below.⁷⁰

2.5 Mutually preferred remedy for breach: the expectation measure. It has been seen that under the expectation measure, but not under other remedies for breach, there will be performance in precisely the contingencies that would have been set out in a mutually optimal completely specified contract. Moreover, as was noted in comment (c) of section 1.3, the size of the pie to be shared by the parties is maximized under the mutually optimal completely specified contract. This suggests that the buyer and the seller would agree, *ex ante*, to employ the expectation measure rather than any other remedy for breach. And, in fact, this is true -- the following proposition can be demonstrated: Given any proposed remedy for breach of contract other than the expectation measure, one can replace the proposed remedy with the expectation

⁶⁸Algebraically, one can see this as follows. Let k be the contract price (75 in the example), c the cost of performance, and v the value of performance to the buyer. Then, if the seller performs, he makes profits of $k - c$ (which may be negative -- that is, losses). If the seller commits breach, he pays damages equal to the buyer's expectancy of $v - k$, which is to say, the seller makes profits of $k - v$. Hence, the seller will perform if and only if $k - c > k - v$, or, equivalently, if and only if $v > c$.

⁶⁹In the version of the example with just three levels of production costs, damages of 50 would not lead to excessive performance, because the only level of costs exceeding 100 is very high, 200; damages would have to be higher than 125 to induce performance in this circumstance. Thus, if production costs are not continuously variable, damages exceeding expectation might not lead to excessive performance if they are not too much higher than the expectation measure.

⁷⁰It should be noted that the statements made in this section apply regardless of whether the seller is risk neutral or risk averse, for the seller makes the decision to commit breach after he knows what production cost is, so that his decision at that point does not involve risk.

measure and adjust the contract price such that both the buyer and the seller would prefer the expectation measure and the modified price to the proposed remedy and the initial price.

To illustrate the reasoning underlying this conclusion, let me show first that, given a contract and the remedy of specific performance, or a very high damage measure, both the buyer and the seller would prefer to use the expectation measure if the contract price is lowered appropriately.

Example 3. Assume that the situation is as described in Example 1 and consider a contract calling for specific performance and a price of, say, 80. The value of this contract to the buyer would be $100 - 80 = 20$, as he will obtain performance for sure. The value of the contract to the seller would be $30\% H(80 - 20) + 50\% H(80 - 60) + 20\% H(80 - 200) = 4$.⁷¹

Now suppose that, instead, the parties consider a contract under which the expectation measure is employed. If the price is not changed from 80, we know that the buyer will be just as well off, for if he does not obtain performance, he will obtain its equivalent in damages. The seller, however, will be better off under the expectation measure, as he will have the opportunity, which he will take, to commit breach when the production cost is 200. More precisely, under the expectation measure, the seller will have to pay 20 if he commits breach (for if there were performance, the buyer would obtain $100 - 80 = 20$). The seller will be better off breaching and paying damages of 20 than sustaining a loss of 120 when production cost is 200 (note that $200 - 80$ is 120). The seller, however, will be led to perform when production cost is 20 or 60. Hence, the seller's valuation of the contract will be $30\% H(80 - 20) + 50\% H(80 - 60) + 20\% H(20) = 24$, so the seller will be better off by 20 under the expectation measure than he would be under specific performance.

Because the seller will be better off by 20 under the expectation measure and the buyer will be just as well off, it follows that if the price is lowered slightly from 80, the seller will still be better off and the buyer will now be strictly better off (rather than just as well off). Suppose, for instance, that the price is lowered to 75. The buyer's valuation will be $100 - 75 = 25$ (for he will obtain either performance or 25 in expectation damages), which exceeds the 20 he would have enjoyed under specific performance. The seller's valuation will be $30\% H(75 - 30) + 50\% H(75 - 60) + 20\% H(25) = 16$, so the seller will be better off than he was under specific performance, when his return was 4.//

The general logic employed in this example is worth restating. Under the expectation measure, the buyer is just as well off as he is under specific performance (or under a very high measure of damages): By definition of the expectation measure, the buyer receives the equivalent of performance if he does not obtain performance. But the seller is better off under the expectation measure than he is under specific performance because under the expectation measure he can commit breach when it would be very expensive for him to perform, whereas under specific performance he must perform. Because the seller is better off under the expectation measure, he will still be better off if he allows the price to be lowered somewhat, and this will make the buyer positively better off by agreeing to use the expectation measure.

Essentially the same logic can be applied to show that that the two parties would prefer

⁷¹If the damage measure were very high, say 150, the seller would perform even if the production cost were 200, for then he would suffer losses of 120, which is better than paying damages of 150. Hence, because he would always perform, the situation would be equivalent to that under specific performance.

the expectation measure to any damage measure exceeding the expectation measure that would result in excessive performance.

Now let us consider an example illustrating that the parties would prefer the expectation measure to a damage measure that is lower than the expectation measure and would result in excessive breach.

Example 4. Consider a contract under which the measure of damages for breach is 5 and the price is 50. Then the seller would perform if the production cost is 20, for he would make a profit of 30; he would commit breach if the cost is 60, for if he performed he would lose 10, and paying 5 in damages is better than that; and he would clearly commit breach if the cost is 200. Therefore the value of the contract to the buyer is $30\% H(100 - 50) + 70\% H 5 = 18.5$. The value of the contract to the seller is $30\% H(50 - 20) + 70\% H 5 = 5.5$.

Now suppose that the buyer proposes that the expectation measure be employed. If the price is not changed, the positions of the buyer and of the seller are as follows. The buyer's expected value is $100 - 50 = 50$. Because the seller would be led to perform when production cost is 20 or 60 and would pay damages of 50 otherwise, his expected profits would be $30\% H(50 - 20) + 50\% H(60 - 50) + 20\% H 50 = 6$, so he would sustain losses of 6.

Notice that the buyer's gain from switching to the expectation measure is 32.5 (namely, $50 - 18.5$) and exceeds the decrease in the seller's position of 11.5 (namely, the difference between gains of 5.5 and losses of 6). This suggests that if the price is raised by enough to compensate the seller for switching to the expectation measure, the buyer can still be left better off than under the lower measure of damages. To verify this, suppose that the price is raised to 70. Then the buyer's expected value is $100 - 70 = 30$, so he is still better off than with the 18.5 under the original contract with low damages of 5. The seller, who would pay 30 if he breached, and would choose to commit breach only when his cost is 200, would value the contract as $30\% H(70 - 20) + 50\% H(70 - 60) + 20\% H(30) = 14$, which exceeds 5.5. Hence, both the buyer and the seller would be better off under the contract with the expectation measure, and the altered price of 70, than under the contract with the low measure of damages of 5.⁷²

In this example, the reason that the buyer and the seller could each be made better off is that raising the measure of damages to the expectation measure induced the seller to perform in a contingency when the value of performance exceeded the cost. That in turn increased the value of the transaction to the buyer by more than it cost the seller. And this made it possible for the buyer to adjust the price by enough to make the seller willing to incorporate higher, expectation damages into the contract.⁷²

⁷²It may be helpful to sketch a proof of the general point that the parties would both prefer the expectation measure over any other measure of damages. This point was first emphasized in Shavell 1980a. Let v be the value of performance and c the uncertain production cost, where c is described by some probability distribution. Let k be the contract price, to be paid at performance. Let d be a damage measure: d determines how much the seller is to pay if he commits breach as a function of variables observable by the courts, and let us say, of c , v , and k . Given $d(c, v, k)$, a seller will breach if and only if $d(c, v, k) < c - k$ (for $c - k$ are his losses if he performs). Let $B_d(k)$ be the expected value of a contract to the buyer given a damage measure d and price k , and let $S_d(k)$ be the expected value to the seller. Let $T_d(k)$ be the total of their expected values.

Under the expectation measure E , the seller will pay $v - k$ if he commits breach, so he will perform when $v > c$ and breach when $c > v$. Hence, the total of the buyer's and seller's expected values $T_E(k)$ under the contract is the maximal possible value M (whatever is the contract price k): the sum of their values in any realization of c is either $v - c$, if there is performance, or 0 if not; thus, the sum is maximized given c if there is performance if and only if $v > c$, which is what occurs under the expectation measure; hence the expected sum is maximized under the

2.6 Generality of results about the expectation measure. The conclusion that the expectation measure is mutually preferred to other measures, and that it leads to the same behavior as would be observed in a mutually optimal completely specified contract, holds more generally than under the assumptions we investigated above. Let me mention three changes of assumption under which the conclusion continues to apply.

First, suppose that not only is the seller's cost uncertain, but also that the buyer's valuation is uncertain. In this context, a new factor is introduced: The buyer may be the party to commit breach. If his valuation falls below the seller's cost, it can be shown that the buyer will be led to breach and pay in damages the profits that the seller would have earned had he performed (which is the expectation measure for buyer breach). This behavior of the buyer is mutually optimal.

Second, suppose that payment is made at the outset instead of at the time of performance. This changes the amount that must be paid at breach under the expectation measure: If, for instance, the price is 70 and the value of performance is 100, a seller who breaches would have to pay the buyer 100 (not 30), for having already paid 70 at the outset, the buyer needs to receive 100 to be made whole. However, behavior under the expectation measure is the same as was described above (the seller obviously will commit breach only if his production cost exceeds 100), and the expectation measure remains the mutually preferred damage measure.

Third, suppose that the losses due to breach can be mitigated by the victim, for example, the buyer might be able to find an alternative supplier of a good or service and thereby limit his losses from breach. In this case, if the expectation measure is interpreted as the buyer's optimally mitigated level of losses, use of that measure will lead to desirable behavior in two regards: The seller will, as usual, perform if and only if the cost of so doing is less than the value of performance to the buyer; moreover, if there is a breach, the buyer will be led to take cost-justified steps to mitigate losses. For these reasons, the expectation measure can again be shown to be the mutually preferred damage measure.⁷³

Fourth, suppose that breach is not an intentional act but a probabilistic phenomenon:

expectation measure. Moreover, M can be divided between the parties in any way, by a suitable choice of the contract price k : it is clear that if k is low enough, the seller's expected value $S_E(k)$ is 0; that $S_E(k)$ rises continuously with k ; and that if k is high enough, $S_E(k) = M$.

Now consider a contract with price k and any damage measure d different from the expectation measure E . Because the damage measure is different from the expectation measure, behavior will in general be different under it -- either there will be performance when $c > v$ or failure to perform when $c < v$, or both sometimes will occur. Hence, the expected sum of values $T_d(k)$ that the parties obtain under d is not maximized and is less than M , that obtained under the expectation measure. For this reason, and because we know that under the expectation measure, M can be divided in any way by a suitable choice of the contract price, there must exist a price, say k' such that, under the expectation measure, $B_E(k') > B_d(k)$ and $S_E(k') > S_d(k)$, that is, such that both buyer and seller are better off under the expectation measure.

⁷³To amplify, let z be the mitigation expenditure of the buyer to raise his post-breach alternative value, say $w(z)$. It can readily be shown that it will be in the parties' joint interests for z to maximize $w(z) - z$; let z^* be this optimal value of z . If y is the gross value of seller performance to the buyer, then we can define v , the net value of performance, as $v = y - (w(z^*) - z^*)$. Thus, expectation damages should equal this v , not the gross value y , if z^* is in fact chosen by the buyer. And if damages equal v , the buyer will choose z^* if he is the victim of a breach, for the buyer's damage payment will be v independently of his choice of z , so he will choose z to maximize $v + w(z) - z$. On this issue of mitigation of the consequences of breach, see Wittman 1981.

suppose that the seller invests in satisfactory performance, such as increasing his level of care to assure timely delivery, and that then performance either occurs or fails to occur due to chance elements. In this case as well, it can be shown that the expectation measure leads to mutually preferred behavior in the sense that the expectation measure induces the seller to exercise the level of care that would have been set forth in a completely specified contract,⁷⁴ and that the expectation measure is the mutually preferred damage measure.

However, the expectation measure is not mutually preferred under all generalizations of our assumptions. As we will see shortly, when parties are risk averse or when reliance decisions are at issue, the expectation measure may not be the mutually preferred remedy.

2.7 Risk aversion. If parties are risk averse, then damage measures play a dual role. Not only do they induce parties to perform or allow them to commit breach, damages measures also allocate risk, and both elements have to be taken into account in determining how well a damage measure serves the parties' purposes.⁷⁵

The expectation measure imposes risk on the seller, who might commit breach, and implicitly insures the buyer against nonperformance.⁷⁶ If the seller is risk neutral and the buyer is risk averse, that allocation of risk would be desirable, so that the expectation measure would be mutually desirable on grounds of both risk sharing and incentives to perform. Indeed, it would result in exactly the sharing of risk that would have been named in a completely specified contract between the parties (see section 1.4).

However, the expectation measure might also be mutually undesirable, because it may impose excessive risk on a risk-averse seller. Suppose, for instance, that the seller is a small company (say a several-person maker of machine tools) and the buyer is a large enterprise (a car manufacturer) that would face a substantial reduction in profits (the production line would have to stop) if the small company does not perform. In this case, the seller could easily be imagined to be unable or unwilling to bear the risk of paying the expectation damages, or else would charge a very high price to compensate it for bearing the risk. In a situation in which the expectation measure has such disadvantages on grounds of risk bearing, another damage measure, a lower one, may be mutually desirable even though it results in performance less often than is ideal.

An alternative approach that the parties may pursue when a damage measure, such as the

⁷⁴In other words, the situation is like that discussed in Part 1 of the book, on accidents, where here an accident is interpreted as a breach. Let x stand for the investment in care that the seller makes, $p(x)$ the likelihood of breach -- failure of successful performance -- and v the value of the contract. It is readily shown that if the parties made a completely specified contract and named x , they would choose the level of x that maximized the expected sum of values, namely, $(1 - p(x))v + x$, which is to say, they would choose the x that minimized $p(x)v + x$; call this x^* . Now if they make an incomplete contract not mentioning x , but stating that the seller should, under the expectation measure, pay the buyer $v - k$ in the event of breach, the seller will choose x to maximize $(1 - p(x))k + p(x)(v - k) + x = k + (p(x)v + x)$, so he will choose x^* as claimed. Probabilistic breach is discussed in Bebchuk and Png 1999, Coote 1985, and Craswell 1988.

⁷⁵On risk aversion and damage measures, see Polinsky 1983 and Shavell 1984a. On ways in which legal doctrines serve to allocate risk in the light of risk aversion, see, for example, Joskow 1977, Posner and Rosenfield 1977 and Sykes 1990.

⁷⁶I am here again assuming that the uncertainty in the contractual situation is only over production cost (not over buyer valuation), so that the party who might commit breach is the seller. The general points to be made will obviously carry over to the more general situation in which the buyer also might commit breach if his valuation falls.

expectation measure, results in excessive risk-bearing is to include more explicit provisions in their contract in order to achieve performance when and only when it is mutually desirable. The contract could read that the seller will perform unless he runs into serious, named problems, and in that case he will be excused from his obligation to perform and need not pay damages. To write such a contract will take some time and will necessitate the buyer's later having to verify that the seller indeed encountered the stipulated problems when he claims that he did. But these difficulties and expenses may be worthwhile for the parties to incur in order to avoid the disadvantages attending the bearing of risk by the seller or the inefficiencies of employing a smaller measure of damages than the expectation measure.

With regard to specific performance as a remedy, a point worth stressing is that it imposes a heavy risk on the seller. Under specific performance, the seller will face a risk that is potentially unlimited, or as large as his cost of performance could be.

2.8 Liquidated damages versus court-determined damage measures. If the parties' mutually preferred measure of damages is one that the courts would apply, then the parties have no need to name damages in their contract and can save themselves the trouble of so doing (although this cost of naming damages would not seem to be great). In fact, the usual measure of court-determined damages is, as was noted, the expectation measure, so that if the parties do not name damages, the measure that would be employed will have attractive properties for them. However, often the parties will not have confidence that the courts will employ the measure that they would desire. This is so even if they want the expectation measure, for the courts may not be able to determine the value of performance; notably, if the courts used too low a level, the parties would suffer from too little performance. Hence, parties will often want to name the measure of damages in a liquidated damages provision.⁷⁷

Courts in fact generally enforce liquidated damages provisions, something that is socially desirable because it amounts to allowing parties to select the mutually preferred remedy for breach of contract. However, courts sometimes refuse to enforce damages that seem excessive (that constitute "penalties") in relation to the value of performance.⁷⁸ This tendency of courts is undesirable if the damages represent what the parties wish, because lower court awards will reduce the frequency of performance and the value of the contract. However, if the contract with high damages is not mutually desirable, but rather is the result of a party's failure to consider the contract carefully, then the courts' refusal to enforce it may be desirable, an implicit form of insurance against mistaken behavior at the time of contracting.⁷⁹

2.9 Asymmetric information and remedies for breach. The buyer's value of performance may not be known to the seller at the time of contracting, and this may influence the remedy for breach that the courts will employ (if the parties do not specify a liquidated damages provision). For example, in a famous English case, a factory owner wanted a broken

⁷⁷Yet that may not be an option for the parties: If the value of performance to the buyer v is uncertain, the parties cannot name it in advance; if the value of having a factory constructed on time will vary, due to market conditions for the product the buyer is going to produce in the factory, then the parties cannot specify the damages to be paid in advance (although they might specify in advance a procedure to determine value).

⁷⁸See, for example, E. Farnsworth 1999, 841-50, on the United States, and Treitel 1988, 223-28, on other countries.

⁷⁹For economic analysis of liquidated damage provisions, see, for example, Clarkson, Miller, and Muris 1978, Goetz and Scott 1977, and the surveys De Geest and Wuyts 2000 and Rea 1998.

mill shaft transported to a manufacturer for replacement, and the value of timely delivery was high because the factory was unable to operate without a replacement; but the shipper did not know that performance was so important.⁸⁰ In such situations, courts tend to employ the value of performance that the seller is likely to have believed when he made the contract -- so that if a buyer's value was higher than usual, he will not receive a higher than usual level of damages. This legal rule regarding damages tends to induce buyers to reveal their high valuations at the time of contracting, for that will lead to more frequent performance, to their benefit, even though it will also result in their having to pay a higher price for their contracts. Thus, the rule under consideration appears to be socially desirable (but there are complicating issues).⁸¹

The seller's lack of knowledge of the buyer's valuation may also influence the choice of a liquidated damages measure. Because the price charged to the buyer may rise if he reveals his valuation or if he does not but asks for a high measure of damages, the buyer may purposely ask for damages in a liquidated damages provision that are lower, perhaps significantly lower, than his valuation. Although he sacrifices a gain in the frequency of performance, he may be better off because he avoids a larger increase in price paid.⁸²

2.10 Contrast to views of legal scholars. The general views developed in this section stand in contrast to those emphasized in most legal scholarship about remedies for breach, and two related points are worth noting.⁸³ The first was mentioned in section 6.3 of chapter 13, that the view of most scholars is that breach is in some sense a morally bad act, because it resembles a broken promise. Breach may be regarded somehow as a practical necessity in some circumstances, but it is not intrinsically a good thing. This view, as was stressed here and before, fails to take into account the difference between the notional, completely specified contract that truly represents the wishes of the parties, and the incomplete contract that the parties in fact make. Against that background, the function of damages measures can be seen to be, in substantial part, to induce performance when and only when the parties would have wanted it. In particular, this means that breach amounts to nonperformance when the parties would have allowed for nonperformance in a detailed contract, and breach is thus not an undesirable act but a desirable, good act from the standpoint of the parties and their true wishes.

The second contrast between legal scholars' orientation and that of the economic view concerns scholars' tendency to focus on the conflict of interest of the parties at the time of a

⁸⁰The case is *Hadley v. Baxendale*.

⁸¹The principal complicating issues are two. First, costs are involved in communicating values, and second, sellers would have an incentive to ask buyers to state their valuations if the legal rule were that damages would equal buyers' actual valuations whether or not they revealed them. When these issues are taken into account, it turns out that, although the legal rule in question is socially desirable when it is socially best for there to be (costly) communication between parties, the rule may be undesirable when that is not true. For economic analysis of these issues, see Ayres and Gertner 1989 and Bebchuk and Shavell 1991.

⁸²In other words, although the joint value of the contract to the parties would be maximized if the buyer were to reveal his valuation and set damages for breach to reflect it, the distribution of the gains from contracting might move unfavorably against the buyer were he to set damages in that way. For further discussion of this and related issues, see Johnston 1990, Spier 1992b, and Stole 1992.

⁸³These points are emphasized in Shavell 1980a.

possible breach. At that time, it is, of course, true that the more that a party has to pay if he commits breach, the worse off he will be, and the better off the party expecting performance will be. This observation leads naturally to the view that the damage remedy is something that reflects a conflict of interest between the parties, and that the best remedy is that which resolves the conflict in a “fair” way. Such an orientation overlooks the desires of the parties at the time that they make contracts and the role of damage measures in furthering these wishes of the parties. In particular, and most significantly, the traditional scholars’ view fails to reflect the point that a measure of damages may be mutually desired by the parties (the expectation measure given our assumptions and risk neutrality, or some other measure given other assumptions). That is, the view fails to recognize that there will be, in general, a measure that both parties would prefer to any other measure, so that naming a different measure, even if deemed fair at the time of breach, would make both parties worse off *ex ante*.

3. Reliance

3.1 Definition of reliance. As was briefly discussed in chapter 13, there are often actions that can be taken in advance of performance that will enhance its value to a party to a contract. A buyer expecting to receive a machine can train his employees to use it, move out an old machine, or do something else that will increase its value to him. Someone expecting a singer to appear at his nightclub can advertise the singer’s appearance, thus increasing the number of people who will come to the club for the appearance. Such value-enhancing actions of parties to contracts will be called reliance actions, or simply *reliance*, because they are taken in reliance on performance.⁸⁴ I will assume for clarity that reliance actions must be undertaken before contractual uncertainty is resolved and thus before it is known whether there will be performance.⁸⁵

3.2 Reliance as the measure of damages. When a party has relied on performance by making an expenditure and the other party commits breach, a possible measure of damages would compensate the victim of the breach for his reliance expenditure -- so that the party is restored to the position he had before he made the contract; the amount that accomplishes this is the *reliance measure* of damages. The reliance measure of damages is ordinarily less than the expectation measure, and it will thus tend to result in more frequent breach than does the expectation measure. Therefore, both parties will prefer the expectation measure to the reliance measure of damages.⁸⁶ Consider the following illustration.

Example 5. Suppose that in Example 1 the buyer must spend 5 in reliance on

⁸⁴Although I will generally interpret reliance to be an expenditure or effort that is made relying on performance, another meaning of the term reliance is an opportunity that is forgone on account of making a contract. For example, if the contract buyer forgoes making an alternative contract because he makes the one in question, then the expected value of the alternative contract might be considered his reliance. What I say below will, in some respects, apply to this alternative form of reliance.

⁸⁵Economic analysis of reliance is introduced in Shavell 1980a, which I largely follow here. The analysis in that article presumed, however, that parties do not renegotiate their contracts. On the important subject of renegotiation and reliance, see section 4.3 below.

⁸⁶This point, that parties prefer expectation damages to reliance, is also true under the interpretation of reliance that equates it with forgone opportunities.

performance in order to obtain the benefit from it of 100. Consider a contract in which damages for breach are governed by the reliance measure, so that damages are 5; and assume also that the contract price is 50. Then the seller will commit breach when the production cost is 60, for he would prefer to spend 5 in damages rather than lose 10, and he will also commit breach when the production cost is 200. The value of the contract to the buyer will thus be $30\% \text{ H}(100 - 50) + 70\% \text{ H } 5 = 13.5$, and the value to the seller will be $30\% \text{ H}(50 - 20) + 70\% \text{ H } 5 = 5.5$.

The parties will both be better off under the expectation measure, presuming that the price is raised suitably to compensate the seller for having to pay larger damages. Suppose that the price is 75 and that expectation damages of 25 are employed. Then the seller will perform when production cost is either 20 or 60. The buyer's value will be $100 - 75 + 5 = 20$, which exceeds 13.5, so he will be better off; even though the price is higher, the increased frequency of performance more than makes up for the price rise. The seller's value will be $30\% \text{ H}(75 - 20) + 50\% \text{ H}(75 - 60) + 20\% \text{ H } 25 = 19$, which exceeds 5.5, so he too is better off.//

As is evident from this example, the reason that the reliance measure is inferior to the expectation measure is that reliance damages are only 5, and thus less than expectation damages. That reliance expenditures, 5 in the example, are less than the value of performance, 100 in the example, is a plausible general assumption because it would be irrational for a contracting party to plan to spend more on reliance than performance is worth; otherwise the contract would have negative worth to the party spending on reliance.⁸⁷

3.3 Comment on views of legal scholars about the reliance measure. Some legal commentators favor the reliance measure over the expectation measure on grounds that we need not review here.⁸⁸ The point worth noting is that because they do not take into account the mutual interests of the contracting parties at the outset, they do not recognize that the expectation measure would be chosen by both parties over the reliance measure in important, general circumstances. Indeed, the contracting parties' preference would be quite strong for the expectation measure over the reliance measure if performance is very important and reliance by the buyer is small, so that under the reliance measure of damages there would be very little assurance of performance. That would reduce the value of the contract substantially, and make the buyer much less willing to pay the seller than otherwise.

3.4 Level of reliance; optimal reliance. A party to a contract often has a choice over whether, or how much, to rely on contract performance (for simplicity, the level of reliance was taken as fixed above). Reliance will be said to be optimal if it maximizes the expected joint value of a contract to the buyer and the seller, that is, if it maximizes the expected value of performance less production cost and less the costs of reliance. Let us consider an example before further discussion.

Example 6. Suppose that in the example that we have been considering, the following is assumed about the level of reliance and the value of performance to the buyer. If reliance is 0,

⁸⁷If the buyer's valuation were uncertain, however, it could turn out that his valuation happens to fall below the level of reliance. This would result in too little breach under the reliance measure. Whether the reliance measure is below the expectation measure or above, the expectation measure will be preferred by both parties given the general assumptions now under consideration.

⁸⁸In a much-cited article, Fuller and Perdue 1936 favor the reliance measure over the expectation measure as a matter of theory, although they often find the expectation measure attractive, mainly for reasons of administrability.

the value of performance is 50; if reliance is 5, the value of performance is 100; and if reliance is 50, the value of performance is 150. (The costs of performance are as has been assumed -- 20 with probability 30 percent, 60 with probability 50 percent, and 200 with probability 20 percent.)

What level of reliance is optimal -- will lead to the highest expected value of performance net of production cost and of the cost of reliance? If reliance is 0, then because the value of performance will only be 50, it will be optimal for production to occur only when the production cost is 20, so the joint value of the contract will be $30\% H(50 \mid 20) = 9$. If reliance is 5, the value of performance will be 100, so production will be optimal when its cost is 20 or 60, so the joint value of the contract will be $30\% H(100 \mid 20) + 50\% H(100 \mid 60) \mid 5 = 39$. If reliance is 50, the value of performance will be 150, so production will again be optimal when its cost is 20 or 60, so the joint value of the contract will be $30\% H(150 \mid 20) + 50\% H(150 \mid 60) \mid 50 = 34$. Thus, reliance of 5 is optimal.

It should be noted that increasing reliance from 5 to 50 involves an extra cost of 45 and results in an enhancement of 50 in the value of performance, from 100 to 150. Why then is 50 not the optimal level of reliance? The answer is that the extra 45 cost of reliance is incurred with *certainty*, whereas the benefit from performance comes only with a *probability*, of 80 percent, so that the expected value of the enhancement in the value of performance is only 80 percent of 50 or 40, which is lower than 45.//

This example illustrates that an expenditure on reliance is optimal only if its *expected* benefits exceed the expenditure; if performance is not optimal, then reliance will have been a waste.

If parties make a mutually optimal completely specified contract and it stipulates the level of reliance, that level will be the optimal level of reliance; for as explained above (see section 1.3(c), the parties will want to arrange the contract to maximize its joint expected value.⁸⁹

3.5 Remedies for breach and the level of reliance. Consider now incomplete contracts and the question of how remedies for breach affect decisions about reliance.

Under the expectation measure, there will be a tendency toward excessive reliance. The reason for this is that, because the buyer is implicitly insured by receipt of damages against losses from breach, he will view his investment in reliance as one with a certain payoff. Yet, as just emphasized in section 3.4, reliance has in fact only a probabilistic payoff, and that needs to be taken into account in determining optimal reliance. The next example shows that under the expectation measure, the buyer will tend to choose more than the optimal level of reliance.

Example 7. Suppose that, in the circumstances of the previous example, the contract price

⁸⁹To be precise about the notion of optimal reliance, let r be the buyer's reliance investment and $v(r)$ be the value of performance given r , where v is increasing in r . The buyer chooses r before the seller learns the production cost c and decides about producing. The optimal decision of the seller is to produce if and only if $c < v(r)$, and the joint value maximizing decision of the buyer is therefore to choose r to maximize

$$\int_0^{v(r)} (v(r) - c)g(c)dc \mid r,$$

where g is the probability density of c . Thus, the optimal r , denoted r^* , is determined by $v'(r)G(v(r)) = 1$, where G is the cumulative distribution of c . The point to note here is that the marginal return to reliance investment is only a contingent return, for the investment pays off only with probability $G(v(r))$, that is, when $c < v(r)$ (when production turns out to be efficient).

is 70 and the buyer is deciding whether to choose 5 or 50 as his level of reliance.⁹⁰ If he selects 5, then because 100 would be the value of performance, the buyer will either obtain performance, and enjoy a net value of 30 or receive 30 in damages; hence the buyer's valuation would be $100 - 70 - 5 = 25$. If he chooses 50 as reliance, 150 would be the value of performance, and his valuation would by similar logic be $150 - 70 - 50 = 30$. Thus, he would choose 50 as his level of reliance. When the buyer increases his level of reliance to 50, he benefits for sure because he effectively receives 150 either in performance or its equivalent in the form of expectation damages.//

As this example illustrates, the reason that the expectation measure leads to too much reliance is that the buyer knows that if he increases reliance, not only will he obtain more if there is performance, he will also obtain more if there is a breach, because the damages put him in the position he would have enjoyed had there been performance.⁹¹

The conclusion under the expectation measure exemplifies a more general point about the receipt of damages: Because damages dull the impact of breach, they tend to lead to excessive reliance, for optimal reliance requires a party to take into account that breach means that reliance is a waste.

Under the reliance measures of damages, there is excessive reliance not only for the general reason just given, but also for another reason having to do with the fact that, because damages are less than expectation, the buyer will be made worse off if there is a breach. Hence, the buyer will want to reduce the likelihood of breach, and this in turn he can accomplish by increasing reliance -- for the higher is reliance, the more the seller will have to pay in damages if he breaches, and thus the less often he will commit breach. For this reason, it can be shown that the level of reliance undertaken under the reliance measure of damages tends to be even more excessive than reliance is under the expectation measure.

The fact that the expectation measure and other damage measures often result in improper reliance complicates the determination of the mutually desirable damage measure. The best measure will represent an implicit compromise between providing proper incentives to rely and proper incentives to perform. There does not exist any damage measure that provides optimal incentives both to perform and to rely, for only the expectation measure provides optimal incentives to perform, yet it does not provide proper incentives to rely.⁹²

3.6 Further on reliance and remedies for breach. Let me make several additional remarks about the theory of reliance and about the interpretation of what has been said.

⁹⁰For simplicity, I will not consider whether he would want to choose 0 as the level of reliance; it is easy to show that he would not.

⁹¹Specifically, if the expectation measure is employed, the buyer will always receive $v(r) - k$ (either he obtains performance, worth $v(r)$, and pays the contract price k , or he receives damages of $v(r) - k$). Hence, he will choose r to maximize $v(r) - k - r$. Consequently, r will be determined by $v'(r) = 1$ rather than by $v'(r)G(v(r)) = 1$ (see note 26), so the buyer will select an inefficiently high r ; the problem is that the buyer does not take into account that investment does not have any value when performance does not occur. This point was first made in Shavell 1980a.

⁹²More precisely, this statement that there does not exist any damage measure that leads to optimal performance and optimal reliance assumes that a damage measure d is a function only of the variables v , r , c , and k -- the value of performance, reliance, production cost, and the contract price. In particular, d may not depend on more information, such as about the function $v(r)$ or the density $g(r)$. See Shavell 1980a.

(a) *Sophisticated damage measures.* On reflection, it can be seen that the tendency toward excessive reliance caused by receipt of expectation damages can be combated if the level of damages is not automatically increased to reflect the actual value of performance. If the expectation measure of damages is sophisticated in the sense that damages are set equal to the level reflecting *optimal* reliance, there will be no incentive to rely excessively, for increasing reliance will not raise damages received in the event of breach. (This will be true in Example 7, for instance, if expectation damages are set presuming that reliance is 5.⁹³) Indeed, in general this sophisticated expectation measure leads to both optimal reliance and to optimal performance.⁹⁴ However, use of such a sophisticated version of the expectation measure requires that the court knows much more than the actual level of reliance and the actual value of performance; it must know the functional relationship between reliance and the value of performance and the entire probability distribution of production costs -- everything about the contractual situation -- in order to calculate optimal reliance. The parties themselves, though, would often be presumed to have approximately enough information to determine optimal reliance (or much more than the court), and so could name the expectation measure given optimal reliance in a liquidated damages provision.

(b) *More general forms of reliance.* In the discussion above, we considered reliance by the buyer that raised the level of performance. However, more generally, the seller can also take actions that will alter the value of the contract, perhaps that will lower his expected production cost or the quality and value of performance for the buyer. These issues introduce new considerations into the analysis of damage measures and reliance. For example, to the extent that seller reliance augments the buyer's value of performance, it may well be beneficial jointly for the seller to undertake that reliance, but he will hardly have an incentive to do so if this were to increase the damages he would have to pay under the expectation measure.

(c) *Damage measures still needed to induce reliance.* Although damage measures may be imperfect and may often lead to reliance exceeding the optimal level, one should not lose sight of the fact that the use of damage measures does generally induce parties to invest in reliance by giving them reason to believe that performance is likely to occur. Were there no contract enforcement, there would be very little reliance.

4. Renegotiation

4.1 Introduction. To this point in the present chapter, the possibility of renegotiation of

⁹³If 5 is reliance, the value of performance will be 100, so the expectation measure will be $100 - 70 = 30$. As a consequence, the seller will perform when the cost is 20 or 60. Hence, the buyer will reason as follows. If he chooses reliance of 0, the value of the contract to him will be $80\% \text{H}(50 - 70) + 20\% \text{H} 30 = 10$. If he chooses reliance of 5, the value of the contract will be $80\% \text{H}(100 - 70) + 20\% \text{H} 30 - 5 = 25$. If he chooses reliance of 50, the value of the contract will be $80\% \text{H}(150 - 70) + 20\% \text{H} 30 - 50 = 20$. Hence, he will choose reliance of 5, the optimal level, as claimed.

⁹⁴Recall from note 26 that r^* denotes optimal reliance. If $d = v(r^*) - k$, the seller will perform when $c < v(r^*)$, so the buyer will maximize his expected return, $(v(r) - k)G(v(r^*)) - (v(r^*) - k)(1 - G(v(r^*))) - r$. Accordingly, r will be determined by $v'(r)G(v(r^*)) = 1$, and this condition is clearly satisfied at r^* . The explanation is that the buyer's choice of r affects his return only when he obtains performance. Hence, r^* will be chosen and performance will also be optimal. This point was first made by Coote 1985.

contracts has not been considered. However, as discussed in chapter 13, renegotiation of contracts sometimes may take place and lead to mutually desirable performance if that would not otherwise occur. Here we will assume that renegotiation of contracts in such circumstances will take place costlessly, unless otherwise mentioned, and we will briefly consider the effect of renegotiation on performance, reliance, and risk aversion.

4.2 Performance.⁹⁵ Given the assumption that renegotiation will take place whenever the mutually desirable outcome as to performance would not otherwise occur, it is essentially a tautology that performance will always be mutually desirable. Thus, regardless of the remedy for breach, performance will be mutually optimal and identical to what would be provided for in the mutually desirable completely specified contract. That is, performance will occur if and only if its value exceeds the cost of performance.

Example 8. In the example that we have been considering, suppose that the contract price is 50 and that the measure of damages for breach is 5. Then the seller would decide to commit breach when production cost is 60 or 200, but when the production cost is 60 there would be an incentive for renegotiation. In particular, if the seller breaches, his losses are the damages of 5, whereas if he performs he spends 60 and receives 50, so loses 10. Hence, if the buyer increases the price he pays sufficiently above 50, say to 65, the seller will agree to perform; if the price is 65, the seller will make a profit of 5. The buyer himself will be better off paying 65 and obtaining performance worth 100 than receiving damages of 5. Such mutually desirable renegotiation for performance is always possible when the cost of performance is less than its value and the seller would otherwise be led not to perform.⁹⁶

If the measure of damages is the expectation measure, namely 50, then we know that the seller will decide to commit breach only when the production cost is 200, which is what is mutually desirable. Hence, there will never be renegotiation of the contract.

If the damage measure is high enough to induce performance when production cost is 200, however, there will be renegotiation, in which the seller pays the buyer in order not to have to perform. Assume for instance that damages for breach are 175. Then if production cost is 200, the seller will lose less, namely 150, if he performs than if he commits breach, so he will perform in the absence of renegotiation. But there is an incentive for the parties to renegotiate. Suppose that the seller pays the buyer 70 in exchange for not performing. The buyer will be better off than if there is performance, in which case he would obtain a net value of 50. The seller will also be better off because he pays 70 rather than losing 150 from performing. Such mutually desirable renegotiation is always possible when the cost of performance exceeds its value and the seller would otherwise be led to perform.⁹⁷//

⁹⁵In this section, I largely follow Shavell 1984a.

⁹⁶To see why algebraically, note that our hypothesis is that the cost of performance c is less than its value v , and that the loss from performance $c - k$ exceeds damages d -- for that is why the seller would commit breach. We want to show that there is an increase t in the price that will lead the seller to perform and also leave the buyer better off. Thus, t must satisfy two conditions: $c - (k + t) < d$ and $v - (k + t) > d$. Now if $t = c - k - d + .5(v - c)$, both conditions will be satisfied, since $v - c > 0$.

⁹⁷In the present situation, our hypothesis is that $v < c$, so that performance is not mutually desirable, and yet that $c - k < d$, so that the seller would perform. We want to show that there exists a payment t that the seller could make to the buyer for not performing that each would prefer to the alternative. Thus, for the seller to be better off, we want $t < c - k$, and for the buyer to be better off, we want $t > v - k$. Let $t = v - k + .5(c - v)$.

Because performance will occur exactly when it is mutually desirable -- when cost is less than its value -- under any remedy for breach, it follows that all remedies for breach will be equivalent ex ante in the eyes of the contracting parties. Specifically, it can be shown that, given any damage measure and price for a contract, a change to any other damage measure will leave both parties exactly as well off if the price is changed appropriately; moreover, no damage measure will be mutually preferable to any other damage measure.⁹⁸ In sum, because damage measures serve no role in inducing mutually desirable performance when renegotiation is a costless and smooth process, damage measures offer no advantage to the contracting parties if performance is the only relevant factor in a contractual situation.

4.3 Reliance. Renegotiation will generally affect the choice of the level of reliance by a buyer, for the terms of the renegotiation will tend to be influenced by the value of performance, which reliance affects. The particular way in which renegotiation affects reliance will depend on the measure of damages, so that the remedy for breach does affect the degree of reliance.

To illustrate, consider initially the case if damages for breach are zero -- so that the seller will want to commit breach whenever the production cost exceeds the contract price that he would receive if he performed. Then the buyer will have to pay the seller to perform when production cost exceeds the price, and the payment he will make will have an effect on his reliance.

Example 9. Suppose that, given some contemplated level of reliance by the buyer, the value of performance would be 100, that the contract price is 50, and that there are no damages for breach. Then the seller will commit breach whenever production cost exceeds 50, so he will want to commit breach when production cost is 60 in the absence of renegotiation. There will, however, be renegotiation in which the buyer pays the seller to perform. If the seller is paid at least 60 he will perform when production cost is 60. Suppose that renegotiation leads to a price of 80 -- the nature of renegotiation is that the two parties split equally the surplus (of $100 - 60 = 40$) that they would gain from renegotiation (how they split it will not matter to the point to be made, as long as the seller obtains some fraction of it).

Now suppose that the buyer thinks about investing in another, higher, level of reliance, which would lead to a value of performance of 120 instead of 100. Then, if production cost is 60, there will be renegotiation and the buyer will pay a price of 90 -- for this will split the higher surplus of 60 (that is, $120 - 60$). Thus, the buyer's net value in the event of renegotiation rises from $100 - 80 = 20$ to $120 - 90 = 30$, which is to say, by 10, when the value of performance goes up by 20, from 100 to 120. The reason is that in the process of renegotiation the seller is extracting half of the reliance-created increase in the value of performance.

⁹⁸The argument that parties will be just as well off under a change in damage measures if the price is adjusted can be outlined as follows. From the previous two footnotes, we know that under any damage measure d , renegotiation will, if necessary, always lead to performance if and only if $c < v$. This implies that the sum of parties values under d , denoted $T_d(k)$ (see note 9) will be maximal, M , regardless of d and of the contract price k . Moreover, for any measure d , M can be divided between the parties in any way through the choice of k . Now let d be a damage measure and k , the price, so that the buyer obtains $B_d(k)$ and the seller $M - B_d(k)$. Change the damage measure to any other measure d' and set the price equal to k' such that $B_{d'}(k') = B_d(k)$. Because $T_{d'}(k') = M$ for any k' , we know that the seller's value under k' must be $M - B_{d'}(k') = M - B_d(k)$, so he is also just as well off as under d and k .

That both parties cannot be better off is clear. Were both parties better off under a change of damage measures, the sum of values would have to exceed M , but that cannot be because M is the maximal sum of values.

As a consequence of this extraction of the value of reliance, the buyer may decide not to increase his reliance even though that would increase the joint value of the contract. For example, suppose that by increasing reliance expenditures by 15, the value of performance would be increased by 20. This would be mutually desirable, but the buyer would not do it.⁹⁹ As this example shows, when the buyer has to renegotiate and bargain for performance, some of the value of reliance to him is extracted in the process, and anticipating this, he will tend to rely too little.

If, however, the remedy for breach is such that the seller has to renegotiate and pay so that he will not have to perform, the buyer will be led to rely too much. The reason is, in essence, that when he increases his level of reliance and thus the value of performance, he will be able to obtain a higher payment from the seller in renegotiation. Because this motive to rely is one that has to do with increasing the payment made to the buyer rather than with increasing realized value from performance, it leads to excessive reliance.

It should also be stated that, under the expectation measure, the incentive to rely is excessive, just as it was before when renegotiation was not at issue. The reason is that, because there is no renegotiation under the expectation measure, its possibility is essentially irrelevant.

The general conclusion is that, because damage measures provide the backdrop against which renegotiation occurs, the choice of damage measures affects reliance decisions. Which damage measure results in the best level of reliance for the parties is a complicated subject and depends on, among other factors, whether both parties have opportunities to rely and the nature of their reliance.¹⁰⁰

4.4 Risk aversion.¹⁰¹ As was discussed above in sections 1.4 and 2.7, when parties are risk averse, the allocation of risk enters into the determination of the mutually beneficial contractual arrangement, and renegotiation affects this point to some degree. In general, renegotiation would seem to reduce risk-bearing that is due to otherwise inappropriate

⁹⁹Because performance will occur when production cost is 20 or 60, and thus with probability 80 percent, the expected value of the increase in the value of performance is $80\% \times 20 = 16$, which exceeds the cost of 15, so that the increase in reliance expenditures does increase the expected net value of the contract. The buyer will obtain a benefit of $30\% \times 20 + 50\% \times 10 = 11$, for when he renegotiates his benefit is only 10. Thus, he will not spend 15 to achieve a personal benefit of only 11.

¹⁰⁰There is a large and growing literature in economics on this subject. Rogerson 1984 first analyzed reliance decisions in the presence of damage measures for breach, assuming costless and perfect renegotiation and that only one party relies. Beginning with Hart and Moore 1988, economists have focused on a more general situation in which both the buyer and the seller rely. This literature furthermore usually supposes that none of the variables (costs, values of performance, reliance investments) are verifiable by courts. Thus, a contract can depend only on what is recorded in it, certain subsequent communications between the parties, whether there has been performance, and, if not, who committed breach. Of note are a number of results establishing the existence of contracts that will produce optimal outcomes, that is, in both parties choosing optimal levels of reliance investment (performance will always be optimal). Aghion, Dewatripont, and Rey 1994 and Chung 1991 demonstrate that reliance will be optimal using a contract in which one party is effectively given the right to make a single take-it-or-leave-it offer to the other in renegotiation. Edlin and Reichelstein 1996, Noldeke and Schmidt 1995, and Rogerson 1992 establish closely related optimality results. However, these optimality results fail to hold if reliance by one party affects the return of the other; see Che and Chung 1999 and Che and Hausch 1999.

¹⁰¹The points of this section are discussed in Shavell 1984a.

performance decisions. Notably, the seller may bear a large risk in the absence of renegotiation if damages exceed the expectation measure or if specific performance is the remedy, for if the seller's production costs are high, he will have to incur them in order to perform. The seller's ability to renegotiate the contract in such circumstances will lower the risk for him.

However, renegotiation cannot alter risk in a substantial way; only ex ante provisions in contracts can do that, as was indicated in chapter 13. If the seller bears a large risk due to damages being high or the use of specific performance, this risk is moderated but not eliminated by renegotiation. The only way to substantially alter the risk to the seller is for the seller to adopt a different damage measure, such as expectation, or else to write a more complete contract governing when performance shall occur and allocating risk independently.

4.5 Summary of conclusions about renegotiation and remedies for breach. Although renegotiation leads to performance when that is mutually desirable regardless of the remedy for breach, remedies for breach exert an important influence on reliance and also on the allocation of risk. Thus, even if renegotiation is costless and operates successfully, remedies for breach remain relevant. In fact, of course, renegotiation is a costly process and may not be successful, so that the role of remedies for breach in promoting mutually desirable performance retains significance.

Chapter 16

OTHER TYPES OF CONTRACT

Here I will discuss two additional types of contract: contracts for the transfer of possession of things that already exist, and donative contracts. These two types of contracts are of substantial importance and present us with features different from those of production contracts.¹⁰²

1. Contracts for Transfer of Possession

1.1 Introduction and description. By contracts for transfer of possession, I refer to contracts for delivery or conveyance of things that already exist rather than things that need to be produced. Examples are contracts for transfer of ownership of real estate, paintings and other art objects, used consumer durables, and goods held in inventory.

The type of uncertainty that we will consider in examining contracts for transfer of possession concerns outside-the-contract bids for the contract good that is supposed to be transferred between the contract seller and the contract buyer. For example, if a house is supposed to be transferred, the uncertainty will concern what an outside party might bid for the house before it is supposed to be conveyed to the contract buyer.¹⁰³

We will distinguish for analytical purposes three categories of situations involving bids of outside parties: (a) when outside parties make bids only to the contract seller; (b) when outside parties make bids only to the contract buyer; and (c) when outside parties can make bids to either the contract buyer or the contract seller. Consider, for example, a contract for the sale of a painting by an art dealer to an individual. One would expect outside bids for the painting to be made primarily to the dealer, because of his business enterprise, rather than to the individual (unless he were a well-known collector). If, however, we consider a contract for the sale of a house, then we might expect that outside bids would be made either to the contract seller or to the contract buyer.

Another point should be added. We will sometimes consider the possibility that there would be two transactions, one in which the outside party buys from the contract seller, and then a second in which the contract buyer purchases from the outside party.

1.2 Completely specified contracts. Let us now consider the nature of mutually beneficial completely specified contracts, that is, contracts that state whether the good will be transferred to the contract buyer as a function of the bid made by outside parties. As with production contracts, we will subsequently examine what remedies for breach will lead to outcomes most closely resembling what the parties would have described in a completely

¹⁰²There are, of course, a variety of other types of contracts of significant interest (such as principal and agent contracts), but they will not be treated here.

¹⁰³There cannot be uncertainty over production cost, of course, because the good is presumed already to exist, so that the issues we discussed in the last chapter surrounding production cost uncertainty are not directly relevant. But, as will be seen, many of the points made there have analogues here. Moreover, much of what is said here will have application in the production contract context, as in that context it is quite possible that outside parties would make bids for the goods that are supposed to be produced.

specified contract.

Bids made only to contract seller. Consider initially the situation in which outside parties make bids only to the contract seller, who for concreteness we can think of as a dealer. In this situation, the mutually desirable completely specified contract would call for performance if and only if the buyer's valuation exceeds the outside bid. This is illustrated in the following example.

Example 1. Suppose that the value of the good to the buyer is 100 and that there are three possible outside party bids, as shown below.

| Outside Bid | Probability |
|-------------|-------------|
| 50 | 50% |
| 90 | 10% |
| 150 | 40% |

Suppose that the parties contemplate a completely specified contract -- one that states whether there shall be performance in each of the three contingencies. Let us show that if the contract does not specify performance when and only when the bids are 50 and 90 -- that is, less than the buyer's valuation of 100 -- then both parties would want to change the contract.

Suppose initially that the parties discuss a contract under which there is to be performance under all three contingencies and a price of, say, 80, to be paid at performance.¹⁰⁴ The value of the contract to the buyer will be $100 - 80 = 20$ since he is assured performance. The value to the seller will be 80. Now suppose that the contract is altered, so that the seller has to perform only when the outside bid is 50 or 90. This is an advantage to the seller, for he will then be able to sell for a high price of 150, rather than at 80, with a probability of 40 percent; indeed, the advantage of receiving an extra 70 is worth $40\% \times 70 = 28$ to the seller. Thus, the seller should be willing to accept a lower price in exchange for being allowed not to perform when the outside bid is high. In particular, suppose that the price is lowered from 80 to 55. Then the seller's expected value will be $60\% \times 55 + 40\% \times 150 = 93$, so he will be better off under the altered contract (93 exceeds 80). Likewise, the buyer's expected value will be $60\% \times (100 - 55) = 27$, so he too will be better off (27 exceeds 20).

Next suppose that the parties initially discuss a contract calling for performance only when the outside bid is 50 and a price of, say, 60. Under this contract, the buyer's expected value will be $50\% \times (100 - 60) = 20$ and the seller's expected value will be $50\% \times 60 + 10\% \times 90 + 40\% \times 150 = 99$. Now consider an altered contract calling for performance whenever the outside bid is 50 or 90. This contract will be better for the buyer since he will receive performance more often, so he should be willing to pay a higher price. Let us suppose that he pays a price of 66. Then the buyer's expected value will be $60\% \times (100 - 66) = 20.4$, so he will be better off. The seller's expected value will be $60\% \times 66 + 40\% \times 150 = 99.6$, so he will also be better off.

Thus, the mutually optimal contract calls for performance whenever the outside bid is 50

¹⁰⁴As in the last chapter, I will assume for concreteness that payment is to be made at the time of performance. Again, the main conclusions would not be altered if payment were to be made at the outset.

or 90; this contract is preferred by both parties either to a contract that calls for performance all the time, or for performance less often than whenever the outside bid is 50 or 90.//

The underlying reasons for the result just illustrated are twofold. On one hand, if the contract calls for a sale to the contract buyer even though the outside bid is 150, the parties would be wasting the joint opportunity to sell at a higher price than the value that the buyer places on the good. On the other hand, the parties obviously do not want to sell the good to the outsider if his bid is less than the buyer's valuation. In other words, the joint value or pie that the contracting parties have to split will be maximized provided that performance occurs when but only when the outside bid is less than the buyer's valuation.

It should be noted that this conclusion does not change if we consider the possibility that, were the good sold to the outside party when the bid is less than the contract buyer's value, the buyer could then repurchase it from him. Suppose for instance that the contract price is 60 and that, when the outsider bids 90, the seller sells to him, but then the buyer repurchases it from the outsider at a price of, say, 95.¹⁰⁵ This means that, although the buyer does obtain the good, he pays more for it than the seller receives -- he pays 95 whereas the seller receives 90. Thus, the two parties to the contract jointly surrender 5 to the outside party in the process of effecting a transfer of the good to the contract buyer -- there is in effect a *leakage of funds* to the outsider that lowers joint value to the contracting parties. For that reason, it is readily demonstrated, along the lines of the logic of the above example, that the buyer and seller would not make a contract under which the outsider would purchase the good for 90; they would continue to find the contract in which there is performance whenever the outside bid is lower than the buyer's valuation mutually desirable. Moreover, this conclusion is reinforced by the fact that an additional transaction, involving extra cost, would be involved were the contract buyer to purchase from an outsider.

Bids made only to contract buyer. Next consider the situation in which outside parties make bids only to the contract buyer, who we can now think of as a dealer. In this situation, the mutually desirable completely specified contract would call for performance under all circumstances. Let me illustrate with the previous example under the present assumption about outside bids.

Example 2. In the previous example, but where the outside bids are now made only to the contract buyer, suppose the parties contemplate a contract in which the good is not always sold to the buyer. For example, suppose that they contemplate a contract with a price of 60 and in which the good is sold to the buyer when the outside bid is 50 or 90, but not when it is 150. The value of this contract to the buyer is $60\% \text{ H}(100 - 60) = 24$. The value to the seller is $60\% \text{ H} 60 + 40\% \text{ H} 10 = 40$, where 10, say, is the value of the good to the seller for his own use.¹⁰⁶ (The seller cannot sell the good to the outside party when he does not convey it to the buyer, for the assumption is that outside bids are made only to the buyer.)

¹⁰⁵That is, we are as usual presuming that the surplus from a transaction is divided between the parties to it. Because the hypothesis is that the outsider bid 90 for the good, he must attach a value of at least 90 to it, so that any transaction that occurs between the outsider and the buyer must be at a price over 90 and not more than 100. (If it so happens that the outsider's valuation exceeds 100, there would be no transaction with the contract buyer, which only strengthens the argument made in the text.)

¹⁰⁶The argument to be given can be shown not to depend on precisely what this value is, only that it is lower than the buyer's valuation -- something which must be true given that the seller is willing to sell to the buyer.

Now consider as an alternative a contract in which the good is also sold to the buyer when the bid is 150. The buyer will be willing to pay more for such a contract, because he will then be able to, and will, sell the good to the outsider for 150, rather than keep it and enjoy a value of only 100. Suppose then that this contract is made and that the price is raised to 70. The buyer's value will be $60\% H(100 - 70) + 40\% H(150 - 70) = 50$, so he is better off, and the seller obtains 70, so he too is better off.//

The reason that the parties want performance all the time is that this means that the two guarantee that they will obtain the highest value from the good: The contract buyer will keep the good for himself if his valuation exceeds the outside bid, and the buyer will sell to the outsider when the outside bid exceeds his valuation. Because the hypothesis is that it is the contract buyer who faces the outside bid, the good must be transferred to the buyer for the good to be sold to the outsider when the bid is higher than the buyer's valuation. (Note that the contract seller too wants this to happen because that enables the buyer to pay him more in the contract price.) The conclusion here can be understood as again reflecting the point that the parties want a contract that leads to the maximization of the joint value that they have to share.

Bids made either to contract buyer or to contract seller. In this situation, the conclusion is that the mutually desirable completely specified contract is not unique: The contract is mutually desirable as long as it calls for performance at least whenever the buyer's valuation exceeds the outside bid -- and in particular if it calls for performance under all circumstances. The logic behind this conclusion can be understood by reference to the arguments of the previous two cases.

Let me first explain why the completely specified contract would require performance when the buyer's valuation exceeds the outside bid, that is, when the outside bid is 50 or 90 in the example. If there were no performance in such a case and the seller sold the good to the outsider, there would be a decrease in joint value, for the buyer values the good at 100. Moreover, even if the buyer can repurchase from the outsider, the joint value will not be maximized because of leakage of funds to the outsider, as explained above.

Let me now explain why does it not matter what the contract specifies when the outside bid exceeds the buyer's valuation, that is, when the bid is 150 in the example. In that case, whoever has the good, the contract buyer or the contract seller, will sell to the outsider for 150, for the assumption is that either party has the opportunity to do so. Thus the opportunity to sell to the outsider when his value exceeds the buyer's will not be lost, and it does not matter to the joint maximization of value who has the opportunity to sell to the outsider. (Who has the opportunity will affect the contract price, of course; if the buyer obtains performance all the time, he will tend to pay more to the seller than if the seller does not have to perform when the outside bid is 150 and thus can sell for the high bid price.)

1.3 Comments. (a) *Mutual desirability and social desirability.* One of the points made about completely specified contracts is that it is mutually desirable for the contract buyer and seller to ensure that there be performance whenever the buyer's valuation exceeds the outside bid -- even if the buyer could, in a second transaction, repurchase the good from the outsider. The reason that the two contracting parties do not want the buyer to obtain the good through repurchase from the outsider is that that would result in a leakage of funds to the outsider. But that is not a *social* reason for there to be performance; it is only a reason applying to the two contracting parties who want to preserve value for themselves. However, the additional transaction costs associated with resale to the contract buyer would constitute a social reason for

there to be performance whenever the buyer's valuation exceeds the outside bid.

(b) *Risk aversion.* The influence of risk aversion on the mutually desirable completely specified contract was not discussed above. Risk aversion does not affect the conclusions about when performance would be mutually desirable. As was the case with production contracts, risk aversion affects only the allocation of risk that the parties would want to effect through their payment arrangements. For example, suppose that outside bids are made only to the contract seller and that he is risk averse and the buyer is risk neutral. Being risk averse, the seller would prefer to have a sum certain than a probability of selling for a high outside bid, such as 150. Thus, under a mutually desirable risk-sharing agreement, the seller would receive a guaranteed price from the contract buyer and give the proceeds from any sale to a high outside bidder to the contract buyer.

Note, however, that the risk at issue in the context of uncertainty about outside bids is a *beneficial* risk -- the opportunity to sell to a high outside bidder. Thus, risk aversion does not seem to be nearly as important a consideration as in the production contract context, where the risk is detrimental, notably, that the production cost might be high.

1.4 Remedies for breach. Let us now inquire about the mutual desirability of different remedies for breach assuming that the contract is not completely specified, and for simplicity that it merely says that the good is to be conveyed to the buyer. Which remedies will be best for the parties will follow in a fairly straightforward manner from the description of completely specified contracts. Let me again discuss the different types of contractual situation separately in order to maintain clarity, and then take stock of the conclusions reached.

Bids made only to contract seller. In this case, the expectation measure leads to the same outcome as in the mutually desirable completely specified contract, and the expectation measure is the mutually preferred remedy for breach. Let me first illustrate this conclusion.

Example 3. Assume in the example that the contract price is 70. Then under the expectation measure, the seller has to pay the buyer 30 in damages, for that would have been his net benefit from performance, namely, $100 - 30$. Under this measure, the seller obviously will not commit breach when the outside bid is 50, for that is less than the price from the buyer, and he would not breach when the outside bid is 90, for this would give him only 20 more than the buyer would pay but he would have to pay damages of 30. The seller will commit breach only when the outside bid is 150, which is the outcome under the mutually desirable complete contract, as discussed in Example 1.

That the expectation measure is mutually preferred to other measures can be shown by the usual kind of reasoning. Suppose, for instance, that the damage measure is only 10 and the contract price 70, so that the seller would breach when the outside bid is 90. In this case, the value of the contract to the buyer is $50\% H(100 - 70) + 50\% H 10 = 20$ and the value to the seller is $50\% H 70 + 10\% H(90 - 10) + 40\% H(150 - 10) = 99$. If the parties change to the expectation measure and the price is raised to 79.5, both will be better off: The value to the buyer will be 20.5, and that to the seller will be $60\% H 79.5 + 40\% H(150 - 20.5) = 99.5$.¹⁰⁷ (It is also

¹⁰⁷The example can be modified to take into account the possibility of repurchase by the contract buyer from the outsider. If damages are 10, the seller breaches and sells to the outsider when his bid is 90, and the contract buyer then buys back from him at a price of 95, again both parties can be made better off. In particular, the value of the contract to the buyer would be $50\% H(100 - 70) + 10\% H(10 + (100 - 95)) + 40\% H 10 = 20.5$, and the value to the seller would be as before, 99. If the parties change to the expectation measure and the price is raised to 79.4, both will be better off: the value to the buyer will be 20.6, and that to the seller will be $60\% H 79.4 + 40\% H(150 - 20.6) = 99.4$.

straightforward to demonstrate that if damages are high enough to result in performance all the time, the parties will both prefer to use the expectation measure.)//

The explanation for the result just illustrated is along lines that are familiar from chapter 15 on production contracts. Under the expectation measure, the seller is induced to perform whenever the outside bid is less than the buyer's valuation (just as the seller was induced to perform whenever the production cost was less than the buyer's valuation in the production contract context). Because this outcome is what would have occurred under the mutually desirable completely specified contract, and maximizes the joint value the parties have to share, the expectation measure turns out to be preferred by the parties to other, lesser damage measures, or to higher ones, or to specific performance.

Bids made only to contract buyer. In this case, specific performance leads to the same outcome as in the mutually desirable completely specified contract and is the mutually preferred remedy for breach. The explanation is immediate from what was said about the completely specified contract. Because it is the contract buyer who has the opportunity to sell to the outside bidder, it is best for the parties to give that opportunity to him, which is what specific performance accomplishes.

Bids made either to contract buyer or to contract seller. Here, the expectation measure or higher measures -- as well as specific performance -- lead to the same outcomes as in mutually desirable completely specified contracts and are equivalent, mutually preferred, remedies for breach. Because the expectation measure or higher measures result in performance at least as often as whenever the outside bid is less than the buyer's valuation, they result in the behavior called for in completely specified contracts. Therefore, these remedies for breach lead to maximized joint value and are equivalent for the parties.¹⁰⁸ However, a damage measure that is less than the expectation measure would be undesirable for the parties, because it might lead to an outcome in which the seller sells to an outsider when his bid is less than the buyer's valuation, resulting in a loss in total value to the two contracting parties. For this reason, such a damage measure is inferior to the others.

The last point, that a damage measure less than expectation measure is not desirable for the parties, has an important implication: *specific performance is mutually preferred to the expectation measure if courts might underestimate the value of performance.* The reason is that, if the expectation were to be underestimated, then the seller might breach when the outside bid is less than the buyer's true valuation. In the example, if the court thought the value of the good to the buyer were 80 (instead of 100), then it would compute expectation damages to be only 10 (instead of 30) when the contract price is 70, so that the seller would commit breach when the outside bid is 90 (instead of performing). Such outcomes reduce joint value to the contracting parties.¹⁰⁹ They can be avoided for sure under specific performance. To put the point differently, under specific performance, it is always the buyer who decides whether to sell to an outside

¹⁰⁸By equivalent is meant that for any contract of one type, there is a contract of the other type that is equivalent for both parties (given a suitable change in the contract price).

¹⁰⁹Note that the possibility of overestimation of the value of performance by courts does not affect the argument under discussion. If overestimation occurs, it only means that the contract buyer will receive the good and then sell to the outside bidder, rather than the seller selling to the outside bidder when the outside bid exceeds the buyer's value but is less than the court's overestimate of the buyer's value. This does not alter the joint value of the contract to the contracting parties.

bidder, and thus he will do so naturally if and only if the outside bid exceeds his valuation. Under the expectation measure, in contrast, the seller decides whether to sell to an outside bidder, and will do so on the basis of a possibly too-low court estimate of the buyer's valuation.

1.5 Comment. It has been assumed throughout this section that the magnitudes of outside bids are given, but it may be that these bids are influenced by the remedy for breach. Notably, other things being equal, one might expect the price that the seller could obtain through bargaining with an outsider to be higher the more the seller would have to pay in damages for breach. This gives the parties to the contract an incentive to employ higher damages for breach than they otherwise would.¹¹⁰

1.6 Interpretation of the discussion of this section: the appeal of specific performance. An interpretation of our discussion of remedies for breach is that the parties will find specific performance appealing. The virtue of specific performance is that it guarantees that the good will never be sold to outsider bidders at a price less than the buyer's valuation. As has been emphasized, to the extent that that occurs, it reduces the joint value of the contract to the contracting parties (even if the buyer repurchases from the outsider). Moreover, such value-reducing sales to outside bidders are a danger under the expectation measure if there is a risk of underestimation of the value of performance. Therefore, if the buyer and seller have equal access to outside bids, or if the buyer alone has access to outside bids, the parties to the contract will prefer specific performance to the expectation measure, and the more so the greater the likelihood of underestimation of the true value to the buyer. Only if the seller has better access to outside bids would the parties prefer the expectation measure.

1.7 Contract law. In Anglo-American law, the usual remedy for breach of contract is, as was stated earlier, the expectation measure. However, specific performance is sometimes employed, mainly for certain types of contracts for transfer possession: for land contracts and for things whose value is idiosyncratic and hard to assess, such as paintings.¹¹¹ In French law, specific performance is the standard remedy for all contracts for transfer of things, while damages are the remedy for breach of contracts to make things.¹¹² In German law, the remedy for all contracts is nominally specific performance.¹¹³ Thus, in important legal systems, specific performance plays a major role as a remedy for breach of contracts for transfer of possession, in contrast to the dominating role of damage measures for contracts to produce things. The mutual appeal of specific performance for contracts for transfer of possession, as I have explained above, may help to explain the pattern of its use in legal systems.

1.8 Contrast with production contracts. Let me contrast the theory presented here with that concerning production contracts. In the case of production contracts, specific performance is undesirable because of its effect on performance: It leads to inefficient production when production cost exceeds the value of performance. Here, by contrast, specific performance has a

¹¹⁰This point was originally made by Diamond and Maskin 1979 and has been developed by, among others, Chung 1992.

¹¹¹See, for example, Calamari and Perillo 1998, 614-17, and E. Farnsworth 1999, 773-77.

¹¹²See, for example, Treitel 1988, 55-63, and Zweigert and Kötz 1998, 475-79.

¹¹³See, for example, Treitel 1988, 51-55, and Zweigert and Kötz 1998, 472-74.

beneficial effect on performance: Its use means that the jointly undesirable outcome of sale to outsiders when bids are less than the buyer's valuation will be avoided (presuming that outside bids are available to the contract buyer). Also, in the case of production contracts, the use of specific performance imposes great detrimental risk on sellers. But here specific performance imposes no risk of loss, only a transfer of the opportunity to avail oneself of the benefit to sell to a high outside bidder. Thus, the mutual desirability of the two types of remedy is quite different in the two contexts.¹¹⁴

Note on the literature. Most writing on the use of specific performance has focused on the inability of courts to estimate the value of performance and has justified specific performance on that ground. Thus, the notion has been that the purpose of remedies is to guarantee that the victim be made whole, and the insufficiency of expectation damages has therefore led to the conclusion that specific performance would be needed to accomplish this purpose.¹¹⁵ This view, note, makes no reference to the mutuality of interests of the contracting parties, to the fact that the potential victim of a breach may not be risk averse, nor to the fact that, in contracts for production, the expectation measure may also be difficult to estimate yet is employed. The explanation given above, concerning the mutual interests of the parties to the contract and avoiding sale to outside bidders at a price less than the buyer's valuation, is developed in Shavell (1984a).¹¹⁶ Kronman (1978b) also considers an explanation for use of specific performance based on its mutual desirability. He suggests that sellers tend to believe that high outside bids are unlikely, whereas buyers believe that high outside bids are likely; hence, buyers will value specific performance substantially and sellers will view it as costing them little. However, the basis for Kronman's assumption of asymmetric beliefs between sellers and buyers is unclear.¹¹⁷

2. Donative Contracts

2.1 Definition and questions to be addressed. By a donative contract is meant an agreement under which one person, the *donor*, binds himself to give a gift to another party, the *donee*, but in which the donee does not agree to do anything in return. For example, a person may arrange to give a gift to a relative or to a charitable or educational institution without such a donee promising anything in return. We will be concerned with three questions in analyzing donative contracts. The first is why gift-giving occurs at all, and we will focus on altruism as the motivation but will consider others at the end of the section. The second question is why donors who want to give gifts would wish to defer them rather than give them immediately. The third question is why donors who wish to defer gifts would want to obligate themselves contractually, that is, why should they not merely wait to give gifts without obligating themselves to do so?

2.2 Altruism as the motivation for gift-giving. By altruism is meant that the donor cares about the well-being of the donee. For example, the donor's utility might be assumed to

¹¹⁴However, it should be noted that the possibility of outside bids is also present in the production context; a good that has to be produced is also one for which an outsider might make a bid. Therefore, what was said here about outside bids also has implications for production contracts. The main implication is that, after a good is produced but before it is delivered, specific performance would seem to have some advantages.

¹¹⁵See, for example, E. Farnsworth 1999, 773-77, and references cited therein.

¹¹⁶This analysis is amplified in Bishop 1985.

¹¹⁷See also Schwartz 1979 and Ulen 1984 on economic analysis of specific performance.

equal the sum of an altruistic component -- the donee's utility multiplied by some factor -- plus the utility the donor derives directly from his own consumption. Such a donor will want to give a gift if the altruistic utility he derives from doing so exceeds the loss in utility he would have obtained from using his gift for his own consumption.

2.3 Why altruistic donors may defer giving gifts. One reason why a donor may wait before making a gift is simply that his assets may not be liquid. Another reason for deferral is that the donor may earn a higher rate of return on funds than the donee over the period before the donee actually needs to use the funds. A third reason for deferral concerns uncertainty. The donor may be uncertain about his financial situation; he may suffer a financial reversal or find that his needs for funds are unexpectedly high. If so, he might turn out not to want to give a gift. Likewise, the donor may be uncertain about the donee. The donee's needs or his financial position may change, or he may reveal characteristics that alter the donor's altruistic feelings about him. By deferring their gifts, donors preserve their options and need not give gifts, or can limit their size, according to the resolution of uncertain factors.¹¹⁸

2.4 Why altruistic donors may -- or may not -- want to make contracts to give gifts. Why would a donor who wishes to defer a gift want to bind himself contractually to give it? Why not merely wait and give the gift? What benefit would the altruistic donor gain by making a contract to do so?

An apparent answer concerns reliance by the donee. If a donee relies on receiving a gift, the value of the gift when given will often be enhanced, and thus will inure to the benefit of the altruistic donor. If a nephew expects to receive a gift financing his college education, he may study harder now and thereby gain admission to a better college or otherwise increase the value of the gift; if a symphony orchestra relies on receiving a future gift by hiring talented new members who are now available to it but will not be later, the value of the gift will be greater than otherwise.¹¹⁹ Thus, if a contract is necessary to induce reliance by the donee, the donor might want to make a contract.

But a contract may not be needed in order to induce donee reliance. Consider the situation of a donor who does not make a contract to give a gift but announces his intentions to do so and whose financial situation and altruistic feelings are known to the donee. The donee may well then find it rational to rely, for he will know that it will probably be rational for the donor to follow through with the gift. If the nephew knows his uncle's altruistic feelings, the nephew knows that it will likely be rational for his uncle to give him the gift; thus the nephew may well have an incentive to study for college. Indeed, it can even be shown that, in the absence of a contract, a donee who understands the altruistic tastes of the donor will have an incentive to rely excessively, in order to induce an even larger gift from the donor.¹²⁰

¹¹⁸The reasons given in this paragraph also help to explain why donors whose motivations are different from altruism may defer gifts.

¹¹⁹It may be clarifying to express the underlying idea of this paragraph formally; for details, see Shavell 1991a. Suppose that $v(x,r)$ is the value to the donee of a gift of magnitude x , given that donee has engaged in the level of reliance r , so that the donee's net utility is $v(x,r) - r$. Let the donor's utility be given by $u(w - x) + \alpha [v(x,r) - r]$, where w is the donor's wealth, u is his utility from his own use of his wealth, and $\alpha > 0$ measures the strength of his altruistically-derived benefit from the donee's net utility. Thus, it is apparent that if raising r increases the donee's net utility $v(x,r) - r$, it will also raise the donor's utility.

¹²⁰The reason for this is, in essence, that, on one hand, the donee knows that the higher his level of reliance, Chapter 16 – Page 9

It follows that for a donor to want to make a contract to make a gift in order to induce reliance, the situation must be that the donee is unsure of the donor's financial situation or unsure of his motivations, perhaps suspecting him of masquerading as an altruist for some other reason, such as to enhance his reputation. Thus, if a person who is not well known to the nephew (an elderly, distant cousin) announces his intention to finance his college education, the nephew might be much less sure that the gift will be forthcoming and not have the same incentive to study for college that he would if the uncle who is known to be fond of him had made the same statement. As a consequence, the distant cousin might have to make a contract to give the gift in order to induce the nephew to study for college.

The conclusion is that there is not always a need for altruistic donors to make a contract to induce reliance; they will need to do that only if the donees are not aware of the donors' financial situation or their altruistic feelings. Moreover, even if donors would need to make a contract to induce beneficial reliance by donees, there is an important reason why a donor will not want to make a contract to give a gift. As discussed above, uncertain events, such as an unanticipated loss of the donor's wealth, might lead him not to want to make a gift after all, so he would not want to be bound to do so. Hence, donors may or may not want to make contracts to give gifts even if that is necessary to induce the best level of reliance.

2.5 Implications for contract law. (a) The possibility that donors may want to make contracts to give gifts in order to induce reliance by donees implies that the law ought to allow contracts to give gifts. For then donors will more often make gifts, raising both their overall utility and that of donees.

(b) However, the point that altruists' announcements of their intentions to give gifts are often sufficient to induce reliance, coupled with the point that donors may not want to be bound due to various uncertainties, suggests that it may well be desirable for the law to allow donors to state their intentions to give gifts without binding them to do so. Otherwise, donors might decide not to give gifts, or not announce them in advance, thus in either case reducing the well-being of both donors and donees. But it is possible that binding donors who state that they will give gifts would not much reduce gift-giving, and would turn out to raise social welfare due to the benefits furnished donees.

2.6 Non-altruistic motivations for gift-giving. Let us now consider two motivations for gift-giving other than altruism and consider their implications for contracts.

(a) *Paternalism.* A paternalistic donor is a person who desires that the donee engage in a

the greater the gift that the donor will be induced to give due to his altruism; and, on the other hand, the donee does not count as a cost to himself (whereas the donor does) the gift itself. In particular, and using the notation of the previous note, observe the following about the situation in which no contract is made, and the donee has knowledge of the donor's utility function and strength of altruism α : (i) If the donee selects r , then the donor will choose his gift x to maximize $u(w - x) + \alpha [v(x, r) - r]$. Call the donor's choice $x(r)$, which can be shown to be increasing in r (presuming that w is above a threshold). (ii) The donee will select r to maximize $v(x(r), r) - r$, as the donee knows the function $x(r)$. The donee's choice, denoted r^* , must obey $v_x(x(r^*), r^*)x'(r^*) + v_r(x(r^*), r^*) - 1 = 0$. (iii) The donor would like the donee to choose r to maximize the donor's utility, namely, $u(w - x(r)) + \alpha [v(x(r), r) - r]$, which for simplicity we will assume is concave in r . The donor's preferred choice, denoted r^{**} , must satisfy $-u'(w - x(r))x'(r) + \alpha [v_x(x(r), r)x'(r) + v_r(x(r), r) - 1] = 0$. (iv) It is apparent that $r^{**} < r^*$; the donor's preferred level of reliance is less than what the donee will actually choose. Specifically, from the first-order condition in (ii), we know that the left-hand side of the condition in (iii) reduces to $-u'(w - x(r^*))x'(r^*)$ when evaluated at r^* . But $-u'(w - x(r^*))x'(r^*) < 0$. Hence, given the assumption of concavity, r must be lower than r^* to satisfy the condition in (iii); thus, $r^{**} < r^*$.

pattern of consumption or of behavior that in general departs from what the donee might choose for himself. For example, a donor to a medical school may want research carried out in a subject area that the school would not select for itself, or a donor may want a donee to take music lessons even though the donee does not wish this. The paternalistic donor does not care about the utility of the donee per se but about something different.¹²¹ It seems that paternalistic donors often have a stronger reason than do altruistic donors to make contracts to give gifts in order to induce behavior that they want. To induce a donee who does not want to take music lessons to do that, the paternalistic donor will have to make a contract to give the gift and tie the gift to the taking of music lessons: Without the contract, the donee will not take the lessons because, by hypothesis, he does not want to, and with the contract, he will do so (presuming the value of the gift exceeds the disutility to him of the music lessons). In contrast, the nephew of the altruistic uncle does not need a contract to induce him to study; his expectation of a gift combined with his desire to go to a good college and further himself will lead to his doing so.

(b) *Enjoyment of appreciation and reputational enhancement.* Another motive for giving gifts is securing appreciation from others or enjoying an enhanced reputation. When a person gives a gift to a relative, for example, an important reason may be to feel the gratitude of the recipient and of other family members; when a donor makes a gift to a university, it may be that what he wants is honor from the university community or from the general public. If, however, the donor has to defer the gift (perhaps because of liquidity problems), he faces a difficulty: He would like to benefit from the appreciation of others during the period before he actually makes the gift, but in the absence of making a contract to give the gift, the donee may not believe that he will receive the gift and therefore not express appreciation for that. For example, someone who states that he plans to give a large sum to endow a university professorship in ten years might not be believed by the university and thus not be honored during the ten year period. Such donors may find it advantageous to make contracts to give gifts in the future so that they can enjoy appreciation before they complete their gifts.¹²²

2.7 The law. In the United States, promises to give gifts are not legally enforceable in many jurisdictions (those that have abolished the seal). However, promises to give gifts that are reasonably relied upon by donees are usually enforceable, and promises to give gifts to charitable institutions are generally enforceable.¹²³ These exceptions suggest that the law does provide donors the opportunity to make binding gifts when they would want to. Note especially that, as explained, the principal reason donors would want to obligate themselves is to induce reliance, and when reliance is induced, they do become bound (presuming the reliance they want is reasonable). At the same time, it is unlikely that a person would become bound against his

¹²¹In some cases what is described as paternalistic, that is, not coincident with the maximization of the donee's utility, might be interpreted to be due to the donee's failure to understand what is in his true interests. For our purposes, however, it will not matter whether this is the case.

¹²²This argument for why donors may want to contract to make gifts must be qualified to the degree that a person who made an unenforceable promise to give a gift would be induced to carry through with it due to a desire to maintain his reputation. But this factor is limited by, among other elements, the possibility that a donor could claim that changes in circumstance forced him not to make the gift. Also, many gifts are made at or near the end of a person's life, and at that point, the importance of the reputational element may be attenuated.

¹²³On the enforceability of donative promises, see, for example, Calamari and Perillo, 1998, 166-71, 175-80, 247-67, Eisenberg 1979, and E. Farnsworth 1999, 50-52, 85-87, 91-101 .

will, for he could state his intent but not promise to give a gift, and thus not be bound even if there were reliance. In France, Germany, and most civil law jurisdictions, a donor obligates himself if and only if he makes a sufficiently formal written promise, typically before a notary; so the donor has effective control over the enforceability of donative promises.¹²⁴

Note on the literature. Posner (1977) and Goetz and Scott (1980) analyze donative contracts from an economic orientation, the latter emphasizing that such contracts may induce donee reliance, to the donor's benefit. Shavell (1991a) asks why gifts are often deferred and for deferred gifts, compares donative contracts with mere statements of intention to give gifts as ways of inducing donee reliance.¹²⁵

¹²⁴See, for example, Zweigert and Kötz, 1998, 389-90.

¹²⁵Other economically-oriented analysis of donative contracts includes Posner 1997; see also Kull 1998.

References

- Aghion, Philippe, and Benjamin Hermalin. 1990. Legal Restrictions on Private Contracts Can Enhance Efficiency. *Journal of Law, Economics, and Organization* 6:381-409.
- Aghion, Philippe, Mathias Dewatripont, and Patrick Rey. 1994. Renegotiation Design with Unverifiable Information. *Econometrica* 62:257-82.
- Ayres, Ian, and Robert Gertner. 1989. Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules. *Yale Law Journal* 99: 87-130.
- Barton, John H. 1972. The Economic Basis of Damages for Breach of Contract. *Journal of Legal Studies* 1:277-304.
- Bebchuk, Lucian A., and Omri Ben-Shahar. 2001. Precontractual Reliance. *Journal of Legal Studies* 30:423-57.
- Bebchuk, Lucian A., and Ivan P. L. Png. 1999. Damage Measures for Inadvertant Breach of Contract. *International Review of Law and Economics* 19:319-31.
- Bebchuk, Lucian A., and Steven Shavell. 1991. Information and the Scope of Liability for Breach of Contract: The rule of *Hadley v. Baxendale*. *Journal of Law, Economics, and Organization* 7:284-312.
- Bernstein, Lisa. 1992. Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry. *Journal of Legal Studies* 21:115-57.
- 1998. Private Commercial Law. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 3:108-14. London: Macmillan.
- Birmingham, Robert L. 1970. Breach of Contract, Damage Measures, and Economic Efficiency. *Rutgers Law Review* 24:273-92.
- Bishop, William. 1985. The Choice of Remedy for Breach of Contract. *Journal of Legal Studies* 14:299-320.
- Calamari, John D., and Joseph M. Perillo. 1998. *The Law of Contracts*. Fourth edition. St. Paul, Minn.: West Group.
- Charny, David. 1990. Nonlegal Sanctions in Commercial Relationships. *Harvard Law Review* 104:373-467.
- Che, Yeon-Koo, and Tai-Yeong Chung. 1999. Contract Damages and Cooperative Investments. *Rand Journal of Economics* 30:84-105.
- Che, Yeon-Koo, and Donald B. Hausch. 1999. Cooperative Investments and the Value of Contracting. *American Economic Review* 89:125-47.
- Chung, Tai-Yeong. 1991. Incomplete Contracts, Specific Investments, and Risk Sharing. *Review*

- of Economic Studies* 58:1031-42.
- 1992. On the Social Optimality of Liquidated Damage Clauses: An Economic Analysis. *Journal of Law, Economics, and Organization* 8:280-305.
- Clarkson, Kenneth W., Roger Leroy Miller, and Timothy J. Muris. 1978. Liquidated Damages v. Penalties: Sense or Nonsense? *Wisconsin Law Review* 1978:351-90.
- Cohen, George M. 2000. Implied Terms and Interpretation in Contract Law. In *Encyclopedia of Law and Economics*, edited by Boudewijn Bouckaert and Gerrit De Geest, 3:78-99. Cheltenham: Edward Elgar.
- Cooter, Robert D. 1985. Unity in Tort, Contract, and Property: The Model of Precaution. *California Law Review* 73:1-51.
- Craswell, Richard. 1988. Contract Remedies, Renegotiation, and the Theory of Efficient Breach. *Southern California Law Review* 61:629-70.
- 1996. Offer, Acceptance, and Efficient Reliance. *Stanford Law Review* 48:481-553.
- De Geest, Gerrit, and Filip Wuyts. 2000. Penalty Clauses and Liquidated Damages. In *Encyclopedia of Law and Economics*, edited by Boudewijn Bouckaert and Gerrit De Geest, 3:141-61. Cheltenham: Edward Elgar.
- Diamond, Peter A., and Eric Maskin. 1979. An Equilibrium Analysis of Search and Breach of Contract, I: Steady States. *Bell Journal of Economics* 10:282-316.
- Edlin, Aaron S. 1998. Breach Remedies. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 1:174-79. London: Macmillan.
- Edlin, Aaron S., and Stefan Reichelstein. 1996. Holdups, Standard Breach Remedies, and Optimal Investment. *American Economic Review* 86:478-501.
- Eisenberg, Melvin A. 1979. Donative Promises. *University of Chicago Law Review* 47:1-33.
- Farnsworth, E. Allan. 1999. *Contracts*. Third edition. New York: Aspen Law and Business.
- Fishman, Michael J., and Kathleen M. Hagerty. 1990. The Optimal Amount of Discretion to Allow in Disclosure. *Quarterly Journal of Economics* 105:427-44.
- Fried, Charles. 1981. *Contract as Promise: A Theory of Contractual Obligation*. Cambridge, Mass.: Harvard University Press.
- Fudenberg, Drew, and Jean Tirole. 1990. Moral Hazard and Renegotiation in Agency Contracts. *Econometrica* 58:1279-1319.
- Fuller, Lon L., and William R. Perdue, Jr. 1936. The Reliance Interest in Contract Damages, Part 1. *Yale Law Journal* 46:52-96.

- Gertner, Robert H. 1998. Disclosure and Unravelling. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 1:605-08. London: Macmillan.
- Goetz, Charles J., and Robert E. Scott. 1977. Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach. *Columbia Law Review* 77:554-94.
- 1980. Enforcing Promises: An Examination of the Basis of Contract. *Yale Law Journal* 89:1261-1322.
- Goldberg, Stephen B., Frank E. A. Sander, and Nancy H. Rogers. 1999. *Dispute Resolution: Negotiation, Mediation, and Other Processes*. Third edition. Gaithersburg, Md.: Aspen Law and Business.
- Greif, Avner. 1998. Informal Contract Enforcement: Lessons from Medieval Trade. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 2:287-95. London: Macmillan.
- Grossman, Sanford J. 1981. The Informational Role of Warranties and Private Disclosure about Product Quality. *Journal of Law and Economics* 24:461-83.
- Grout, Paul A. 1984. Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach. *Econometrica* 52:449-60.
- Hadfield, Gillian K. 1994. Judicial Competence and the Interpretation of Incomplete Contracts. *Journal of Legal Studies* 23:159-84.
- Hart, Oliver D. 1987. Incomplete Contracts. In *The New Palgrave Dictionary of Economics*, edited by John Eatwell, Murray Milgate, and Peter Newman, 2:752-59. London: Macmillan.
- Hart, Oliver D., and Bengt Holmström. 1987. The Theory of Contracts. In *Advances in Economic Theory*, Fifth World Congress, edited by Truman Bewley, 71-155. New York: Cambridge University Press.
- Hart, Oliver D., and John Moore. 1988. Incomplete Contracts and Renegotiation. *Econometrica* 56:755-85.
- Johnston, Jason S. 1990. Strategic Bargaining and the Economic Theory of Contract Default Rules. *Yale Law Journal* 100:615-64.
- Jolls, Christine. 1997. Contracts as Bilateral Commitments: A New Perspective on Contract Modification. *Journal of Legal Studies* 26:203-37.
- Joskow, Paul L. 1977. Commercial Impossibility, the Uranium Market and the Westinghouse Case. *Journal of Legal Studies* 6:119-76.
- Kant, Immanuel. [1785] 1998. *Groundwork of the Metaphysics of Morals*. Translated and edited by Mary Gregor. Cambridge: Cambridge University Press.

- Kaplow, Louis. 2002a. Economic Analysis of Law. In *Handbook of Public Economics*, edited by Alan J. Auerbach and Martin Feldstein, 3:1661-1784. Amsterdam: Elsevier.
- 2002c. On the Superiority of Corrective Taxes to Quantity Regulation. *American Law and Economics Review* 4:1-17.
- Kaplow, Louis and Steven Shavell. 2002a. Economic Analysis of Law. In *Handbook of Public Economics*, edited by Alan J. Auerbach and Martin Feldstein, 3:1661-1784. Amsterdam: Elsevier.
- Katz, Avery. 1990b. The Strategic Structure of Offer and Acceptance: Game Theory and the Law of Contract Formation. *Michigan Law Review* 89:215-95.
- 1990c. Your Terms or Mine? The Duty to Read the Fine the Print in Contracts. *Rand Journal of Economics* 21:518-37.
- 1993. Transaction Costs and the Legal Mechanics of Exchange: When Should Silence in the Face of an Offer Be Construed as Acceptance? *Journal of Law, Economics, & Organization* 9:77-97.
- 1996. When Should an Offer Stick? The Economics of Promissory Estoppel in Preliminary Negotiations. *Yale Law Journal* 105:1249-1309.
- 1998. Contract Formation and Interpretation. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 1:425-32. London: Macmillan.
- Klein, Benjamin, and Keith B. Leffler. 1981. The Role of Market Forces in Assuring Contractual Performance. *Journal of Political Economy* 89:615-41.
- Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian. 1978. Vertical Integration, Appropriable Rents, and the Competitive Contracting Process. *Journal of Law and Economics* 21:297-326.
- Kronman, Anthony T. 1978a. Mistake, Disclosure, Information, and the Law of Contracts. *Journal of Legal Studies* 7:1-34.
- 1978b. Specific Performance. *University of Chicago Law Review* 45:351-82.
- Kull, Andrew. 1998. Gratuitous Promises. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 2:203-07. London: Macmillan.
- Landes, William M., and Richard A. Posner. 1978. Salvors, Finders, Good Samaritans, and Other Rescuers: An Economic Study of Law and Altruism. *Journal of Legal Studies* 7:83-128.
- Mahoney, Paul G. 2000. Contract Remedies: General. In *Encyclopedia of Law and Economics*, edited by Boudewijn Bouckaert and Gerrit De Geest, 3:117-40. Cheltenham: Edward Elgar.

- Mathios, Alan. D. 2000. The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market. *Journal of Law and Economics* 43:651-77.
- Milgrom, Paul R. 1981. Good News and Bad News: Representation Theorems and Applications. *Bell Journal of Economics* 12:380-91.
- Nöldeke, Georg, and Klaus M. Schmidt. 1995. Option Contracts and Renegotiation: A Solution to the Hold-up Problem. *Rand Journal of Economics* 26:163-79.
- Okuno-Fujiwara, Masahiro, Andrew Postlewaite, and Kotaro Suzumura. 1990. Strategic Information Revelation. *Review of Economic Studies* 57:25-47.
- Polinsky, A. Mitchell. 1983. Risk Sharing Through Breach of Contract Remedies. *Journal of Legal Studies* 12:427-44.
- Posner, Eric A. 1997. Altruism, Status, and Trust in the Law of Gifts and Gratuitous Promises. *Wisconsin Law Review* 1997:567-609.
- Posner, Richard A. 1972a. *Economic Analysis of Law*. Boston: Little, Brown.
- 1977. Gratuitous Promises in Economics and Law. *Journal of Legal Studies* 6:411-26.
- 1998. *Economic Analysis of Law*. Fifth edition. New York: Aspen Law & Business.
- Posner, Richard A., and Andrew M. Rosenfield. 1977. Impossibility and Related Doctrines in Contract Law: An Economic Analysis. *Journal of Legal Studies* 6:83-118.
- Pound, Roscoe. 1959. *An Introduction to the Philosophy of Law*. Revised edition. New Haven: Yale University Press.
- Rasmusen, Eric, and Ian Ayres. 1993. Mutual and Unilateral Mistake in Contract Law. *Journal of Legal Studies* 22:309-43.
- Rea, Samuel A., Jr. 1998. Penalty Doctrine in Contract Law. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 3:23-27. London: Macmillan.
- Rogerson, William P. 1984. Efficient Reliance and Damage Measures for Breach of Contract. *Rand Journal of Economics* 15:39-53.
- 1992. Contractual Solutions to the Hold-Up Problem. *Review of Economic Studies* 59:777-94.
- Rose-Ackerman, Susan. 1985. Inalienability and the Theory of Property Rights. *Columbia Law Review* 85:931-69.
- Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.
- Schoenbaum, Thomas J. 2001. *Admiralty and Maritime Law*. Third edition. St. Paul, Minn.: West Group.

- Schwartz, Alan. 1979. The Case for Specific Performance. *Yale Law Journal* 89:271-306.
- 1992. Relational Contracts in the Courts: An Analysis of Incomplete Agreements and Judicial Strategies. *Journal of Legal Studies* 21:271-318.
- 1998. Incomplete Contracts. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 2:277-83. London: Macmillan.
- Searle, John R. 1964. How to Derive “Ought” from “Is.” *Philosophical Review* 73:43-58.
- Shavell, Steven. 1980a. An Analysis of Causation and the Scope of Liability in the Law of Torts. *Journal of Legal Studies* 9:463-516.
- 1980b. Damage Measures for Breach of Contract. *Bell Journal of Economics* 11:466-90.
- 1984a. The Design of Contracts and Remedies for Breach. *Quarterly Journal of Economics* 99:121-48.
- 1991a. An Economic Analysis of Altruism and Deferred Gifts. *Journal of Legal Studies* 20:401-21.
- 1994. Acquisition and Disclosure of Information Prior to Sale. *RAND Journal of Economics* 25:20-36.
- 1998. Contracts. In *The New Palgrave Dictionary of Economics and the Law*, edited by Peter Newman, 1:436-45. London, Macmillan.
- Smith, Janet Kiholm, and Richard L. Smith. 1990. Contract Law, Mutual Mistake, and Incentives to Produce and Disclose Information. *Journal of Legal Studies* 19:467-88.
- Spier, Kathryn E. 1992b. Incomplete Contracts and Signalling. *RAND Journal of Economics* 23:432-43.
- Spier, Kathryn E., and Michael D. Whinston. 1995. On the Efficiency of Privately Stipulated Damages for Breach of Contract: Entry Barriers, Reliance, and Renegotiation. *RAND Journal of Economics* 26:180-202.
- Stole, Lars A. 1992. The Economics of Liquidated Damage Clauses in Contractual Environments with Private Information. *Journal of Law, Economics, and Organization* 8:582-606.
- Sykes, Alan O. 1990. The Doctrine of Commercial Impracticability in a Second-best World. *Journal of Legal Studies* 19:43-94.
- Trebilcock, Michael J. 1993. *The Limits of Freedom of Contract*. Cambridge, Mass.: Harvard University Press.
- Treitel, G. H. 1988. *Remedies for Breach of Contract: A Comparative Account*. Oxford: Clarendon Press.

- Triantis, George C. 2000. Unforeseen Contingencies. Risk Allocation in Contracts. In *Encyclopedia of Law and Economics*, edited by Boudewijn Bouckaert and Gerrit De Geest, 3:100-16. Cheltenham: Edward Elgar.
- Ulen, Thomas S. 1984. The Efficiency of Specific Performance: Toward a Unified Theory of Contract Remedies. *Michigan Law Review* 83:341-403.
- Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.
- Wils, Wouter P. J. 1993. Who Should Bear the Costs of Failed Negotiations? A Functional Inquiry into Precontractual Liability. *Journal des Economistes et des Etudes Humaines* 4:93-134.
- Wittman, Donald. 1981. Optimal Pricing of Sequential Inputs: Last Clear Chance, Mitigation of Damages, and Related Doctrines in the Law. *Journal of Legal Studies* 10:65-91.
- Zweigert, Konrad, and Hein Kötz. 1998. *An Introduction to Comparative Law*. Third edition. Oxford: Clarendon Press.