

---

CHAPTER ON THE LAW & ECONOMICS OF CONTRACTS

BENJAMIN E. HERMALIN, AVERY W. KATZ, AND RICHARD CRASWELL

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The economic motive for contracts . . . . .	2
1.2	Law & economics issues in contracting . . . . .	7
1.3	What this chapter is not . . . . .	11
1.4	Organization of the chapter . . . . .	12
<b>2</b>	<b>Freedom of Contract</b>	<b>13</b>
2.1	Freedom of contract defined . . . . .	13
2.2	The economic case for freedom of contract . . . . .	16
2.3	The economic case against freedom of contract . . . . .	24
2.4	Other arguments for regulating private contracts . . . . .	40
2.5	Legal doctrines regulating freedom of contract . . . . .	42
<b>3</b>	<b>Formation of Contracts</b>	<b>52</b>
3.1	Pre-contractual behavior . . . . .	52
3.2	Avoiding miscommunication . . . . .	56
3.3	Legal doctrines addressing contract formation . . . . .	57
<b>4</b>	<b>Interpretation of Contracts: Contractual Incompleteness</b>	<b>63</b>
4.1	Modeling incomplete contracts . . . . .	64
4.2	The sources of contractual incompleteness . . . . .	70
4.3	Consequences of contractual incompleteness . . . . .	76
4.4	Legal doctrines addressing contractual incompleteness . . . . .	82
4.5	Overall assessment of the law of contract interpretation . . . . .	94
<b>5</b>	<b>Enforcement of Contracts</b>	<b>94</b>
5.1	General issues in enforcement . . . . .	95
5.2	Monetary damages for breach of contract . . . . .	97
5.3	Complications in determining monetary damages . . . . .	110
5.4	Private enforcement of contracts . . . . .	116
5.5	Other law bearing on contract enforcement . . . . .	123
<b>6</b>	<b>Conclusions</b>	<b>123</b>

## 1 Introduction

The essence of a free-market economy is the ability of private parties to enter into voluntary agreements that govern the economic exchange between them. Consequently, the law that governs such agreements is critical to the functioning of such economies. While the law of property determines the configuration of entitlements that form the basis of production and exchange, and the law of torts protects those entitlements from involuntary encroachment and expropriation, it is contract law that sets the rules for exchanging individual claims to entitlements and, thus, determines the extent to which society is able to enjoy the gains from trade. Accordingly, economists interested in the welfare properties of specific institutions in particular, or the micro-foundations of exchange generally, have good reason to take account of the law of contracts.

This chapter, accordingly, surveys the main issues arising in the economic analysis of contract law. We discuss both the main features of contract law as they relate to the problem of economic exchange, and how relevant legal rules and institutions can be analyzed from an economic perspective. In this introductory section, we set out the basic scope, methodology, and organization of the discussion to follow. Subsection 1.1 discusses why formal and informal contracts exist, and what economic functions they serve. Subsection 1.2 distinguishes between positive and normative issues in the economic analysis of contract law, and discusses some methodological problems associated with applying standard economic analysis to legal institutions and when engaging with legal scholarship. Subsection 1.3 identifies limits on the chapter's scope and provides bibliographic recommendations for material we don't cover; and subsection 1.4 sets out the organization of the remainder of the chapter.

A caveat is in order at the outset: although it is conventional to present contract law as a discrete field, one should understand that, to a significant extent, the operation of the rules and institutions discussed below will depend on other aspects of the law, including the fields of tort, bankruptcy, procedure, and evidence. Lawyers have a cliché that describes this interdependence; they say that “the law is a seamless web.” It is useful to keep in mind that many issues that economists would regard as contractual, including some important limits on contractual freedom, are governed not by contract but by tort law. Additionally, the rules relating to certain categories of exchange, such as consumer, employment, insurance, and information-licensing contracts, have developed specialized content to the point that they are often treated as distinct legal fields. Finally, the practical ability of contracting parties to assert their legal entitlements depends importantly on the procedural rules that govern courts and other enforcement institutions. Many of the specific features of contract law that we discuss below cannot be understood except as a response to the costs and other limitations of such institutions.

		Seller	
		Buy insurance	Don't buy
Buyer	Buy insurance	$-p_b, -p_s$	$-p_b, 0$
	Don't buy	$0, -p_s$	$-\ell_b, -\ell_s$

**Figure 1:** Coordination of insurance payments

## 1.1 The economic motive for contracts

In a neoclassical exchange economy of the sort analyzed by Walras (1874) or Arrow-Debreu (Arrow and Debreu, 1954; Debreu, 1959), there is little need for contracts or contract law, since buyers and sellers can exploit all gains from trade through spot transactions. Indeed, in spot markets, such as public bazaars, the parties manage reasonably well without formal contracting. Contracting becomes worthwhile when there is a temporal element to their exchange or one party, at least, is unsure as to what her counterparty will do. For example, when the item to be exchanged needs to be produced or the service being rendered takes time. Absent a contract, the parties could be reluctant to trust each other to complete the agreed upon exchange at the called-upon time, and thus valuable exchange is forgone. Conversely, contracts can be worthwhile even in non-exchange settings, as when advance commitment enhances the value of a gift by enabling reliance by the beneficiary (R. Posner, 1977, and Shavell, 1991) or when a supplier's commitment to remain in a market notwithstanding short-run losses deters competitive entry by rivals (*e.g.*, Rasmusen et al., 1991) or encourages entry by producers of complementary goods. The central question then becomes why commitment is valuable, to which there are several answers.

### 1.1.1 Coordination

The most straightforward reason to use contracts is to coordinate independent actions in situations of multiple equilibria. As an illustration, consider the game depicted in Figure 1, in which a buyer,  $b$ , and a seller,  $s$ , are independently deciding whether to purchase insurance against the loss of a good in transit. The parties' payoffs net of this decision are normalized to zero. Denote the insurance premium when purchased by party  $i$  by  $p_i$  and the *expected* loss suffered by party  $i$  when no insurance is purchased by  $\ell_i$ . Assume that  $\ell_b + \ell_s > p_b > p_s$ ; that is, going uninsured is more expensive than buying insurance, but it is cheaper for the seller to buy the insurance than for the buyer to do so. (This is perhaps because the seller, who packages the goods for shipment, is better able to control moral hazard and thus can obtain a better rate.)

If the parties make their choices independently, there are three Nash (1951) equilibria to this game: one in which the seller buys insurance, one in which the buyer buys insurance, and a mixed-strategy equilibrium in which both buy

insurance with positive probability. Of these, the first equilibrium is the efficient one (and it is also more efficient than any disequilibrium outcome). A contract to play this efficient equilibrium could serve to ensure this outcome is achieved.

Formal contracts are of course not the only way for parties to coordinate among multiple equilibria. The efficiency of the seller-buys equilibrium could make it a focal point for the parties (Schelling, 1960). The literature on “cheap talk” (*e.g.*, Farrell, 1987a, 1993) suggests that such coordination can, in principle, also be achieved by having the parties announce their intentions in advance.

In actual institutional settings, however, contracts offer more stability than focal points or mere announcements. In particular, they provide permanent authoritative records that can be used by parties who suffer from imperfect recall or by those who need to delegate performance to their agents or successors. Note that when contracts are used for pure coordination purposes, they are self-enforcing in the sense that it is in each party’s private interest *ex post* to comply with the chosen equilibrium. Hence, those designing such contracts can devote most of their attention to problems of formation and interpretation, and relatively little attention to problems of enforcement. This coordination function of contracts has been rather less discussed in the law and economics literature than the incentive mechanism functions discussed below, but it may be by far the most important purpose that contracts serve in practice. As Myerson (2004) suggests, coordination games could be the best models through which to understand legal institutions generally.

### 1.1.2 Implementing exchange over time

A second reason for using contracts is to implement exchanges that depend on future events. For instance, consider an insurance contract that covers a loss that occurs in state 1, but not in state 0. Under this contract, the insured pays a premium to the insurer, in exchange for a casualty payment received in state 1, but not in state 0.

In the standard model of risk allocation, goods are state-contingent commodities; for example, an apple in state 0 is considered a different good than an apple in state 1. Treating goods as state-contingent commodities has the advantage of allowing direct application of standard analyses of exchange, but has the disadvantage of abstracting from real institutional issues. In particular, exchange that is mutually beneficial *ex ante* may not be *mutually* beneficial *ex post*. In the insurance example, the insured will not wish to pay the premium *ex post* if state 0 is realized, and the insurer will not wish to make the casualty payment *ex post* if state 1 is realized. Such exchanges cannot, therefore, be implemented in spot markets and require some form of advance commitment.

In the typical insurance context, the transaction is motivated by the insured’s risk aversion. But the need to contract across different states is more general than that and is not reliant on risk aversion. For instance, consider purely speculative exchange between risk-neutral parties, in which trade is motivated by differences of opinion regarding the probability of future events. If one trader thinks the price of orange juice will rise next year and another thinks it will fall,

they can make themselves better off *ex ante* by entering into a forward exchange in which the second promises to deliver to the first. As with the insurance contract, this exchange requires a commitment mechanism since *ex post* one of the parties is sure to regret the deal.

An analogous problem arises with the rental of capital assets or the extension of credit. Even though the owner of an asset may not be its highest-value, she may be unwilling to yield possession to a higher-value user for fear that she will be unable to get it back at the end of the rental period. The law of property provides a partial solution to this problem by entitling the owner to reclaim her asset, but evidentiary difficulties make this alternative an imperfect one (as illustrated by the old maxim, “possession is nine tenths of the law”). The party in possession could claim, for instance, that the transaction was a gift or a sale, or that the agreed lease period had not yet expired. In such settings, a contract that specifies the parties’ relative rights and duties makes the borrower’s promise to return the asset more credible, facilitating exchange.

More generally, some form of commitment is necessary in any exchange in which performance is sequential, because the party who performs first is effectively extending credit to the party who performs second. It may be possible to structure the exchange so that each stage of a party’s performance is timed to coincide with the performance of her counterparty (*e.g.*, an installment sale of goods in which each shipment is delivered C.O.D.), but in many instances such timing may be infeasible or costly. For instance, consider a grocery that requires regular delivery of a perishable commodity such as milk. The costs of making and receiving payment (keeping cash on hand, updating accounts, preventing embezzlement, etc.) generate substantial scale economies if disbursements for multiple shipments are combined into a single monthly payment.

Contracts can also be useful in situations of hidden information. In Akerlof’s (1970) lemons model, for example, adverse selection can prevent efficient exchange when the quality of the good to be traded is known to the seller but not to the buyer, even if the buyer values the good more. This problem can be overcome if the seller of a high-quality good can signal its quality by taking an action that is cheaper for her to take than it would be for the seller of a low-quality good. A common action in this regard is to offer a warranty against the good’s proving to be substandard (Grossman, 1981). Conversely, the buyer could screen for quality by offering a premium to any seller who agrees to provide a warranty. The signal (screen) works only if the seller is bound to honor the warranty, because a low-quality seller can offer (agree to) a worthless warranty just as cheaply as a high-quality one can. Some form of commitment is, thus, needed to implement the exchange.

### 1.1.3 Implementing production over time

Finally, contracts are valuable in promoting production in advance of exchange. Advance production typically increases the surplus available from exchange, but requires sinking resources in ways that may be unrecoverable if the contemplated exchange is not completed. For example, a clothing manufacturer can increase

the price it receives for its products by producing them to meet the needs of its buyers, either very specifically (*e.g.*, custom-tailored suits) or only moderately so (*e.g.*, cutting them so they will be in style for a limited time only). Once the materials used to make the clothing are combined in a particular way, however, they can no longer be easily reconfigured to produce other items. In such settings, producers will be reluctant to sink such expenditures up front unless they can be assured that they will recover their costs *ex post* (see Williamson, 1975, for a seminal analysis of holdup problems).

As Katz (1996a) discusses, suppliers typically cannot capture all the surplus their upfront (pre-trade) investments generate; some of this surplus will go to the buyer. Absent binding purchase commitments prior to investment, suppliers' incentives to invest will be suboptimal, possibly to the point that no investment and, so, no trade occur. Binding contracts can restore proper incentives. Conversely, buyers can also increase their surplus from exchange by making up-front investments, whether out-of-pocket (*e.g.*, buying complementary inputs) or implicit (*e.g.*, ceasing to maintain alternate sources of supply). Because such investments are often relationship specific, however, buyers will not make them unless they can be assured that the exchange price will stay sufficiently low. In some cases, the parties may be able to provide such assurance by manipulating property rights (*e.g.*, Grossman and Hart, 1986; Hart, 1989) or industrial structure (*e.g.*, Shepard, 1987) *ex ante*, but when they cannot do so cheaply, contracts could be a cost-effective alternative.

#### 1.1.4 Limitations of contracts as commitment devices

While contracts are often useful for achieving commitment, they can be imperfect devices for doing for some of the following reasons.

**SPECIFICATION COSTS.** Because it is costly to foresee or to write down all the potential contingencies that might be relevant to the performance of the parties' contractual obligations, actual contracts are often left incomplete. Incompleteness has at least two meanings: first, the contract could simply fail to provide for certain contingencies, in which case a tribunal called upon to enforce the contract, or the parties themselves, would have to decide after the fact what to do if such contingencies arise. Second, the contract could cover all relevant contingencies, but not in as fine-tuned a manner as would be ideal insofar as the contract does not distinguish finely enough, in terms of consequent obligations, among the possible contingencies. In either event, the contract will, with positive probability, fail to assure commitment or commit the parties to a course of action that is suboptimal *ex post*.

**ENFORCEMENT COSTS.** It is never costless to hold a party to his commitment if he is inclined to try to escape it. If the contract is being enforced through the courts, for instance, lawyers must be hired and evidence assembled, and performance or damages are likely to be awarded only after some delay. Such costs make enforcement incredible when the damages from breach are relatively

small and parties can exploit this lack of credibility by holding the level of breach below the threshold necessary to provoke suit (Menell, 1983; Priest, 1978).

**UNOBSERVABLE AND UNVERIFIABLE ACTIONS.** Even if legal commitment has been established and the means for its enforcement are available, the beneficiary of a contractual promise may be unable to determine whether the promise has been kept or broken. For instance, the typical purchaser of a complex consumer product is not in a position to tell whether the product has been manufactured according to warranted specifications; at most she can observe whether the product works as she expected. Even if a promisee can determine that there has been a breach, she may nevertheless be unable to demonstrate that fact to a third-party enforcer at reasonable cost. For instance, a supplier might deliver substitute goods that appear reasonably equivalent to a lay person or to a generalist court, but which the parties themselves know to be substandard. In such situations, the promisee's inability to prove that the promise has been breached renders it ineffective as a method for assuring commitment. On the other hand, as we discuss in §4.3.1, the parties can sometimes contract around the court's lack of expertise (see, *e.g.*, Hermalin and Katz, 1991; Maskin and Tirole, 1999).

**DYNAMIC INCONSISTENCY.** In cases where the purpose of contractual commitment is to promote specific investment, the parties' incentives to stick with their deal may change after the investment has been completed (*e.g.*, Laffont and Tirole, 1988; Aghion et al., 1994). In particular, the parties may all wish to modify or renegotiate their bargain. But if the parties anticipate that renegotiation will take place, it could prove impossible to induce them to undertake efficient investments *ex ante*.

**THE NEED FOR PRE-CONTRACTUAL COMMITMENT.** Some commitments, in order to serve their purpose, must be undertaken before the parties are in a position to engage in voluntary contracting. For example, parties may spend resources on finding contractual partners or on determining whether exchange is worthwhile. Even once an available partner and potential transaction are identified, it typically takes time and expense to negotiate terms; and commitments are often less valuable if they are delayed until bargaining is completed. (For an extreme example, consider the case of an emergency paramedic who must decide whether and how to treat an unconscious accident victim who is not carrying an insurance card.) In ongoing or repeated relationships it is possible for the parties to agree to accept liability in advance of a final bargain, but in one-shot or new relationships it is not.

The law of contracts has recognized most of these problems and has devised a variety of doctrinal arrangements to deal with them; and the succeeding sections of this chapter will discuss such arrangements in more detail. The reader should appreciate at the outset, however, that because these legal arrangements are themselves imperfect, parties will often want to use legal contracts in com-

ination with other legal and non-legal commitment devices, such as deposits, third-party guaranties, reputational bonds, repeated dealing, mutual threats, hostage exchange, investing in altruistic preferences, and the like. The success of formal contract law, accordingly, depends importantly on how well it functions in combination with these substitute and complementary devices, and not just on how well it works in isolation.

## 1.2 Law & economics issues in contracting

### 1.2.1 Normative issues

Much normative discussion relating to contract law revolves around the issue of freedom of contract—to what extent will unregulated private contracting lead to desirable social consequences? We discuss this issue, and its relationship to standard issues in welfare economics, in §2 below. At the outset, however, it is worth observing that the dominant normative consideration here, even more so than in other fields of law and economics, is transactional efficiency. In part, this dominance follows the implicit assumption, shared by most commentators, that externalities and analogous market failures are a less significant phenomenon in this field of law than they are in, say, tort law. But the focus on efficiency also stems from the general recognition that much of contract law, putting aside specialized categories such as consumer and employment contracts, is designed for the purpose of facilitating exchange between business firms and analogous commercial entities. Such entities are motivated primarily by economic gain as opposed to nonpecuniary considerations, enter into legal obligations deliberately and at arms length, and are rational in the standard economic sense. It is thus easier to justify applying the efficiency norm to such voluntary arrangements than to the typical tort case involving persons drawn together involuntarily and outside of market institutions.

A more complete account of social welfare, however, would consider competing normative values such as fairness, equity, etc. to the extent they affect social well-being. While there has been relatively little economic analysis of contract law in this regard, we will discuss these values insofar as they are relevant to specific analytical and doctrinal topics.

### 1.2.2 Positive issues

While most work on the economics of contract law has sought, at least in part, normative conclusions, there is a segment of the literature devoted to predicting and explaining how different contractual rules affect private transactions, and why contracting parties might choose one contractual device rather than others. For example, a variety of authors (*e.g.*, Joskow, 1987; Crocker and Masten, 1988; Pirrong, 1993) have investigated the connection between the use and duration of contractual agreements and the extent of relationship-specific investments. Other authors (*e.g.*, Klein, 1980; Goldberg and Erickson, 1987; Hadfield, 1990; Gergen, 1992) have sought to explain the common use of indefinite or open terms in otherwise clearly negotiated agreements; and still others

(*e.g.*, Weinstein, 1998; Goldberg, 1998, 2000) have sought to explain particular risk-sharing or option terms. In the field of commercial contracts, there is a vigorous literature discussing the determinants of secured lending (see, *e.g.*, Scott, 1986; Schwartz, 1989; Triantis, 1992; Mann, 1997a,b). The antitrust literature has considered whether certain contractual practices are more likely to have efficiency or anticompetitive motivations (*e.g.*, Cheung, 1969; Kenney and Klein, 1983, 2000; Crocker and Masten, 1988; Klein and Murphy, 1988; Masten and Snyder, 1993). While a full survey of this literature is beyond the scope of this chapter, the reader should be aware that many of the issues discussed in the later sections have been discussed empirically. At the same time, however, it also true that the empirical study of contract is relatively less developed than the theory, leaving much room for future researchers.

### 1.2.3 Economic versus non-economic theories of contract law

In recent years, the majority of contracts scholarship in the legal academy has reflected a methodology based on economic analysis, and most legal scholars in the field have become conversant with economic concepts such as efficiency, moral hazard, adverse selection, and the like. (Conversely, over the same period, economic theorists have increasingly come to appreciate the importance of legal concepts such as principal and agent.) Nonetheless, a considerable amount of discussion in legal circles continues to reflect alternative conceptual frameworks; and economists engaged in interdisciplinary work should be aware of these competing frameworks and their underlying assumptions. Three major competing perspectives are worth brief discussion here; we denote these as the corrective justice, liberal autonomy, and social constructivist perspectives.

**CORRECTIVE JUSTICE.** Corrective justice, the most longstanding of these perspectives, has intellectual roots that trace back to classical writers such as Aristotle. This perspective holds that judicial institutions are only justified in acting to redress unjust or wrongful situations. Examples of such redress in the contractual setting would include restitution of unjustly received benefits or compensation for wasted expenditures incurred in reliance on a broken promise.

The corrective justice approach seeks the restoration of some past or proper state of affairs, and thus can stand at odds with the economic approach to law, which tends to regard past gains and losses as sunk and to emphasize incentives for future behavior (*e.g.*, Easterbrook, 1984). It is less clear, however, that the corrective justice approach has any implications at all for *ex ante* analysis of legal problems; and many legal writers in this tradition (*e.g.*, Dworkin, 1980) have distinguished between the use of economics in judicial settings, which they regard as requiring decisions according to principle, and its use in legislation and contractual planning, which may legitimately be designed to promote goals of social policy or private advantage.

**LIBERAL AUTONOMY.** The liberal or autonomy-based perspective (see, *e.g.*, Barnett, 1986) emphasizes the individual as opposed to the collective interest.

From this point of view, individual rights should take priority over more general concerns of society. Such a perspective is plainly more consonant with the economic approach than is the corrective justice perspective. Standard economic measures of welfare are based on aggregates of individual utility, and under conventional assumptions of welfare economics, liberal freedoms tend to lead to desirable economic outcomes (see §2.2 below). There will be tension between the liberal and economic approaches, however, whenever market failures or transaction costs prevent the completion of efficient exchanges, as Sen (1970) has shown through social choice theory. In a contractual setting, for instance, liberal theorists might argue that obligations to which the parties have not knowingly consented may not be imposed on them even when those obligations would be both efficient and distributionally equitable. Conversely, libertarians might argue that the law should protect even anticompetitive agreements, such as price-fixing, on the basis that the conduct is voluntary and that consumers have no inherent right to trade with producers on any particular terms.

**SOCIAL CONSTRUCTIVISM.** The social constructivist perspective takes the position that the normative goals of society—and in the view of many writers, its descriptive categories as well—are determined by collective and ongoing deliberation among its citizens, and accordingly, that the main goal of legal institutions should be to provide adequate opportunity for such decision making. So defined, this perspective can be seen to encompass a variety of more specific normative positions, including those of writers who emphasize civic virtue (Kronman, 1987), democratic self-government (Pildes and Anderson, 1990), an egalitarian distribution of political and economic power (Kennedy, 1982), or the primacy of particular substantive values, which economists would call merit goods (Radin, 1987). On such views, the constituted citizenry may well decide to pursue economic goals such as efficiency, but may, with equal legitimacy, decide to pursue other procedural or substantive goals. For example, if the citizenry decided the well-being of local manufacturers were sufficiently important to outweigh productive efficiency or economic liberty, this decision would legitimize restricting interregional trade. It follows that there is no particular reason why individualist institutions, like contract or the market, are more legitimate venues for such decision making than collective or political ones.

**RELATIONSHIP BETWEEN ECONOMIC AND NON-ECONOMIC THEORIES.** These three rival perspectives can and often do combine to provide overlapping arguments against the use of economic analysis in contract law. For example, the most prominent alternative theory of contract law to be put forward in recent years, the so-called “will theory” (*e.g.*, Fried, 1981), holds that promises ought to be kept for their own sake, in part because promise-breaking is a deontological wrong that needs to be rectified as a matter of corrective justice, and in part because enforcing promises is necessary to respect the autonomy of the promisee (and arguably the promisor as well). In response to such arguments, more economically oriented writers (*e.g.*, Craswell, 1989a; Shavell, 1991) have

replied that most economic analysis of contract law is aimed at filling the gaps in incomplete agreements and setting default rules that operate when the parties have expressed no preference regarding a particular issue. If we accept that most business contracts have economic purposes, economic analysis will help those interpreting contracts to better implement the parties' will.

The will theorists might offer, as a rejoinder, the observation that there is a difference between what most contracting parties subjectively understand when they make and receive promises, and what would be most efficient for them to do. For instance, most people, even those with substantial business experience, intuitively understand promises to bear inherent moral force and believe that the mere fact that it turns out to be suboptimal to carry them out should not count as an excuse for non-performance. If this claim is true—and it should be plain that it is an empirical claim—then will theorists (*e.g.*, Charny, 1991) would argue that it violates the parties' autonomy and dignity for the state to enforce the efficient bargain, which they perhaps should have made, rather than the actual arrangements that they did make.

An obvious response for economists to such non-economic considerations is to incorporate them into a larger welfare analysis in which the relevant non-economic values are interpreted as arguments to be traded off against one another in a Bergson-type social welfare function (SWF). Shavell and Kaplow (2002) have presented such an analysis at length, but have achieved mixed success in persuading non-economically-oriented legal scholars of its merits. Observe, in this regard, that the three main categories of non-economic theories vary in their compatibility with the SWF approach. Pure corrective-justice theorists would reject the SWF on the grounds that legally appropriate actions can only be justified with reference to right and wrong, and never with reference to consequences. Liberal autonomy theorists tend to be somewhat more accepting of the SWF approach, but they would reject the idea that rights could be traded off against each other or against economic values (formally, this is equivalent to adopting a lexicographical preference ordering for the SWF). Social constructivists would view the whole SWF approach as logically circular, since they view the social values that would underlie any SWF as endogenously determined by political and cultural processes of which economic policy discussions are an integral part. In their view, starting with a SWF and then attempting to maximize it is putting the cart before the horse.

Most importantly, economists who study legal institutions should recognize that lawyers and legal commentators often employ different methodologies than they do. In particular, legal scholars do not typically draw the same sharp distinction between positive and normative analysis as economists do. On the contrary, many seemingly descriptive statements made by lawyers or appearing in legal texts are understood in context to carry significant normative overtones, and *vice versa*. Non-lawyers who are insufficiently appreciative of the mixed nature of such discourse may miss an important part of what is being said, and may be led to model the phenomena under discussion in an incorrect or misleading way (Katz, 1996a). Similarly, lawyers' customary method of reasoning inductively from individual cases, rather than deductively from general principles,

makes many lawyers reluctant to accept some of the standard methodological practices of economists, including formal modeling and the use of statistical aggregation. When presenting positive arguments or analyses to legal audiences, accordingly, it is generally necessary to lay out one's methodological assumptions explicitly and at the outset. Otherwise, one risks being misunderstood by those lawyers who are accustomed to assuming a normative subtext or an individualist perspective whenever legal topics are being discussed.

### 1.3 What this chapter is not

#### 1.3.1 A guide to contract theory

Not surprisingly, there is a strong link between the law and economics of contracts and the economic sub-discipline of contract theory.<sup>1</sup> Contract theory provides a framework to analyze the scope and limits of what contracts can accomplish, at least at a theoretical level. It is, however, beyond the scope of this chapter to offer a comprehensive guide to contract theory. To be sure, some contract theory will be discussed as appropriate within the context of the issues discussed below; but no one should mistake such a scatter-shot approach for an attempt at a systematic treatment of the subject.

For readers interested in contract theory, there are a number of excellent introductions. Most new graduate microeconomics texts devote at least a chapter to the subject (see, *e.g.*, Kreps, 1990; Mas-Colell et al., 1995; Varian, 1992). Good book-length introductions are Laffont and Martimont (2002) and Bolton and Dewatripont (2005). For those looking for cheap—in fact free—introductions, one of us (Hermalin) hosts two introductory manuscripts (Caillaud and Hermalin, 2000a,b) on his web page.

#### 1.3.2 A guide to the law of contracts

While this chapter discusses most of the important economic issues relating to the law of contracts, it does not attempt to present a survey of major legal doctrines in the field. It is worth noting, as a general observation, that many legal doctrines may seem indistinguishable from an economic viewpoint. Nevertheless, they vary in their specific content and their differences are of practical importance to lawyers and clients. For example, the doctrines of mistake, impossibility, and frustration of purpose all excuse contractual liability in extreme or unexpected circumstances and thus allocate risk to the recipient of a contractual promise, but their legal application varies.

Economists interacting with lawyers or pursuing research in contract law, accordingly, should be aware of distinctions such as these and of the associated views of major legal authorities. The most useful general treatise on US contract law is probably Farnsworth (2004), available in both one- and three-volume editions; a shorter introduction to the subject can be found in Chirelstein (2006).

---

<sup>1</sup>Whether this link is a valuable one is, however, a matter of debate. See, for instance, the *Yale Law Journal* debate among E. Posner (2003a), Ayres (2003), and Craswell (2003).

For statutory material applicable to US contracts for the sale of goods, the best source is White and Summers (2000).

Less attention has been paid by economists to non-US legal systems, making them a potentially fruitful source for future research; here leading English-language treatises include Honnold (1999) and Schlechtreim (1998) on international sales contracts, Bonell (1997) on other international commercial contracts, Atiyah (2003) and Treitel, ed. (2003) on English contract law, and Lando and Beale, eds. (2002) on the contract law of the civil law systems of continental Europe. In this chapter, however, most discussions of legal doctrine will be restricted to US law.

#### 1.4 Organization of the chapter

The remainder of this chapter is divided into four sections corresponding to the major conceptual divisions of contract doctrine. Section 2, entitled “Freedom of Contract,” discusses the scope of private parties’ power to create binding contractual obligations. It analyzes the major doctrines that govern which bargains will be recognized and enforced by state legal institutions, and which parties are empowered to create enforceable bargains. It also considers how doctrinal limits on freedom of contract correspond to the economic criteria for determining whether decentralized trade will lead to optimal welfare outcomes.

Section 3, entitled “Formation of Contracts,” discusses the extensive body of legal doctrine that governs the procedural mechanics of exchange, as well as the rules that govern the parties’ obligations before they enter into exchange. These formal rules, by attaching consequences to the various acts and omissions that bargainers can choose from in searching for and negotiating with potential contractual partners, affect parties’ incentives to make and to respond to offers, to delay, to bluff, and to communicate with one another in the first place.

Section 4, entitled “Interpretation of Contracts: Contractual Incompleteness,” discusses the problems that arise when it is unclear whether those parties who are empowered to create binding contracts have actually done so, and if they have, what specific obligations they have created. Recent work in microeconomic theory has also been concerned with this problem, especially as it relates to the ability of third-party enforcers to verify the parties’ bargain. From a legal perspective, the problem is governed by the various doctrines dealing with contract interpretation, and this section shows how the legal rules in this area affect and respond to the economic problem of incomplete contracts.

Section 5, entitled “Enforcement of Contracts,” discusses how the foregoing rules and institutions are translated into effective costs and benefits that can motivate parties to comply with their obligations and to insure against others’ lack of compliance.

Finally, the concluding section offers some overall perspectives on the entire discussion, relates its main points to analogous or complementary doctrines in related fields of law, and offers some speculations regarding the path of future legal and economic developments in the area.

## 2 Freedom of Contract

The threshold issue in any discussion of contract law is freedom of contract—the extent to which the law sanctions the use of contracts as a commitment device. No legal system enforces all voluntary private agreements, but in the US and other industrial democracies, most contracts that support legitimate economic exchange are at least presumptively enforceable. Still, the limits of freedom of contract vary among Western countries and are an important element of regulatory policy. This section, accordingly, analyzes and evaluates those limits in economic terms. Subsection 2.1 defines the scope of the issues included under the framework of contractual freedom; subsection 2.2 reviews the presumptive economic case in favor of freedom of contract; subsections 2.3 and 2.4 discuss the main arguments, economic and otherwise, that are typically used to justify limits on private contracting; and subsection 2.5 outlines the major doctrinal limitations on freedom of contract that are in force in the US and in related systems, and relates those doctrines to the economic arguments set out in the prior subsections.

### 2.1 Freedom of contract defined

#### 2.1.1 State regulation versus state enforcement

The concept of contractual freedom encompasses a number of distinct considerations. One important distinction is between negative and affirmative government sanction: are the parties permitted to enter into a given contract versus will the law enforce it? These two questions are not equivalent: there are many agreements that cannot be enforced in the courts but that can still be useful as commitment devices if the parties can manage to implement them privately. For instance, prior to the passage of the Sherman Act, contracts in restraint of trade were unenforceable under US common law; this reduced their incidence, but not to the point of elimination. Under modern antitrust statutes, in contrast, government disapproval of anticompetitive conduct goes beyond non-enforcement to include active interference through civil liability and, in some cases, criminal prosecution.

In the remainder of this section, accordingly, we focus on those limits on contractual power that are motivated by regulatory concerns that the agreement itself is socially undesirable for reasons of inefficiency, inequity, and other substantive objections.

#### 2.1.2 Positive versus negative contractual freedom

The freedom to enter into contractual liability would be rather less meaningful were it not accompanied by the complementary freedom to avoid liability for contracts into which one does not wish to enter. In general, this negative freedom applies to most types of contractual obligations, but not all. In traditional common law, for instance, some businesses (*e.g.*, mills, ferryboats, railroads, and the like) are designated as common carriers and are obliged to enter into

exchange on standard terms with anyone who wishes. Modern statutes have expanded such duties in a variety of ways: for instance, the essential-facilities doctrine in antitrust law requires vertically integrated firms to make certain stages of production available on a contractual basis to their non-integrated competitors; and anti-discrimination laws require businesses to deal with customers and suppliers on an equal basis without regard to race, religion, or the like.

Additionally, some rules of tort and property law have the effect of requiring rightholders, under some circumstances, to transfer their entitlements to persons in need or to the general public. For instance, the doctrine of eminent domain requires landowners to convey their property to the state when required for an appropriately defined public use. (As in the US, the government could be constitutionally obliged to pay just compensation when it takes private property.) Similarly, the tort doctrine of necessity allows parties in dire need to make use of others' property when voluntary contracting is not feasible (as with an endangered hiker who breaks into an empty cabin to find food or shelter) so long as compensation is paid *ex post*.<sup>2</sup> Such doctrines are typically conceptualized by legal theorists as internal limits on the underlying entitlement at issue rather than as restrictions on contractual freedom; however, they usually amount in practice to restrictions on contractual freedom because of the infeasibility of contracting around them. One could theoretically imagine contracting in advance with the government not to exercise its right of eminent domain, but such contracts are rare and it is doubtful whether they are immune to abrogation by subsequent governments.

Finally, some doctrines of contract law impose promissory liability even when the promisor has not actually intended to enter into a contractual exchange. The doctrine of promissory estoppel, for instance, holds parties to promises on which a reasonable person would foreseeably rely, at least to the extent necessary to protect the promisee's reliance; and the doctrine of trade usage holds parties to contracts that experienced market participants would view as legally enforceable under similar circumstances, even if the parties had not themselves subjectively intended to be bound in the particular instance. For the most part, such doctrines provide rules of interpretation rather than of substantive contractual freedom, in that it is possible to avoid liability by being sufficiently explicit in one's communications. In some cases, however, the law does not allow one to disclaim liability for one's representations or promises; for instance, commercially sophisticated parties dealing with less sophisticated counterparts can find themselves bound to statements made to the less experienced party even if sophisticated persons would understand such statements do not entail legal obligation. Such restrictions on disclaimers are usually motivated by considerations of market failure, transaction cost, or distribution, as discussed below.

---

<sup>2</sup>Modern legal theorists, following Calabresi and Melamed (1972), use the concept of property and liability rules, further discussed in this *Handbook*, to identify situations in which such involuntary exchanges are authorized by law: entitlements subject to such imposition are said to be protected by a liability rule, while entitlements that are immune from such imposition are said to be protected by a property rule.

### 2.1.3 Mandatory versus default terms

Finally, as the example of promissory estoppel illustrates, it is useful to distinguish between mandatory contract terms, which the parties are not legally free to change, and default terms, which the parties are theoretically free to change but which govern the contract to the extent the parties are silent. An example of the former would be the constitutional prohibition on involuntary servitude, which, for instance, prevents people from pledging their future labor for long periods, even when they might find wish to do so, such as to provide collateral for a loan. An example of the latter would be the various warranties of quality that are implied under current US law in contracts for the sale of goods. Sellers are generally permitted to disclaim such warranties, subject to some limits imposed by consumer-protection and product-liability law, by following the requisite procedures, which usually require some specific notification to the buyer.

Because it is costly to write complete contracts, all systems of contract law must provide default terms to cover the issues over which the parties do not specifically bargain (see §4.4.1 below). The regulatory effects of default terms, however, are bounded by the costs of contracting around them. It could be a reasonable approximation to ignore such effects in many instances, especially in the commercial setting where parties are sophisticated and have access to legal advice. In some cases, however, re-contracting costs are substantial and the choice of default rule will have the effect of privileging one outcome over others. For example, in mass transactions in which parties communicate through standard forms, it is impractical to reconcile all the discrepancies between the various forms; and under modern doctrine, the legal default applies to all issues on which the forms do not agree. This often has the consequence of providing broad product warranties and leaving the seller open to liability for damages following breach, even if the seller attempted to disclaim such liability.

As with the decision to withhold enforcement, the provision of default rules may be motivated either by regulatory purposes or by the desire to conserve on transaction costs. In situations where there is some doubt about whether a specific contract term should be discouraged, for example, a default rule supplies some deterrent effect while still allowing parties whose gain from the term is sufficiently high to opt out of the default at a price. For example, Camerer et al. (2003) advocate using default rules as a relatively libertarian method of regulating against poor decisions caused by bounded rationality. Or, as Ayres and Gertner (1989, 1992) have suggested, and as we discuss further in §4.4.1 below, a default rule may be employed as a screen to induce parties to reveal private information that might be relevant to *ex post* interpretation or to the *ex ante* decision whether to enter into exchange.

## 2.2 The economic case for freedom of contract

### 2.2.1 Welfare economics

An economic case for or against freedom of contract is based on the consequent welfare implications. In this section, we briefly review what a welfare analysis of markets suggests about freedom of contract.

Economists typically use two welfare criteria. One, known as Pareto efficiency, evaluates a proposed allocation among a set of actors by asking whether there exists a second allocation that (i) none of the actors prefer less than the proposed allocation and (ii) at least one of the actors actually prefers to the proposed allocation. If such a second allocation exists, the proposed allocation is deemed inefficient (alternatively, Pareto inferior or Pareto dominated). The second allocation in this case is deemed Pareto superior. If no such second allocation exists, the proposed allocation is deemed efficient.

While the Pareto criterion is useful for ruling *out* undesirable allocations, it doesn't always serve as a useful guide for selecting a desirable allocation. For instance, it offers no guidance as to who should receive an indivisible object; any allocation other than throwing it away is Pareto efficient because a switch to another allocation would not be favored by the party losing the object.

An alternative welfare measure is to consider a function that aggregates, in some way, the preferences of the actors in question. A full discussion of welfare functions is beyond the scope of this chapter.<sup>3</sup> We will limit attention to the utilitarian welfare function,  $W = \sum_{i \in \mathcal{I}} u_i$ , where  $\mathcal{I}$  is an index set over the actors and  $u_i$  is the utility the  $i$ th actor enjoys from the proposed allocation (if there is a stochastic aspect to the allocation,  $u_i$  should be understood to be  $i$ 's *expected* utility). Any allocation that maximizes social welfare,  $W$ , must be Pareto efficient;<sup>4</sup> but Pareto efficiency does not necessarily imply that social welfare is maximized. For instance, as noted, any allocation of an indivisible object (other than throwing it out) is Pareto efficient, but only the allocation that awards it to the person who values it most is welfare maximizing.

A stronger connection between the two welfare criteria can be achieved if one accepts the existence of a transferable good (typically taken to be money). Now the "losers" from moving to a social welfare-maximizing allocation can receive payments from the "winners" as compensation. If preferences can be captured by a quasi-linear utility function (*i.e.*, of the form  $u + y$ , where  $y$  is money and  $u$  is utility from other goods), then an allocation is Pareto efficient

<sup>3</sup>The interested reader is directed to Chapter 22 of Mas-Colell et al. (1995) or Chapter 6 of Laffont (1988). Also see Arrow's (1963) classic book.

<sup>4</sup>Proof: Suppose not. Then, although the allocation  $\mathbf{a}$  maximizes  $W$ , there is another allocation  $\tilde{\mathbf{a}}$  that Pareto dominates  $\mathbf{a}$ . Let  $u_i$  and  $\tilde{u}_i$  denote the utilities under the allocations  $\mathbf{a}$  and  $\tilde{\mathbf{a}}$  respectively. By the definition of Pareto dominance,  $\tilde{u}_i \geq u_i$  for all  $i$  and there is at least one  $i$  such that  $\tilde{u}_i > u_i$ . But, then,

$$\sum_{i \in \mathcal{I}} \tilde{u}_i > \sum_{i \in \mathcal{I}} u_i,$$

which contradicts the assertion that  $\mathbf{a}$  maximizes social welfare.

if and only if it maximizes welfare.<sup>5</sup> For this reason, economists are generally satisfied with social welfare (total surplus) as an appropriate welfare standard when transfers are feasible.<sup>6</sup> Note that this analysis relies on the marginal utility of the transferable good being constant across individuals, so that we are still seeking to maximize  $\sum_{i \in \mathcal{I}} u_i$ ; transfers among the actors are irrelevant to maximizing welfare because the benefit one actor gets from receiving a dollar is completely offset by the cost another incurs transferring that dollar.

Competitive markets are typically seen as doing well with respect to the Pareto efficiency criterion. Under somewhat stringent conditions—in particular, (i) that complete markets for all commodities (including commodities such as clean air and water) exist, (ii) that no actor has market power (*i.e.*, acts as a price setter rather than a price taker), and (iii) symmetric information—a general equilibrium of the economy will be Pareto efficient (see, *e.g.*, §6.3 of Debreu, 1959). This result is known as the First Welfare Theorem.<sup>7</sup>

The First Welfare Theorem applies to the economy as a whole. If the entire economy is an Arrow-Debreu economy and in equilibrium, then any particular market within the economy must be “efficient,” insofar as any change in it

<sup>5</sup>Proof: Footnote 4 *supra* established that welfare maximization implies Pareto optimality. Consider a Pareto optimal allocation of real goods and money. Let  $u_i$  and  $y_i$  denote the utility components under this Pareto optimal allocation. Suppose this allocation does not maximize welfare. Then, there exists another allocation of real goods, with utility components  $u_i^*$  such that

$$\sum_{i \in \mathcal{I}} u_i^* > \sum_{i \in \mathcal{I}} u_i.$$

Because the “ $u_i$ ” allocation is Pareto efficient, it cannot be that  $u_i^* \geq u_i$  for all  $i$ ; that is, a change in allocation must create “losers,” for whom  $u_i < u_i^*$ . Let  $\mathcal{L}$  be the set of losers and  $L$  be the number of elements in  $\mathcal{L}$ . As just noted,  $L > 0$ . Similarly, there must be winners (*i.e.*, those for whom  $u_i^* \geq u_i$ ). Let  $\mathcal{W}$  be the set of winners. For losers, define  $\tau_i = u_i - u_i^*$ . For winners, define  $t_i = u_i^* - u_i$ . Because the “ $u_i^*$ ” allocation is welfare maximizing,

$$\sum_{i \in \mathcal{W}} t_i - \sum_{i \in \mathcal{L}} \tau_i \equiv G > 0.$$

Finally, consider the allocation of real goods that produces utilities  $u_i^*$ , allocates  $y_i - t_i$  in money to each winner and  $y_i + \tau_i + G/L$  to each loser (note the additional transfers sum to zero, hence are feasible). For a winner,  $u_i^* + y_i - t_i = u_i + y_i$ , so winners are indifferent. For a loser,  $u_i^* + y_i + \tau_i + G/L = u_i + y_i + G/L$ , so losers are strictly better off. But, then, we have an allocation (including money) that Pareto dominates our original allocation, contradicting its Pareto optimality. The result follows by contradiction.

<sup>6</sup>The Kaldor-Hicks criterion is even more flexible: An allocation,  $\alpha$ , over real goods is a Kaldor-Hicks efficient allocation if there is no other allocation over real goods  $\alpha'$  and no profile of transfers,  $\mathbf{t}$ , such that the overall allocation  $(\alpha', \mathbf{t})$  Pareto dominates the overall allocation  $(\alpha, \mathbf{0})$ , where  $\mathbf{0}$  means no transfers are made. It is sufficient that  $\mathbf{t}$  exist—whether or not these transfers are made—for  $\alpha'$  to dominate  $\alpha$  according to the Kaldor-Hicks criterion. If all individuals have utility  $u_i + y_i$ , where  $u_i$  is an individual’s utility over real goods and  $y_i$  is his allocation of money (transfer), then an allocation of real goods is Kaldor-Hicks efficient if and only if it maximizes  $\sum_{i \in \mathcal{I}} u_i$ . See Chapter IV of Arrow (1963) for details.

<sup>7</sup>There is also a second Welfare Theorem that has to do with the ability of prices to serve as appropriate incentive devices in a competitive (Arrow-Debreu) economy; that is, any Pareto efficient allocation can be supported as a general equilibrium of the economy by selecting the appropriate prices (see §6.4 of Debreu for details).

changes at least part of the overall allocation and no other overall allocation is Pareto superior. If, however, one doubts the Arrow-Debreu model adequately models the *entire* economy, one can still ask, in certain circumstances, whether a particular market in it is achieving an efficient allocation.

A single competitive market (*i.e.*, a market in which neither buyers nor sellers exercise market power) will achieve an equilibrium at the price that equates demand and supply. Under normal assumptions, this equilibrium is unique. Assume the demand curve is a good approximation for the social marginal benefit curve (*i.e.*, there are essentially no positive externalities from the good in question and income effects are *de minimis*).<sup>8</sup> Assume, further, that the supply curve is a good approximation for the social marginal cost curve (*i.e.*, there are essentially no negative externalities from the good in question). Then, as is well known, welfare is the area beneath demand and above supply from 0 to the number of units traded. This area—that is, total welfare—is maximized by exchanging the quantity that corresponds to the intersection of supply and demand. But this quantity is precisely the quantity that will be exchanged in competitive equilibrium—the competitive equilibrium maximizes total welfare and, thus, achieves a Pareto-efficient allocation.

### 2.2.2 Theoretical justifications for freedom of contract

The welfare-theoretic arguments of the previous subsection rely on the assumption of competitive markets. Many writers, however, have had the intuition that freedom of contract is desirable much more generally.

In the law and economics literature, this intuition is most prominently associated with the work of Ronald Coase and his widely-cited “Coase Theorem” (Coase, 1960). Despite its formal sounding name, the Coase theorem is not a theorem in the traditional sense (nor did Coase suggest it was). Indeed, as Medema and Zerbe (2000) point out, there is not even an agreed upon statement of it. We offer the following version:

**Theorem 1 (Coase theorem)** *Consider a bilateral contracting situation in which (i) the parties are rational with respect to their individual self-interests; in which (ii) the parties can agree on any contract without incurring transaction costs; and in which (iii) the parties’ utilities are additively separable over the allocation of real goods and monetary transfers (i.e., are of the form  $u_i + y_i$ , where  $u_i$  is party  $i$ ’s utility from the allocation of the real goods and  $y_i$  is his or her net transfer). Then the allocation of real goods after contracting will maximize total welfare regardless of the initial allocation of real goods.*

In this formulation, the Coase theorem is a true theorem with the following proof: Let  $\alpha$  be any real-good allocation that does *not* maximize welfare. We need to show that no such  $\alpha$  will be implemented via contracting. Let  $\alpha^*$  be a welfare-maximizing allocation. Clearly, if  $u_i(\alpha^*) > u_i(\alpha)$  for both parties

---

<sup>8</sup>For a detailed discussion of measuring consumer benefit and the consequence of income effects for such measurements see Chapter 10 of Varian (1992). Also see Willig (1976).

$i$ , then  $\alpha$  will not be implemented—because both parties act rationally and they can costlessly contract, they won't settle for  $\alpha$ . Consider the only other possibility:  $u_i(\alpha^*) > u_i(\alpha)$  for one party and  $u_j(\alpha^*) \leq u_j(\alpha)$  for the other. Let  $\tau_M = u_i(\alpha^*) - u_i(\alpha)$  and  $\tau_m = u_j(\alpha) - u_j(\alpha^*)$ . The optimality of  $\alpha^*$  entails  $\tau_M > \tau_m$ . Pick any transfer  $\tau$  such that  $\tau_m < \tau < \tau_M$ . Then a contract that selected allocation  $\alpha^*$  and had  $i$  transfer an additional  $\tau$  to  $j$  would be preferred by both parties to a contract that implemented allocation  $\alpha$ . Costless contracting and the parties' rationality thus rule out  $\alpha$  being the contracted-for allocation.

**Corollary 1** *Under the assumptions of the Coase Theorem, interference or restrictions on the contract the parties sign cannot increase total welfare; that is, under these assumptions, there should be freedom of contract if the only welfare issue is the total welfare of the parties to the contract.*

How strong a case the Coase Theorem makes for freedom of contract depends on the appeal of its assumptions. Consider, first, assumption (iii); that the parties have quasi-linear utility functions. That assumption is common to most welfare analyses of partial-equilibrium settings. Although the assumption can frequently be justified, it is not always.<sup>9</sup> Suppose, for instance, that one party's utility is  $u + y$  if  $y \geq 0$ , but  $-\infty$  if  $y < 0$ . An interpretation is that this party simply cannot survive if made to make transfers. Now the Coase result could fail to hold if this party is the one who must transfer to ensure a welfare-maximizing outcome. On the other hand, if this party is the recipient of transfers, then welfare-maximization would continue to hold.

As Medema and Zerbo (2000) note, the Coase Theorem has two conclusions. One is an efficiency conclusion—private contracting will lead to a welfare-maximizing solution. The other is an invariance conclusion—the initial allocation is immaterial for whether a welfare-maximizing solution is reached. As just discussed, relaxing condition (iii) can undermine both conclusions, but, in a sense, it primarily undermines the invariance result. Efficiency, as judged by the Pareto criterion, will generally still be attained:

**Theorem 2 (Modified Coase Theorem I)** *Consider a bilateral contracting situation in which (i) the parties are rational with respect to their individual self-interests, but are not mean-spirited; and in which (ii) the parties can agree on any contract without incurring transaction costs. Then the allocation after contracting will be Pareto efficient regardless of the initial allocation.*

**Proof:** We need to show that the parties will never settle on an allocation,  $\mathbf{a}$ , that is Pareto dominated (note, now, an allocation may include the transferable good). Consider such an allocation. By definition, there exists at least one other allocation,  $\mathbf{a}^*$ , such that

$$u_i(\mathbf{a}^*) \geq u_i(\mathbf{a}) \tag{1}$$

<sup>9</sup>Willig (1976) provides justifications for assuming quasi-linear utility that are applicable in many contexts.

for both  $i$  and such that  $u_i(\mathbf{a}^*) > u_i(\mathbf{a})$  for at least one  $i$ . Case 1: The inequality in expression (1) is strict for both  $i$ . Then  $\mathbf{a}$  will not be implemented—both parties act rationally and they can costlessly contract, so they won't settle for  $\mathbf{a}$ . Case 2: (1) is an equality for one  $i$ . Then that  $i$  will refuse to implement  $\mathbf{a}^*$  over  $\mathbf{a}$  only if he is mean-spirited, which we have assumed he isn't. Hence, a Pareto-dominated allocation will not be implemented. ■

The modicum of altruism in the modified Coase Theorem—that the parties not be mean spirited—is unnecessary if we assume a strict-tradeoff condition:

**Condition 1 (Strict Tradeoff)** *Let  $\mathcal{A}$  be the set of feasible allocations and  $\mathbf{a}_1$  and  $\mathbf{a}_0$  be two elements of  $\mathcal{A}$ . Then, if  $u_1(\mathbf{a}_1) \geq u_1(\mathbf{a}_0)$  and  $u_2(\mathbf{a}_1) \geq u_2(\mathbf{a}_0)$  with at least one inequality holding strictly, there exists an  $\hat{\mathbf{a}} \in \mathcal{A}$  such that  $u_1(\hat{\mathbf{a}}) > u_1(\mathbf{a}_0)$  and  $u_2(\hat{\mathbf{a}}) > u_2(\mathbf{a}_0)$ .*

The strict-tradeoff condition entails that if  $\mathbf{a}$  is a Pareto-dominated allocation, then there exists a feasible allocation that strictly Pareto dominates  $\mathbf{a}$ ; in which case, we need only Case 1 of the proof of the modified Coase Theorem and we can dispense with the case that relied on no mean-spirited behavior:

**Theorem 3 (Modified Coase Theorem II)** *Consider a bilateral contracting situation in which (i) the parties are rational with respect to their individual self-interests; in which (ii) the parties can agree on any contract without incurring transaction costs; and in which (iii) the set of feasible allocations satisfies the strict-tradeoff condition. Then the allocation after contracting will be Pareto efficient regardless of the initial allocation.*

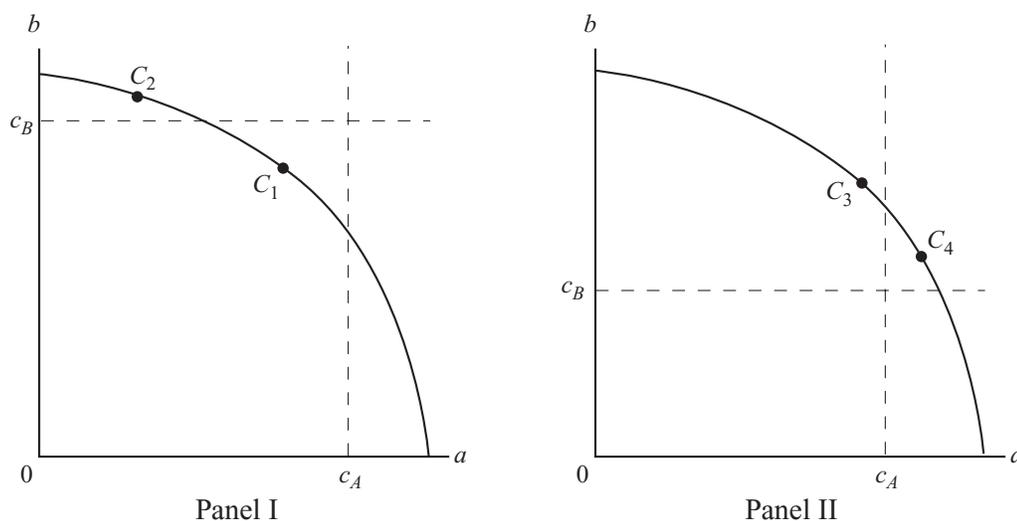
With respect to freedom of contract, we have

**Corollary 2** *Under the assumptions of the modified Coase Theorems, interference or restrictions on the contract the parties sign cannot increase the Pareto efficiency of the contracted-for outcome; that is, there should be freedom of contract if the only welfare issue is the efficiency of the outcome achieved by the contract from the perspective of the parties to the contract.*

Clearly, all the Coase theorems rely on the rationality assumption. We explore the consequences of relaxing this assumption later (see §2.3.4 and §4.2.1).

The no-transactions-cost assumption is also important. In the proofs, it serves to guarantee that the parties will not agree to a non-optimal contract because they can, without cost or impediment, choose an optimal contract instead. Relaxing this assumption would, thus, seem to have the potential for undermining the conclusions of these theorems—a point made by a number of authors (see, *e.g.*, Farrell, 1987b).

As Farrell notes, the economic literature on bargaining is generally sanguine about the prospects of an efficient outcome when the parties bargain under symmetric information. The literature is much more pessimistic, however, when



**Figure 2:** Consequence of transactions costs on bargaining outcomes under symmetric information.

they bargain under asymmetric information.<sup>10</sup> We discuss the consequences of bargaining under asymmetric information in §2.3.2. The rest of this section considers what can still potentially go wrong under symmetric information and concludes with Coase-theorem-like propositions that account for the potential of costly bargaining.

The potential consequences of transactions costs with symmetric information are illustrated in Figure 2. Suppose two parties,  $A$  and  $B$ , wish to enter into a contract. To do so, however, they must expend costs  $c_A$  and  $c_B$ , respectively. Once those costs are sunk (*e.g.*, lawyers are retained), the parties bargain over the contract terms. Let the payoffs, in money, be  $a$  and  $b$ , respectively. Because bargaining is conducted under symmetric information, theory predicts that the outcome will generally be efficient (see *e.g.*, Farrell, 1987b, or Sutton, 1986);<sup>11</sup> that is, the contract will be on the Pareto frontier, which is shown in the two panels of Figure 2 as a solid curve.

The first problem, illustrated in Panel I, is that  $c_A$  and  $c_B$  are so large that one or both sides prefer not to enter into negotiations. For instance, if bargaining would result in contract  $C_1$ , then neither side would be willing to negotiate; at  $C_1$ , we have  $a < c_A$  and  $b < c_B$ . Moreover, while there are contracts that, if adopted, would induce one side to participate (*e.g.*,  $C_2$ , for which  $b > c_B$ ), there

<sup>10</sup>See the survey on bargaining by Sutton (1986).

<sup>11</sup>Although efficiency is expected, it should be noted that one can construct perverse bargaining games that don't achieve efficiency even under symmetric information; see Hermalin and Katz (1993).

is no contract that would induce both to participate.

The problem illustrated in Panel I can serve to justify *default* contracts. That is, if the law stipulated that, absent a contract, the implicit contract was  $C_1$ , then that would clearly be an improvement. Observe that this improvement stems from two factors. First, the parties get to the Pareto frontier when they otherwise wouldn't. Second, they avoid the expenditures  $c_A$  and  $c_B$ . Note, however, that this analysis does *not* rely on  $C_1$  being *mandatory*; this is not an argument for mandatory contracts.

A related problem, and one that *could* justify mandatory contracts, is illustrated in Panel II. Suppose that bargaining would lead the parties to contract  $C_3$ . Because  $a < c_A$  at  $C_3$ , party  $A$  would refuse to enter into negotiations. This is undesirable, especially as there exist contracts, such as  $C_4$ , that, were they the outcome of bargaining, would induce both parties to negotiate. Now there is scope for limitations on contracts. If terms were limited, so that the only contracts that could be considered were on the arc segment between the dotted lines (*i.e.*, the segment containing  $C_4$ ), then both  $A$  and  $B$  would be willing to enter into negotiations and they would get an outcome on the Pareto frontier.

Of course, one may wonder how the law would know that  $C_1$  is a good default contract or to limit contracts to those in the neighborhood of  $C_4$ . If the law incurs costs arriving at default or mandatory contract terms, or lacks the information to design optimal contract, then it could be better to leave things in private hands. In other words, while it is true that restrictions on private contracts can possibly enhance efficiency when the private parties incur transactions costs, one must assess that observation in light of real-life limitations on what the legal system can do and the cost at which it can do it.

Before leaving the issue of transactions costs, it is worth considering how far we can go in relaxing the no-transactions-cost assumption and still establish a Coase-like case for freedom of contract. Our objective is to have a precise statement of the form *if bargaining leads the parties to a second-best efficient<sup>12</sup> contract and does so in a way that minimizes bargaining costs, then the legal system can do no better than to leave the private parties' choice of contract unrestricted.*

To establish such a result, it is helpful to switch from working with allocations to working directly with contracts. Let  $C$  denote an arbitrary contract. Contract terms are assumed not only to fix the allocation of real goods (possibly contingently), but also the allocation of transfers (possibly contingently) between the parties. Let  $U_i(C)$  denote party  $i$ 's utility (possibly expected) should the parties agree to contract  $C$ .

We can restate the strict tradeoff condition as

---

<sup>12</sup>Second-best efficiency refers to the optimal outcome taking into account the informational constraints faced by the parties. For instance, in the standard hidden-action agency problem, the first-best outcome would entail the agent expending some ideal level of effort for a flat wage. But that outcome is typically infeasible once the constraint that the agent's action is hidden from the principal is taken into account; the second-best solution in such cases typically requires the agent to bear risk, which is inefficient from a first-best perspective.

**Condition 2 (Strict Tradeoff)** *Let  $\mathcal{C}$  be the set of feasible contracts and consider any two contracts  $C_0$  and  $C_1$  in  $\mathcal{C}$ . If  $U_1(C_1) \geq U_1(C_0)$  and  $U_2(C_1) \geq U_2(C_0)$ , with at least one inequality holding strictly, then there exists a contract  $C_2 \in \mathcal{C}$  such that  $U_1(C_2) > U_1(C_0)$  and  $U_2(C_2) > U_2(C_0)$ .*

Given the Strict Tradeoff condition, the following Coase-like theorem can be established:

**Proposition 1 (Hermalin and Katz (1993))** *Suppose the two parties to a contract are symmetrically informed prior to and during bargaining and that bargaining consists of alternating offers. Assume the costs of delay in achieving an agreement are due to discounting. Assume that the set of possible contracts,  $\mathcal{C}$ , is invariant over rounds of bargaining and that the set  $\{(U_1(C), U_2(C)) | C \in \mathcal{C}\}$  is convex and compact. Assume that there is at least one  $C \in \mathcal{C}$  that each party strictly prefers to no agreement (the status quo). Finally assume either the Strict Tradeoff condition is satisfied or both parties are risk neutral and non-contingent (lump-sum) transfers in any amount are feasible. Then there is an essentially unique subgame-perfect equilibrium in which bargaining ends in the first round with agreement on a Pareto-efficient contract.*

The qualifier “essentially unique” captures the fact that there could be more than one contract that yields the unique equilibrium utility levels.

**Proof:** Follows from Lemma A2 and Proposition 4 of Hermalin and Katz (1993). ■

Because bargaining ends in the first round, there are no transactions costs on the equilibrium path. Hence, there is no possibility of increasing efficiency by reducing transactions costs. Moreover, because the resulting contract is efficient, there is no scope for increasing efficiency with respect to the contract chosen. In sum, under the assumptions of Proposition 1, there is no gain to be had from restricting private contracts; that is, Proposition 1 makes a case for freedom of contract.

On the other hand, Proposition 1 relies on a number of assumptions. Fortunately, some of these can be relaxed if one is willing to impose refinements on the subgame-perfect solution concept as applied to bargaining games. Specifically, we wish to rule out the possibility that the parties fear proposing efficient contracts because the equilibrium specifies a continuation game following the proposal of an efficient contract by party  $i$  that is unfavorable to party  $i$  (see Hermalin and Katz, 1993, for an example of such a perverse game). To that end, consider the following equilibrium refinements:

**Condition 3 (Monotone Acceptance Condition)** *Suppose that, at some date (round)  $t$ , a party would accept an offer of contract  $C_0$ . Suppose that another contract,  $C_1$ , would yield that party a higher expected utility level. Then, he or she will also accept contract  $C_1$  at date  $t$ .*

**Condition 4 (Stolen Thunder Condition)** *Suppose that the equilibrium entails one party's making an offer of contract  $C$  at date (round)  $t$  on the equilibrium path for some  $t > 1$ . Then he or she would accept an offer of  $C$  by the other party in round  $t - 1$ .*

As Hermalin and Katz (1993) discuss, both refinements seem reasonable for bargaining games of complete information.

Given these refinements one can establish a stronger result:

**Proposition 2 (Hermalin and Katz (1993))** *Suppose the two parties to a contract are symmetrically informed prior to and during bargaining and that bargaining consists of alternating offers. Assume the costs of delay in achieving an agreement are due to discounting or from per-round fixed costs. If the Strict Tradeoff, Monotone Acceptance, and Stolen Thunder conditions are satisfied, and there is at least one contract that each party strictly prefers to no agreement (the status quo), then bargaining ends with their agreeing to a Pareto-efficient contract in the first round of bargaining.*

**Proof:** This is Proposition 5 of Hermalin and Katz (1993). ■

Propositions 1 and 2 make the case that, when society is solely concerned with the wellbeing of the parties to the contract and those parties are symmetrically informed at the time of contracting, there is no reason to believe that restricting the parties' freedom of contract will improve efficiency.

## 2.3 The economic case against freedom of contract

So far, we have focused on the economic case *for* freedom of contract. Now we review the case against. Our discussion of the Coase Theorem suggests two potential grounds on which to argue against (complete) freedom of contract: (i) actors who are not party to a contract (third parties) are affected by externalities resulting from the contract; and (ii) problems in negotiating a contract prevent the parties from writing the optimal contract.

### 2.3.1 Third-party externalities

As we saw in §2.2.2, the efficiency of markets and private contracting is contingent on there being no third-party externalities. For instance, the market equilibrium with a competitive, but heavily polluting, industry does not maximize welfare—the supply of the good in question is determined by the private costs incurred by the manufacturers rather than the social costs that account for both those private costs and the harm the pollution imposes on society. Because social costs are greater than private costs, more than the welfare-maximizing quantity gets sold.

The inefficiency of the market when externalities are present can justify restrictions on private contracts. For instance, to deal with negative externalities, society does better by restricting the freedom of buyer and seller to set price;

there exists a price floor above the market-equilibrium price that forces trade to occur at the welfare-maximizing level.<sup>13</sup>

A strong believer in the Coase Theorem might object to this conclusion, arguing that polluters and their victims could contract to set the pollution level optimally. While that might be plausible in the context of a single polluter and a single victim (*e.g.*, noise pollution issues between neighbors), most situations of interest involve multiple polluters and millions of victims. It is difficult to imagine that significant expenditures of time and effort aren't required for a multitude of parties to reach an agreement on the terms of a contract. Moreover, as Farrell (1987b) notes, the unknown intensity of the parties' preferences typically means that any such bargaining would occur under asymmetric information. When such real-life transactions costs are accounted for, restrictions on contracts could be the more efficient means of solving the externality problem.

The transactions-cost issue is worse when the victims of the externality are not even known at the time parties enter into a contract. Aghion and Bolton (1987) nicely illustrate the problem: A monopolist (*e.g.*, a manufacturer), concerned about entry into its market, signs long-term exclusive-dealing contracts with buyers (*e.g.*, retailers). An entrant enters only if it is more efficient (has a lower marginal cost) than the incumbent monopolist. Whether such an entrant will exist is unknown *ex ante* by the incumbent monopolist and any given buyer;<sup>14</sup> but both know the distribution of the potential entrant's marginal cost. Because of the potential for entry, the buyer can expect to earn more surplus in the future than it currently does; as noted by Bork (1978) and R. Posner (1976), this will make the buyer reluctant to enter into an exclusive dealing contract with the seller. However, some of the surplus generated by the entry of a more efficient entrant is captured by the entrant itself. Hence, if the buyer and incumbent seller collude—that is, enter into an exclusive-dealing contract with a liquidated-damages provision that the buyer must pay the incumbent seller if it switches to the entrant—then buyer and incumbent seller can capture some of the entrant's surplus. The entrant must lower its price *vis-à-vis* the price it would have charged absent any liquidated damages provision by the amount of the liquidated damages to induce the buyer to switch to it. If the entrant lowers its price, then surplus is being transferred from it to the buyer-seller combination. The problem with such exclusive-dealing contracts is that the buyer-seller combination is like a non-discriminating monopolist,<sup>15</sup> it sets the liquidated damages provision too high—in the combination's desire to capture more surplus from the most efficient entrants, it deters entry from those that

---

<sup>13</sup>A price floor is not the only policy tool that could improve efficiency relative to the unfettered market. Other possibilities are an excise tax on the good, a pollution tax, or permitting the manufacturers some ability to cartelize their industry.

<sup>14</sup>Later in their article, Aghion and Bolton relax this assumption and assume the incumbent monopolist has superior information; the implications of that assumption will be addressed later in §2.3.2.

<sup>15</sup>For a discussion of the welfare issues connected to a non-discriminating monopolist, see §2.3.3 *infra*.

are only moderately more efficient. Consequently, prohibiting exclusive-dealing contracts increases expected welfare by increasing the probability of entry by a more efficient producer.

Observe that it is difficult to invoke the Coase Theorem in response to the Aghion and Bolton model. Because the entrant is unavailable at the time the incumbent seller and buyer contract, there is no possibility of them signing a three-way contract that achieves efficiency.<sup>16</sup>

Following Rasmusen et al. (1991), exclusive dealing can illustrate another externality problem. Consider an incumbent monopolist who sells to  $N$  buyers. Normalize the surplus that each buyer enjoys under monopoly pricing to zero. Assume that, if there were entry, each buyer would enjoy surplus  $s > 0$ . Assume that entry is feasible only if an entrant can attract at least  $\hat{N}$  buyers, where  $1 < \hat{N} < N$ . Observe, therefore, that if the incumbent can lock up at least  $N - \hat{N} + 1 \equiv N^*$  buyers through exclusive-dealing contracts, the incumbent blocks entry. Consider a point in time prior to the arrival of an entrant; and consider the following offer made simultaneously by the incumbent to each of the  $N$  buyers: In exchange for signing an exclusive-dealing, the incumbent provides the buyer surplus  $\varepsilon > 0$  (*e.g.*, the incumbent cuts its price by a small amount). The buyers respond independently and simultaneously to the incumbent. Take  $\varepsilon$  to be small enough that  $N\varepsilon$  is smaller than the amount the incumbent would stand to lose should entry occur; that is, the incumbent does better paying out  $N\varepsilon$  and keeping its monopoly than not paying that amount and facing competition. Note this will often entail  $\varepsilon < s$ .

**Proposition 3** *There is a Nash equilibrium in which all  $N$  buyers sign an exclusive dealing with the incumbent.*

**Proof:** If a given buyer believes that the other  $N - 1$  buyers will sign, then that buyer believes entry has been blocked (recall  $\hat{N} > 1$ ). Hence, that buyer expects to get 0 if she doesn't sign and  $\varepsilon$  if she does. Because  $\varepsilon > 0$ , it is, thus, a best response to sign. ■

The equilibrium of Proposition 3 is undesirable insofar as social welfare is reduced by the deadweight loss resulting from the preservation of monopoly pricing. Limitations on freedom of contract (*i.e.*, a prohibition on exclusive-dealing contracts) would be welfare enhancing.

As, however, Rasmusen et al. note, the equilibrium of Proposition 3 is not unique if  $\varepsilon < s$ . Another Nash equilibrium is for all buyers to refuse the contract—if no other buyer will sign, then signing would mean forgoing  $s$  in exchange for  $\varepsilon$ . Moreover, as Segal and Whinston (2000) point out, if one allows the buyers to form “coalitions,” then the only Nash equilibrium will be the

<sup>16</sup>There is also a further problem insofar as the entrant's cost is its private information, so there is an asymmetry of information problem that would impede efficient contracting even if the entrant were known *ex ante*. We elaborate on this point in §2.3.2.

- If you, a buyer, sign and
1. if fewer than  $N^*$  buyers (including you) sign in total, then you will receive  $\varepsilon$ , but you will be *released* from the exclusive-dealing provision (*i.e.*, you can buy from the incumbent or the entrant); alternatively,
  2. if exactly  $N^*$  buyers (including you) sign in total, then you receive  $s + \varepsilon$  and you must buy from the incumbent; alternatively,
  3. if more than  $N^*$  buyers (including you) sign in total, then you receive  $\varepsilon$  and you must buy from the incumbent.

**Figure 3:** A Clever Contract

one in which all buyers refuse the contract.<sup>17</sup>

Unfortunately, as Segal and Whinston go on to show, there is still scope for entry-detering exclusive-dealing contracts. For instance, it is possible that  $N^*s$  is smaller than the amount the incumbent would stand to lose should entry occur. Hence, there exists an  $\eta > 0$  such that  $N^*(s + \eta)$  is smaller than the amount that entry would cost the incumbent. Then the incumbent could offer just  $N^*$  buyers an exclusive dealing contract in which each buyer received  $s + \eta$ . Because  $s + \eta$  is greater than what a buyer receives even if entry occurs, it is a dominant strategy for each of these  $N^*$  buyers to accept this contract.

Indeed, building on ideas in Dal Bó (2003), it is possible for the incumbent to induce all the buyers to sign an exclusive-dealing contract at a cost to itself that is arbitrarily close to zero. Consider the contract in Figure 3.

**Proposition 4** *If the incumbent offers the Figure 3 contract to all buyers, then it is a dominant strategy for each buyer to accept the contract. Hence, the unique Nash equilibrium is for all buyers to accept the contract, which entails the incumbent's paying a total of  $N\varepsilon$  to block entry.*

**Proof:** Consider a given buyer and let  $n$  be the number of other buyers it expects to sign (so  $N - n - 1$  is the number of other buyers it expects not to sign). There are three cases to consider:

1.  $n < N^* - 1$ . If the given buyer signs, she will be released from the exclusive dealing and her total surplus will be  $s + \varepsilon$ . If she doesn't sign, it will be just  $s$ . Hence, she should sign.

<sup>17</sup>The word coalition in this context has a specific game-theoretic meaning; roughly a coalition in this context refers to a self-enforcing agreement. That is, here, an agreement to reject the incumbent's offer would be self-enforcing because it is a Nash equilibrium for all buyers to reject. For the situations below in which the incumbent offers a contract to only  $N^*$  buyers or the incumbent uses a "clever contract," then an agreement among the buyers not to accept would *not* be self-enforcing—because it is a dominant strategy for some or all buyers to accept, they would not honor their agreement with their fellow buyers; that is, those contracts are robust to the formation of "coalitions." We elaborate on this point below.

2.  $n = N^* - 1$  (*i.e.*, the given buyer is pivotal). If the given buyer signs, she will be obligated to buy from the incumbent, but she will be paid  $s + \varepsilon$  in surplus. If she doesn't sign, then entry will occur, but her surplus will be just  $s$ . Hence, she should sign.
3.  $n \geq N^*$ . If the given buyer signs, she will be obligated to buy from the incumbent, but she will be paid  $\varepsilon$  in surplus. If she doesn't sign, then, because entry is blocked, she will get no surplus. Hence, she should sign.

Because, as shown, signing is best regardless of what the buyer thinks  $n$  is, signing is a dominant strategy. Because the given buyer was arbitrary, this holds for all buyers; that is, all buyers will sign. Because more than  $N^*$  buyers sign, the incumbent blocks entry, and does so at a cost of only  $N\varepsilon$ . ■

Although we have presented the inefficiencies illustrated by Propositions 3 and 4 in the context of exclusive dealing, they are, in fact, examples of a broader phenomenon. Segal (1999b) considers the general issue of a single contract proposer,  $P$ , who can enter into a number of bilateral contracts with other actors,  $A_1, \dots, A_N$ , and derives conditions under which a set of unconstrained bilateral contracts will fail to maximize welfare due to externalities.<sup>18</sup> Other contexts include vertical relations (*e.g.*, exclusion of retailers or resale price maintenance), takeover battles ( $P$  is a raider and  $A_1, \dots, A_N$  incumbent shareholders), debt workouts ( $P$  is “equity,” which offers a debt-equity swap to the creditors, the  $A$ s), and network externalities ( $P$  sells a network good and the  $A$ s purchase it).

Again, a strong believer in the Coase Theorem might object to these conclusions on two grounds. First, a single grand contract among all the participants would achieve efficiency if there is no asymmetry of information. Second, a *binding* agreement among the  $N$  actors (*e.g.*, the retailers) in advance of bilateral contracting with  $P$  (*e.g.*, the incumbent monopolist) would ameliorate, if not eliminate, the problem. There are, however, a number of counter-objections:

- If  $N$  is large, then the transactions costs are likely to be so large as to make contract restrictions more efficient.
- Some of the  $N$  actors could be unknown or not yet exist (*i.e.*, similar to the problem in Aghion and Bolton).
- Contracting among the  $A$ s could generate other concerns. For instance, there could be legitimate antitrust concerns if all the retailers of a good or in an area were allowed to write a contract among themselves.

### 2.3.2 Asymmetric information

As has been known since Akerlof's (1970) seminal work, asymmetric information between parties can result in market distortions. A number of authors

---

<sup>18</sup>For the latest on the general theory of bilateral contracting with externalities, see Segal and Whinston (2003).

(Aghion and Bolton, 1987; Aghion and Hermalin, 1990; Johnston, 1990; Spier, 1992; Hermalin, 2002, among others) have applied this idea to the issue of contract design; showing that asymmetric information between the parties at the time a contract is negotiated can lead to distortions in the resulting contract *vis-à-vis* the contract that would have been negotiated under symmetric information. Unless the equilibrium under the symmetric-information contract is, itself, second best, such distortions must imply a loss of welfare *vis-à-vis* the symmetric-information benchmark.

Whenever the parties negotiate imperfect contracts, the question arises whether there is scope for the legal system to improve matters, either by restricting the set of possible contracts *ex ante* or through appropriate court action *ex post*. Aghion and Hermalin (1990) explore the former possibility in the context of a signaling game.<sup>19,20</sup> Their analysis can be motivated as follows. A well-known restriction on debt contracts is that the contract cannot contain an enforceable waiver by the debtor to her right to declare bankruptcy. Indeed, many states in the US impose even further protections, allowing a bankrupt debtor to keep certain assets (*e.g.*, a car or house) under certain circumstances. Can such restrictions enhance efficiency or total welfare?

To be concrete, consider an entrepreneur who needs to raise capital for a project. She knows how likely it is that her project will succeed; that is, whether it is a good project, which has a probability of failing of  $F_g$  or a bad project, which has a probability of failing of  $F_b > F_g$ . Using a short-hand common to information economics, call the entrepreneur with a good (alt. bad) project the good-type (alt. bad-type) entrepreneur. While the entrepreneur knows her type, a potential investor does not. His knowledge is limited to knowing that there is a probability  $\theta \in (0, 1)$  that the project is good. Assume that if a project fails, it is impossible for the entrepreneur to repay the investor fully.

---

<sup>19</sup>Signaling games, first studied by Spence (1973), are games of asymmetric information in which the better informed party takes actions that have the potential to convey—“signal”—her information to the less well informed party. The classic example (Spence) is a worker who signals information about her ability to potential employers through the amount of education she acquires. An equilibrium of a signaling game is called *separating* if the equilibrium actions of the informed player vary with her information (*e.g.*, workers who know themselves to be more talented acquire more education than workers who know themselves to be less talented). A *pooling* equilibrium is one in which the equilibrium actions of the informed player do not vary with her information (*e.g.*, all workers get the same level of education).

<sup>20</sup>Section II.B.2 and Appendix C of Johnston (1990) also considers the implications of signaling on contract formation in the context of evaluating limited-liability rules. Unlike Aghion and Hermalin, Johnston is concerned with *default* rules rather than binding restrictions. However, as Ayres and Gertner (1992) argue, Johnston’s emphasis on default rules undermines his arguments; Ayres and Gertner show that the choice of *default* rule is irrelevant in a world with costless contracting.

A recent paper by Anderlini et al. (2003) is another contribution to this literature. They focus on *ex post* actions by the courts, specifically whether the court should void certain contracts in some states of the world. Unlike Aghion and Hermalin, who focus on how restrictions can shift the equilibrium from an inefficient separating equilibrium to a more efficient pooling equilibrium, Anderlini et al. show how the expectation of the court’s *ex post* actions creates the possibility of an efficient separating equilibrium when otherwise the equilibrium would be a less efficient pooling equilibrium.

Investing is, therefore, risky and the risk is greater investing in a bad project than in a good project. For this reason, the entrepreneur can get more generous terms from an investor, the more likely he thinks it is that her project will succeed. Consequently, the entrepreneur has an incentive to signal the investor that her project is good through the terms of the debt contract she offers the investor. Specifically, because the expected cost of a large payment to be paid if the project *fails* is greater for a bad-type entrepreneur than a good-type entrepreneur, a good-type entrepreneur can signal that she has a good project by promising a large payment to the investor should the project fail. The cost of signaling in this manner is that the entrepreneur exposes herself to considerable risk (*e.g.*, losing her house if the project fails).

Restricting the amount the entrepreneur can promise to repay in the case of failure can potentially generate Pareto superior outcomes. To see why, note that, because of the additional risk, an entrepreneur with a good project might prefer not to signal if silence were interpreted by the investor as meaning her project was “average” (*i.e.*, had a failure probability of  $\theta F_g + (1 - \theta) F_b$ ). The difficulty is that, under a reasonable solution concept for the game,<sup>21</sup> the investor will interpret silence as evidence that the project is *bad*; and given the choice between looking good (signaling) and looking bad (not signaling), an entrepreneur with a good project will prefer to look good. If, however, signaling is restricted (*e.g.*, bankruptcy laws limit what the entrepreneur can pay in the event of failure), then not signaling is no longer informative. The investor will, therefore, treat all entrepreneurs as if they have an average project. Both types of entrepreneur are better off—an entrepreneur with a bad project now looks average, while an entrepreneur with a good project avoids the additional risks imposed by costly signaling. Because the investor is always held to his reservation utility conditional on his equilibrium beliefs, he is no worse off. Thus, restricting the possible terms of the contract would be Pareto superior.

Aghion and Hermalin formalize this argument in the entrepreneur-investor context. They go on to suggest, but not model formally, that the idea of contract restrictions eliminating “wasteful” signaling is more general than the entrepreneur-investor example. For instance, it could justify limits on penalties for breach of contract: To signal that she is very likely to be able to deliver a product on time, a good-type supplier might offer “too high” a penalty to be paid were she to be late; where “too high” means that she would be happier with a lower promised penalty if only the buyer wouldn’t interpret that lower penalty as indicating she was the bad-type supplier. Barring excessive penalties would prevent the buyer from making that interpretation, which in turn would lead to a contract that both good and bad-type suppliers preferred.

While Aghion and Hermalin prove that restrictions on contracts *can* be welfare enhancing in the context of signaling models, they do not establish that restrictions will always be welfare enhancing. For some set of parameters, restrictions enhance welfare; for others, they don’t. Moreover, in the latter

---

<sup>21</sup>Specifically, a Perfect Bayesian Equilibrium (PBE) satisfying the Intuitive Criterion of Cho and Kreps (1987).

case, the imposition of binding restrictions will reduce welfare (see Figure 5b of Aghion and Hermalin and connected discussion). Intuitively, there exist parameter values such that the separating PBE under asymmetric information replicates the equilibrium that would hold under symmetric information. In those situations, given our earlier discussion, it is not surprising that restrictions can only reduce, not enhance, welfare. Furthermore, if the contracting situation is already problematic, the fact that the informed player must signal can improve matters: Aghion and Bolton (1987), for instance, point out that the introduction of asymmetric information in their model pushes down the average liquidated damages penalty, thereby increasing the likelihood of efficient entry.

The lack of a clear normative conclusion from Aghion and Hermalin admittedly limits the practical application of their results. Eric Posner (2003a) sees this as a fatal flaw, supporting his overall indictment of contract theory as a guide to the law of contracts. But, as Craswell (2003) notes, there is no reason to require economic analysis to reach “all or nothing” conclusions before the analysis is useful normatively.

A signaling game is a particular kind of game of asymmetric information; in the context of contract design, it corresponds to a situation in which the informed party (*e.g.*, the entrepreneur) makes a contract offer to the uninformed party (*e.g.*, the potential investor). But there are other possible contract-offer games. An obvious alternative is for the uninformed party to make a contract offer to the informed party (a game known as *screening*). Can restrictions on contracts improve the efficiency of the outcomes of screening games?

The answer, as demonstrated by Hermalin and Katz (1993), is effectively no. The argument is as follows, let  $\mathcal{C}_U$  be the set of unrestricted contracts and let  $\mathcal{C}_R$  be the subset of restricted contracts (*i.e.*,  $\mathcal{C}_R \subset \mathcal{C}_U$ ). Let  $C_i^*$  be the contract offered by the uninformed party and accepted by the informed party in the game where the relevant contract space is  $\mathcal{C}_i$ .<sup>22</sup> Let  $v_P(C)$  be the uninformed party’s expected utility under contract  $C$ . Because the uninformed party cannot signal information, changing the contract space cannot change the informed party’s acceptance rule. Hence, by the nature of optimization,

$$v_P(C_U^*) \geq v_P(C_R^*). \quad (2)$$

Clearly, if the inequality is strict, then restrictions on contracts cannot be Pareto improving. If expression (2) is an equality, then there is no reason for the uninformed party not to offer  $C_R^*$  if it Pareto dominates  $C_U^*$  and, hence, it is unclear why we shouldn’t expect  $C_R^*$  to be offered even absent restrictions. Moreover, if the Strict Tradeoff condition (Condition 2) holds, then  $C_R^*$  cannot Pareto dominate  $C_U^*$  if expression (2) is an equality. To see this, were  $C_U^*$  dominated by  $C_R^*$ , then, by the Strict Tradeoff condition, there would exist

<sup>22</sup>As is typical in contract theory, we can, without loss of generality, add the “refusal contract” to the set  $\mathcal{C}_R$  (and, thus, to  $\mathcal{C}_U$ ); where the refusal contract stipulates the same payoffs to the parties as would result if the informed player refused the uninformed player’s offer. In other words, there is no loss of generality in assuming acceptance of some contract in equilibrium.

a third contract  $\hat{C} \in \mathcal{C}_U$  that the uninformed player strictly preferred to  $C_U^*$  and which the informed player would accept. But the existence of such a  $\hat{C}$  contradicts the optimality of  $C_U^*$ . Therefore,  $C_R^*$  cannot Pareto dominate  $C_U^*$ . A robust conclusion, therefore, is

**Proposition 5** *Restrictions on contracts cannot be Pareto improving in screening situations if either (i) the uninformed party is not mean spirited; or (ii) the contract space satisfies the Strict Tradeoff condition (Condition 2).*

Why do signaling and screening models yield different conclusions? Externalities offer an explanation. Although the informed player in a signaling game is a single entity, one can nonetheless view her as being two (*e.g.*, the bad-type entrepreneur or the good-type entrepreneur). The fact that a bad type is potentially willing to pretend to be a good type forces the good type to distort the contract she offers so that she won't be mistaken for the bad type (*i.e.*, select a contract, though not ideal for her, that the bad type would *not* be willing to offer). The problem is that there is no way for the bad type to internalize this externality that her potential mimicry imposes on the good type. When, however, it is the *uninformed* party who makes the contract offer, he is in a position to internalize the costs and benefits of attempting to differentiate the different types of the informed player. In essence, signaling can impose an externality, while screening cannot.

We can also consider social welfare in screening models of contract bargaining. Obviously, if restrictions are Pareto improving they also enhance welfare. However, as we've just seen, it will generally be the case that restrictions will not be Pareto improving with screening. Hence, we limit attention to the case in which expression (2) is a strict inequality. The question of whether restrictions can enhance welfare then boils down to whether the informed party's (average) gain from the restrictions exceeds  $v_P(C_U^*) - v_P(C_R^*) > 0$ .

To see that restrictions can enhance welfare, consider the following example. The uninformed player is a seller and the informed player is a potential buyer. The buyer's private information is his knowledge of the benefit,  $b$ , he derives from a single unit of the good being sold by the seller. Assume that  $b$  is drawn prior to contracting from a uniform distribution on  $[0, 1]$  and this fact is common knowledge. For convenience, assume the good holds no intrinsic value for the seller, so her cost is zero. While the buyer knows his benefit from purchase at the time of contracting, the seller knows only it was drawn from the uniform distribution. The unconstrained contract offered by the seller will be a contract to sell the good at a price of  $1/2$ , which yields total expected welfare of  $3/8$  (of this  $1/4$  is expected profit for the seller and  $1/8$  is the expected surplus captured by the buyer).<sup>23</sup> Consider the restriction that  $p \leq 0$ . Within this constrained

<sup>23</sup>At price  $p$ , the buyer will buy if  $b \geq p$ . The probability  $b \geq p$  is  $1 - p$ . The seller's expected profit from a price  $p$  is, thus,  $p(1 - p)$ , which is maximized by  $p = 1/2$ . If the unit is traded, welfare is just  $b$ ; hence, expected welfare is  $\int_{1/2}^1 b db = 1/2 - 1/8 = 3/8$ .

space, the seller will set a price of  $p = 0$ . All types of buyer buy, so expected welfare is  $1/2$  (all of which is captured by the buyer).<sup>24</sup>

Observe that the welfare loss that arises in an unrestricted world occurs because of asymmetric information. If the seller knew the buyer's valuation, then the seller would set  $p = b$ . All types of buyer would buy, so expected welfare would be  $1/2$  (all of which would be captured by the seller). Alternatively, if the buyer did not know his valuation at the time of contracting—so the parties are symmetrically informed insofar as they both know only that  $b \sim U[0, 1]$ —then welfare would be maximized by a contract that set  $p = 1/2$  and the buyer always bought; the seller gets  $1/2$  for sure and the buyer's expected surplus is zero, so total welfare is  $1/2$ . Note this last result is consistent with the view that asymmetries of information that arise *after* contracting are *not* justifications for restrictions; a point made more generally by Hermalin and Katz (1993)—see Propositions 1 and 2 above.

In §2.2.1, we observed that the welfare criteria of Pareto efficiency and social welfare often coincide when transfers between the parties are feasible. It is worth, therefore, considering why that coincidence breaks down with asymmetric information. When there is asymmetric information, transfers are called upon to serve double duty. They continue to be a means of transferring surplus so that the welfare-maximizing allocation might be viewed as Pareto efficient by the parties. But, with asymmetric information, they are also a means of screening the different types. As our simple example illustrates, this second duty impedes transfers from doing the first optimally.

To summarize: In a signaling situation, restrictions on contracts can lead to Pareto superior outcomes and, thus, can increase total welfare. In contrast, in a screening situation, restrictions on contracts generally cannot be expected to generate Pareto improvements, although they can increase total welfare.

### 2.3.3 Market power

As discussed above, competitive markets can be expected to maximize welfare in the absence of externalities. When, however, one or more entities have *market power*, the market can no longer be expected to yield the social welfare-maximizing allocation. It is well known that a firm with market power (*i.e.*, that faces a downward sloping firm-specific demand) produces less than the welfare-maximizing quantity. So, at least in the standard static framework, market power reduces welfare.<sup>25</sup> Consequently, public policy should generally oppose contracts that promote market power over competition, such as cartel agreements. This logic extends to other contracts, such as exclusive-dealing con-

<sup>24</sup>Expected welfare is  $\int_0^1 b db = 1/2$ .

<sup>25</sup>In a dynamic framework, one may need to consider the incentive effects of monopoly profits for innovation; that is, in some contexts, without the monopoly profits that intellectual property protection (*e.g.*, patents) afford, the innovator would lack sufficient incentive to innovate. No innovation means zero units are traded, which yields even less welfare than the monopoly outcome.

tracts,<sup>26</sup> that firms might sign to maintain, establish, or extend market power.

It is important to recognize that the welfare loss that comes from market power (*i.e.*, non-discriminating or simple monopoly pricing) is an example of the adverse welfare consequences of asymmetric information in a screening model (see §2.3.2).<sup>27</sup> While buyers know their valuations for each unit, the monopolist does not. Hence, the monopolist sets her price both to affect the transfer of surplus from buyers to herself and to screen the buyers. As before, asking a single instrument to serve two roles leads to distortions in welfare.

The fact that the monopoly-pricing problem is a screening problem also implies that if the monopolist knew each buyer's valuation schedule for all the units he could conceivably wish to purchase, then it could achieve the welfare-maximizing allocation. That is, as is well known (see, *e.g.*, Tirole, 1988, §3.1), a perfect price-discriminating monopolist maximizes social welfare.<sup>28</sup> Hence, the welfare loss typically seen with monopoly stems not from market power alone, but from the combination of market power and asymmetric information.

Price discrimination in the real world is imperfect. Depending on the form of discrimination, the structure of demand, and other circumstances, allowing a monopolist to engage in price discrimination versus simple monopoly pricing can enhance or diminish welfare (see Chapter 3 of Tirole, 1988, for a review of the welfare consequences of imperfect price discrimination). Consequently, it is difficult to assess, at a general level, what policy should be towards the enforceability of contractual terms that facilitate price discrimination.

For example, it is not obvious from a welfare perspective whether airlines should be free to issue tickets that contain restrictions (*e.g.*, an obligation to stay a Saturday night before returning to one's point of origin).<sup>29</sup> Such matters need to be studied on a case-by-case basis.

Note too that market power is connected to bargaining power. Indeed, the screening problem associated with a non-discriminating monopolist can be interpreted as stemming, in part, from the seller having all the bargaining power,

---

<sup>26</sup>Recall the discussion of Aghion and Bolton (1987) and Rasmusen et al. (1991) in §2.3.1.

<sup>27</sup>It follows from the Coase Theorems (Theorems 1–3) or Propositions 1 & 2, as appropriate, that bargaining power should *generally* be irrelevant in a standard welfare analysis *absent* asymmetries of information.

This is not to say that there couldn't be other social reasons for concern about inequities in bargaining power, such as distributional concerns (consider, *e.g.*, Kennedy, 1982). As E. Posner (2003a) notes, courts have been known to declare contract unconscionable because of unequal bargaining power. See also §2.4.1 and §2.5.2 *infra*.

<sup>28</sup>Of course, because a perfect-discriminating monopolist captures 100% of the surplus, there could be distributional or equity grounds for this outcome not to be favored.

<sup>29</sup>Saturday night restrictions are a form of second-degree price discrimination whereby an airline can screen business travelers (those with a high value of flying and a high cost of staying over Saturday) from non-business travelers (those with lower values of flying and lower costs of staying over Saturday). If banning such restrictions caused the airlines to price so that far fewer non-business travelers flew, then the ban would almost surely be welfare reducing. If, however, the ban led the airlines to price so as to keep non-business travelers flying, then eliminating the distortionary effects of the Saturday-night restriction would almost surely be welfare enhancing.

so that she gets to make a take-it-or-leave-it offer to buyers. The welfare loss from monopoly would disappear if it were the buyers who could make take-it-or-leave-it offers (assuming they knew the seller's marginal-cost schedule).

### 2.3.4 Capacity and bounded rationality

The parties in traditional law & economics analyses are presumed to be sophisticated and possess the requisite capacity. Consequently, any contract into which they voluntarily enter is, in rational expectation, superior for them to their no-contract (*status quo*) position. That is, each party correctly estimates that their expected utility from contracting exceeds that from not.

Observe that the rationality being assumed has two components. First, neither party would enter into an agreement that he or she thought would make him or her worse off, in expectation, than not contracting. Second, each party is forming these expectations in an objectively correct manner. For instance, if you respond to some get-rich-quick spam email, you presumably expect to enrich yourself, but such expectations are not rational; that is, you are rational in the first sense, but not the second.

It is difficult to argue that people aren't rational in the first sense (putting aside certain pathologies such as compulsive self-destructive behavior), but there has long been some unease with the assumption that people are rational in the second sense (see, *e.g.*, Simon, 1972, or Rubinstein, 1998, for a discussion). More recently, a movement has arisen within law & economics generally (see, *e.g.*, Jolls et al., 1998; Korobkin and Ulen, 2000) that also questions the assumption of rationality in the second sense. This movement has been labeled "behavioral law and economics," and it resembles related work in economics in its emphasis on cognitive errors, framing effects, time-inconsistent patterns of discounting, and similar phenomena. Within the formal modeling of contracts, however, assumptions of *bounded rationality* have been largely limited to the issue of explaining incomplete contracts (or justifying assuming that contracts are incomplete). We take up this use of bounded rationality in §4.2.1 below.

While, to the best of our knowledge, capacity and sophistication have not been formally modeled in the context of contract theory, such issues have received attention by law & economics scholars in law reviews (see, *e.g.*, Eisenberg, 1995; Korobkin and Ulen, 2000; Bar-Gill, 2005).<sup>30</sup> Korobkin and Ulen, for instance, argue that mandated contractual terms can be justified when one side lacks sophistication. They take as an example the former practice of insurance companies to cover only one day of hospital stay for maternity and the legislative reaction that required these companies to cover longer stays. They argue that legislative action was necessary because "the possible permutations of coverages that could, in principle, be provided by a given health insurance policy are numerous and . . . can overwhelm even sophisticated consumers . . . [and their] agents . . . who might be making purchasing decisions" (Korobkin and

---

<sup>30</sup>Although, as we discuss *infra* some formal models (*e.g.*, Katz, 1990b; Rasmusen, 2001) touch on related issues.

Ulen, p. 1082).

While it is difficult to argue against the proposition that the majority of consumers do not fully understand the provisions of their insurance contracts,<sup>31</sup> the example of maternity benefits seems a poor one from which to argue for intervention on the basis of inattention. The legislative reaction to this issue, and the strong public outcry that prompted it, suggest that the limited coverage insurers had previously been offering was not due to any consumer inattention. A more likely explanation was insufficient willingness to pay.<sup>32</sup> Put a bit differently, given the apparent awareness of the market to how much maternity stay was being covered, one would have expected the “unregulated market” to provide the desired level of maternity benefits if maternity benefits had been important to those who actually decided about policies.

A more plausible variant of the Korobkin and Ulen idea is developed by Eisenberg (1995). Although Eisenberg presents his argument verbally rather than mathematically, it is helpful to add some degree of formalism.<sup>33</sup> Suppose there are two possible future states, a rare one, which occurs with probability  $r$ ; and a more likely one, which occurs with probability  $1-r$ . Eisenberg is interested in the case where  $r$  is small, but not zero. Consider contracting between two parties,  $P$  and  $A$ . For convenience, assume that the bargaining between them always results in a contract that yields monetary benefits  $B_i$  (gross of direct transfers) to party  $i$  in the more expected state (*i.e.*, the state that occurs with probability  $1-r$ ). Suppose that the contract can contain a stipulation,  $s \in \mathcal{S}$ , that governs what happens in the rare state. Let  $b_i(s)$  denote party  $i$ 's monetary benefit gross of direct transfers in the rare state under stipulation  $s$ . Assume that the stipulation that maximizes  $P$ 's benefit in the rare state,  $\hat{s}$ , is not the stipulation that would maximize the sum of the parties' benefits. Let

---

<sup>31</sup>See, for example, Eisenberg's (1995) discussion (p. 242) of *Gerhardt v. Continental Insurance Cos.* (48 N.J. 291, 295 A.2d 328), in which he quotes the justices of the New Jersey Supreme Court expressing their difficulties in understanding the terms of the insurance contract in question.

<sup>32</sup>Korobkin and Ulen cite statistics (their footnote 111) that 90% of insured Americans get their health insurance through their employer. This fact, however, is potentially problematic for their argument. The employers could be quite sophisticated and are simply shopping for a bargain in insurance. That is, the problem is due to agency (the employers obtaining insurance on their workers behalf) rather than bounded rationality. Korobkin and Ulen might argue that workers suffer the same cognitive limitations bargaining with their employers; but, depending on worker-employer bargaining, the relative political power of workers versus employers and insurers, and the incidence of the cost of increased medical coverage, it is nevertheless possible that everything can be explained within the rational-actor paradigm.

<sup>33</sup>Katz (1990b) and Rasmusen (2001) are two articles that also model “reading costs” formally.

that stipulation be  $s^*$ . Mathematically,

$$\begin{aligned} b_P(\hat{s}) &\geq b_P(s) \text{ for all } s \in \mathcal{S}, \\ b_P(\hat{s}) &> b_P(s^*), \\ b_A(s^*) + b_P(s^*) &\geq b_A(s) + b_P(s) \text{ for all } s \in \mathcal{S}, \text{ and} \\ b_A(s^*) + b_P(s^*) &> b_A(\hat{s}) + b_P(\hat{s}). \end{aligned} \tag{3}$$

Observe that there must be  $s \in \mathcal{S}$ , including  $s^*$ , that  $A$  prefers to  $\hat{s}$ . Define  $\Delta_i = b_i(s^*) - b_i(\hat{s})$ . By construction,

$$\Delta_A > \Delta_A + \Delta_P > 0 > \Delta_P.$$

Eisenberg is interested, in part, in situations in which  $P$  proposes the stipulation for the rare state and it costs  $A$  some amount,  $k$ , to evaluate or understand the stipulation that  $P$  has proposed. One can view  $k$  as the cost of consulting an attorney or the cost of time spent trying to understand the small print. If  $A$  is particularly unsophisticated, one might even set  $k = \infty$  to reflect the idea that  $A$  is simply incapable of understanding what  $P$  has proposed. Suppose that if  $A$  does understand the stipulation, then  $P$  and  $A$  would bargain to some term that  $A$  found acceptable. In keeping with our earlier analysis of bargaining under symmetric information, we may as well assume that they would then agree on  $s^*$ . A critical assumption, however, is that

$$k > r\Delta_A; \tag{4}$$

in other words, the expected gain that  $A$  stands to reap from understanding the stipulation for what should happen in the rare state is smaller than the cost of obtaining that understanding (we will see in a moment that, when  $A$  doesn't understand the stipulation in the rare state,  $P$  should propose  $\hat{s}$ ). Observe that expression (4) holds if  $k$  is large or if  $r$  or  $\Delta_A$  are small; Eisenberg can be read as being primarily concerned about the situation in which  $r$  is small—we will return to this point later.

To close the model, suppose that the bargaining process is such that  $P$  can be seen as proposing a contract subject to  $A$ 's acceptance, where  $A$  accepts only if his expected net benefit at least meets a reservation level  $\beta_A$ ; that is,

$$(1-r)B_A + r\tilde{b}_A - p \geq \beta_A, \tag{5}$$

where  $\tilde{b}_A$  is the benefit  $A$  *expects* in the rare state and  $p$  is a transfer to  $P$  (if  $p < 0$ , then the transfer is from  $P$  to  $A$ ).  $P$ 's utility is

$$(1-r)B_P + r\tilde{b}_P + p,$$

where  $\tilde{b}_P$  is what she gets in the rare state.

Suppose that  $A$  will not expend the  $k$  necessary to understand the stipulation for the rare state; then it is irrelevant for  $A$ 's acceptance rule (5) what stipulation  $P$  *actually* makes because  $A$  won't understand it and, thus, his expectation,

$\tilde{b}_A$ , cannot depend on it. Given this, it is clearly a best response for  $P$  to choose  $\hat{s}$ .<sup>34</sup> In equilibrium,  $A$  must form his expectations correctly,<sup>35</sup> so  $\tilde{b}_A = b_A(\hat{s})$ . Observe, given (4), it is indeed a best response for  $A$  not to expend  $k$  on understanding even if  $A$  anticipates  $P$  is proposing  $\hat{s}$ . In equilibrium, total welfare will, thus, be

$$\widehat{W} \equiv (1-r)(B_A + B_P) + r(b_A(\hat{s}) + b_P(\hat{s})).$$

From expression (3), it follows that total welfare is *not* being maximized.

If we suppose that  $P$  has sufficient bargaining power so that expression (5) is binding, then  $P$ 's equilibrium utility is

$$(1-r)(B_A + B_P) + r(b_A(\hat{s}) + b_P(\hat{s})) - \beta_A,$$

from which we see that it is  $P$  who bears the full cost of this loss in welfare. Normally when a party bears the full cost, that party would take steps to eliminate the efficiency loss. However, here,  $P$  can't—there is no credible way for her to commit to  $A$  that she has stipulated  $s^*$ ; that is, any promise of  $s^*$  is cheap talk because  $P$  knows  $A$  will not understand if  $P$  substitutes the stipulation that favors her.

There are some ironies in this result. First, while Korobkin and Ulen and Eisenberg appear to suggest that it will be the *unsophisticated* party who suffers in these situations, the truth could well be that it is the *sophisticated* party who suffers. Second, even though, as noted by Schwartz and Wilde (1979), the sophisticated party has a motive to internalize the value of the superior contract term, there is no way for her to do so.<sup>36</sup>

Conversely, suppose that, due to competition among the  $P$ s, it is the  $A$ s who capture the surplus; that is,

$$(1-r)B_P + rb_P(s) + p = K_P,$$

where  $K_P$  is  $P$ 's reservation utility (cost). Observe this expression can be rearranged as

$$p(s) \equiv K_P - (1-r)B_P - rb_P(s). \quad (6)$$

From expression (6), it follows that the price a  $P$  charges if the term is  $\hat{s}$  is lower than if it is  $s^*$ ; that is,  $p(\hat{s}) < p(s^*)$ . In this environment, we can assess

<sup>34</sup>One might wonder why  $P$  bothers to worry about what to stipulate if the state really is rare. One answer, suggested by Eisenberg, is that  $P$  deals with many  $A$ s (*e.g.*, the contract is a form contract that  $P$  uses with many customers); hence, even though the expected gain to  $P$  from stipulating  $\hat{s}$  over  $s^*$  is small for any one transaction, in aggregate it is large enough to induce  $P$  to invest in choosing the best stipulation for her.

<sup>35</sup>This imposition of rationality on  $A$  might seem at odds with the bounded rationality approach of Eisenberg. We will discuss this point later.

<sup>36</sup>In Katz (1990b), the sophisticated party can take actions costly to it that lower the unsophisticated party's cost of understanding the contract terms. On the other hand, if those costs are incurred on a per-transaction basis (*e.g.*, they represent the expense of a person explaining terms to a potential buyer), then such actions could well fail to be cost effective, especially if the terms in question pertain to a rare event.

Schwartz and Wilde’s claim that if some of the *As* have a low enough cost of evaluating the stipulation for the rare state, then competition will lead the *Ps* to stipulate  $s^*$  rather than  $\hat{s}$ . Suppose there are two kinds of *As*, those with low evaluation costs,  $k_\ell$ , and those with high evaluation costs,  $k_h$ . For the moment, set  $k_\ell = 0$ , so the  $\ell$ -type *As* always evaluate. For the  $h$ -type *As*,  $k_h$  satisfies expression (4). Suppose the presence of the  $\ell$  types led all *Ps* to stipulate  $s^*$ . Competition among the *Ps* would then require them each to charge  $p(s^*)$  and earn zero profits. Consider a deviation from this candidate equilibrium in which a *P* stipulated  $\hat{s}$ . Although she would lose all business from  $\ell$  types, she would, at least in the short term, keep the  $h$  types. Moreover, at a price of  $p(s^*)$ , she would realize an expected gain of  $-r\Delta_P > 0$  from a contract with an  $h$  type. Hence, deviating from the candidate equilibrium—stipulating  $\hat{s}$  rather than  $s^*$ —is profitable for that *P*, which means the candidate equilibrium isn’t an equilibrium at all. In other words, there is no equilibrium in which all *Ps* stipulate  $s^*$  for sure.<sup>37</sup>

In fact, if we set  $k_\ell > 0$ , we can see a second reason why an  $s^*$ -only equilibrium cannot arise. Define  $\tilde{\Delta}_A = b_A(s^*) - \bar{b}_A$ ; that is,  $\tilde{\Delta}_A$  is the expected gain for *A* that he expects from negotiating the contract from the expected stipulation in the rare state to  $s^*$ . Observe, even for an  $\ell$  type, it is only worth evaluating the stipulation if

$$k_\ell \leq r\tilde{\Delta}_A.$$

Suppose all *Ps* were expected to stipulate  $s^*$ ; then,  $\tilde{\Delta}_A = 0$ . Hence, no types of *A* would bother to evaluate the stipulation. But, as we’ve seen, if *A* won’t evaluate, then *P* should always stipulate  $\hat{s}$ . The conclusion is, thus, that it is impossible for *A* to ever be certain that *P* has stipulated  $s^*$  unless *A* checks (we might dub this the “Eisenberg uncertainty principle”).<sup>38</sup>

Although this model is somewhat rudimentary, it does demonstrate that the problem identified by Eisenberg and others is unlikely to be eliminated by private action. This suggests that intervention could be beneficial. Such intervention could be *ex ante*, *i.e.*, the law could disallow any stipulation but  $s^*$ ; or the intervention could be *ex post*, *i.e.*, the courts could grant *A* damages should it be shown that the stipulated term was other than  $s^*$ .

As we observed earlier, a critical condition is (4), which makes it rational for *A* not to seek to understand the terms stipulated for the rare state. If condition (4) fails, as it will if the relevant terms ( $r$  or  $\Delta_A$ ) are large, then the argument for intervention collapses.<sup>39</sup> While either  $r$  or  $\Delta_A$  small enough is sufficient for condition (4) to hold, the situation in which it holds because  $\Delta_A$  is small

<sup>37</sup>What *is* the equilibrium in this game depends on what additional assumptions we make. If, for instance, we assumed that  $\ell$  types could, in negotiations, change the contract from  $\hat{s}$  to  $s^*$  in exchange for an increase in price from  $p(\hat{s})$  to  $p(s^*)$ , then an equilibrium can be constructed in which all *Ps* offer contracts at  $p(\hat{s})$ , but end up negotiating different terms with the  $\ell$ -type *As*.

<sup>38</sup>Katz’s Lemma 2 makes the same point.

<sup>39</sup>Although, even when (4) fails, there could still be cause to be concerned due to the Eisenberg Uncertainty Principle.

could be relatively uninteresting—presumably  $A$  will need to take some action to enforce restrictions on  $s$  (*e.g.*, sue to have an illegal contract voided or to recover damages); but if  $\Delta_A$  is small, then  $A$ 's motive to do so could be small as well. That is, it could be difficult to implement intervention in cases in which  $\Delta_A$  is small. The situations in which  $r$  is small are, as Eisenberg suggests, the more important ones. The *ex post* injustice could be large, but the motive of a single  $A$  to address it *ex ante* could be too small to make feasible a private solution.

Admittedly, we have imbued  $A$  in our formal model with rather more rationality than Eisenberg intended. Like others (*e.g.*, Jolls et al., 1998; Korobkin and Ulen, 2000; Bar-Gill, 2005), he observes that research in psychology and behavioral economics suggests that most people are poor at estimating probabilities correctly, especially probabilities of small events.<sup>40</sup> To the extent that they systematically underestimate the probability of a rare event (*e.g.*, likelihood of an insurance claim), they will use too small an  $r$  in deciding whether to invest in understanding what the contract stipulates for rare states. While the addition of such cognitive biases have the potential to enrich the analysis, our simple model demonstrates that they are not essential to obtain the main conclusion.

Katz (1990b), among others, notes there is a relation between the analysis in this subsection and Akerlof's lemons model. Although, here, we have a problem of moral hazard—the contract proposer *chooses* her “type”—whereas Akerlof is concerned with a problem of adverse selection—the informed party's type is *exogenously* determined—it is true in both that the uninformed party expects adverse terms of trade. A major difference, however, is that here all valuable exchanges take place, but just not under the optimal contract. In the lemons model, in contrast, buyer suspicion about quality can push the price below the level at which high-quality sellers are willing to sell, leading to inefficient exit of the high-quality sellers from the market. Admittedly, if the terms in question were important enough to the uninformed party, but yet not important enough to incur the reading costs, then it is theoretically possible that no trade could occur because of the uninformed's suspicions about the adverse nature of the contract. These issues have been discussed informally in the legal literature as well (see, for example, Kennedy, 1982; Craswell, 1993, 2000; and Eisenberg, 1995).

## 2.4 Other arguments for regulating private contracts

### 2.4.1 Distributional fairness

The Second Theorem of Welfare Economics holds that given convex production and utility functions, complete markets, and costless redistribution, any Pareto efficient allocation can be implemented as the general equilibrium of a competitive economy (see, *e.g.*, §6.4 of Debreu, 1959). In the real world, however, these

---

<sup>40</sup>See Rabin (1998) for a survey.

necessary assumptions are frequently violated, so that the same policy instruments must be used to promote both efficiency and distributional justice, with the two goals being traded off in a second-best fashion (Okun, 1975). Accordingly, much economic regulation has in practice some distributional component (R. Posner, 1971, and Polinsky, 1974).

In the field of contracts, however, as opposed to other bodies of law such as tort or property, it is relatively difficult to use private law rules as instruments of distribution. This is because contractual obligations are in general voluntarily undertaken, so that changes in legal rules will be accompanied by changes in the parties' reservation prices for exchange, tending to shift the incidence of a regulation to its intended beneficiary (*e.g.*, Buchanan, 1970). But restrictions on price will redistribute between marginal and inframarginal market actors, and in cases of market power, can redistribute between the two sides of the market (as can readily be seen from the example of a monopolist subject to a price ceiling set between the monopoly price and marginal cost). And restrictions on non-price terms can similarly redistribute between marginal and inframarginal actors if they differ in the value they attach to such terms (Spence, 1975; Craswell, 1991).

Similarly, policies that internalize externalities have distributional effects as well, and from a political economy viewpoint such effects—because they are larger—are probably more important than any efficiency gains in explaining the pattern of regulation. For example, in situations of adverse selection (the most important externality arising in the contractual setting), policies that force information revelation or limit entry or exit from a market can cause redistribution across informational types.

#### 2.4.2 Liberty and autonomy

In the context of contract law, the liberal perspective tends to support arguments for contractual freedom, although on libertarian rather than economic grounds. As indicated in §1.2.3, however, liberal and economic theories of contract can diverge in cases of market failure or transaction costs. A canonical illustration is provided by the phenomenon of standard-form contracts. Standardized contracts are an inevitable by-product of a mass production economy, but many legal commentators (*e.g.*, Kessler, 1943; Rakoff, 1983) have regarded them with distrust on the grounds that most persons presented with standard forms quite reasonably do not bother to familiarize themselves with the specific contents, relying instead on the drafter's reputation and on the knowledge that other contracting parties regularly do business on like terms. Such commentators take the position that these facts vitiate the non-drafting party's consent and justify state regulation of fine-print terms, although other writers, more influenced by an economic or commercial perspective (*e.g.*, Llewellyn, 1960, pp. 362–371) are willing to read implied consent into the overall situation. This issue is discussed at greater length in §3 and §4 below, where we take up the problems of contract formation and interpretation.

### 2.4.3 Inalienability and commodification

Finally, both popular morality and legal institutions commonly limit transactions dealing with matters thought to be fundamental to citizenship or personal identity; common examples include prohibitions on slavery, sexual prostitution, and the transfer of political rights such as suffrage or military service. The entitlements subject to such restriction are often described in political terms as inalienable, a concept not easily incorporated into economic accounts of exchange. Sometimes, as in the case of sexuality or body parts, restrictions are imposed on market exchange but not on other transfers: a situation that Radin (1987) labels “market-inalienability.”

In some cases, restrictions on alienability can be justified in terms of market failure such as asymmetric information (*e.g.*, the use of in-kind benefits to screen out those not in need, as in Blackorby and Donaldson, 1988) or externality (when Pigouvian taxes are administratively infeasible). But many such restrictions are better explained by the idea that exchange of the relevant entitlement injures some fundamental interest of the restricted agent not captured by his utility function, or some social interest that cannot be translated into material or pecuniary terms. In economics, this idea finds historical roots in the classical Marxian idea of commodification, or the change in the social meaning of a good that occurs when it becomes the subject of economic exchange.

The concept of commodification can be translated into neoclassical economic terms by interpreting it as sort of a cultural externality (*e.g.*, when sexual prostitution is said to diminish the quality of other people’s relationships) or as a lexicographical preference ordering on the social welfare function (so that equality in the distribution of certain goods trumps other allocative or distributional considerations, as in Tobin, 1970). But this interpretation does not do justice to the sense in which those who advocate the concept are concerned with the social construction of perceptions and preferences. Accordingly, the concept is best explained (see, *e.g.*, Kelman, 1981) as relating to potential changes in the content of individual utility functions or of the cumulative social welfare function.

## 2.5 Legal doctrines regulating freedom of contract

To a considerable extent, the actual legal doctrines restricting freedom of contract can be understood in economic terms, in that most doctrinal restrictions roughly correspond to situations of market failure or high transaction costs, of the sort discussed in the previous subsections. This correspondence is not exact, however; and a survey of the main restrictions will illustrate the roughness of the relationship.

### 2.5.1 Formal requirements for contracting

The law imposes a number of formal requirements that must be satisfied in order for contractual agreements to be enforceable in court. In general, as

discussed in §2.1.1 above, failure to satisfy these requirements results merely in denial of public enforcement rather than any negative sanction. Under US law, for instance, a contract for the sale of goods with market value above \$500 (a threshold that has not been adjusted for fifty years) is not ordinarily enforceable in the absence of a writing.<sup>41</sup> Nonetheless, there are no penalties for entering into an oral contract of this sort and in fact such bargains are regularly concluded and performed without incident.

In many other situations, though, denial of public enforcement operates as a binding sanction, especially in one-shot deals where the potential gain or loss from breach of contract is high compared to the value of future business relations. For instance, few businesses would use oral contracts to govern the production and sale of custom-made machinery; and fewer lawyers would advise their use. In such cases, formal requirements can have important regulatory consequences; and their costs can deter exchanges from taking place (as when a business decides not to go public because of the burden of complying with the SEC's disclosure and auditing requirements).

Most legal formalities, like many default terms of the sort discussed above in §2.1.3 are justified, at least in part, by the need for coordination or by administrative needs of the legal system. The rules of offer and acceptance, for instance, discussed below in §3, are typically interpreted as social conventions that serve to help contracting parties ensure that they attach similar meaning to their words and actions and that this meaning will be understood by third parties interested in the agreement. Similarly, the law encourages parties to structure their arrangements and to create and present evidentiary records in a way that lowers the *ex post* costs of contract enforcement—both because of the economies of scale involved in setting up any particular administrative regime, and because it is judged infeasible or politically undesirable to tailor court fees to the individual costs of dispute resolution.

Many formalities, however, also serve regulatory purposes. For example, the requirement that litigation documents be submitted on standard size paper (a formality imposed by virtually all court systems) serves purely administrative purposes, but the requirement that a minor's application for a marriage license be signed by a parent or guardian is also intended to help prevent the minor from making a hasty and irrevocable decision. And some formalities, such as the doctrinal rule that silence in the face of an offer does not ordinarily constitute acceptance, serve both coordinating and regulatory functions.<sup>42</sup> Such a rule indeed comports with standard interpretive conventions, but it also serves to discourage parties from sending purely opportunistic offers in the hopes that the recipient will somehow overlook them.

When viewed as restrictions on contractual freedom, formalities are plausibly justifiable on at least three of the theories of market failure discussed in

---

<sup>41</sup>UCC 2-201.

<sup>42</sup>See RESTATEMENT (SECOND) OF CONTRACTS 69 (1979). In exceptional circumstances, however, such an inference is justified, as when there has been a course of dealing that leads the offeror to expect a response.

§2.3 above. First, formalities can operate to correct informational problems. Parties entering into contracts may not always realize that they are doing so, especially if they are amateurs or newcomers to the relevant commercial community; similarly, they may not understand all of the specific obligations entailed by contracting. Formalities that warn such parties against assuming unintended or unwanted obligations can thus prevent inefficient exchanges as well as undesired distributional transfers from the uninformed to the informed. (On the other hand, as many anti-formalist critics have argued, hidden formalities can also mislead naïve agents into thinking they have entered into binding legal obligations when in reality they have not.)

Second, formalities may provide an effective response to bounded rationality if their presence triggers some cognitive or institutional process that operates as a safeguard against the specific dysfunctional behavior at issue. For example, the federal regulation requiring a “cooling-off” period before completion of a consumer door-to-door sale allows time for additional reflection at a remove from any pressure imposed by the salesperson; and in an organizational setting, writing requirements allow principals more easily to know when an obligation has been undertaken and to monitor their agents for making bad bargains.

Third, formalities can help to correct various negative externalities, especially those that are moderate in magnitude or related to informational asymmetry, by attaching an additional cost to the externality-producing behavior. In markets affected by adverse selection, for instance, formalities that make it harder to disclaim warranties or opt out of medical coverage may keep some high-quality agents from exiting the market, supporting a higher quantity of exchange. Additionally, uncertainty regarding the existence and enforceability of contracts may in some settings adversely affect third parties whose economic fortunes are linked with the contracting agents. Such externalities may explain the traditional common-law rule requiring a writing for contracts for the sale of land or in consideration of marriage—both transactions with potentially significant implications for other members of the contracting parties’ family and community. Finally, formalities can help deter rent-seeking by opportunistic agents hoping to take advantage of the informational and behavioral problems discussed just above.

To illustrate these possibilities, we briefly discuss two of the more important formalities imposed by the common law of contracts: the doctrine of consideration, and the statutory requirement that certain contracts be executed in the form of a writing.

**THE DOCTRINE OF CONSIDERATION.** In the US and in other legal regimes that descend from the English common law, contractual promises are not enforceable at law unless supported by what lawyers call consideration. The precise meaning of this concept has long been debated, but its overall import is to require the contract to relate to an exchange that is understood as such by both parties. A promise to make a future gift is not enforceable under the law of contracts, for instance, although it may be possible to achieve a similar result by using some

other legal device such as an equitable trust. While the consideration doctrine has at times been viewed as a substantive limit on contractual freedom, modern commentators (following Holmes, 1897, and Fuller, 1941) view it largely as matter of form, in that it can often (though not always) be avoided by the use of some other formal device such as a special writing, the transfer of a nominal sum, or delivery of a symbolic token. As such, it raises the relative cost of entering into those contracts to which it applies.

Although its grounding in exchange would seem to suggest a close connection between the consideration doctrine and the promotion of economic welfare, the doctrine would seem to diverge from simple efficiency in at least two respects. First, many non-exchange promises also enhance economic welfare. A donor's commitment to make a gift, for example, enables the beneficiary to engage in specific anticipatory investment, thus lowering the donor's cost of providing the beneficiary with any given level of utility (R. Posner, 1977, and Shavell, 1991). Second, the lawyer's understanding of what counts as an exchange is narrower in practice than an economist's would be. In traditional common law, for instance, a promise to guarantee the debt of another person was not enforceable unless the party seeking to enforce could show that the promisor received a specific promise or benefit in exchange for the guaranty; it did not suffice to observe that commercial parties do not typically make such guaranties unless they stand to gain from the extension of credit and are hoping to induce it. Similarly, certain open-ended promises, such as promises to buy or sell goods in a quantity to be specified by the promisee, were traditionally deemed unsupported by consideration, on the theory that the promisee retained discretion to fix terms that would eliminate all benefits received by the promisor.

The doctrine of consideration has long been controversial; and civil law systems, such as those in continental Europe, do not employ it. In its application to donative promises, the doctrine is probably best justified as a response either to bounded rationality or to *ex ante* rent seeking by potential beneficiaries. It is at least plausible that impulsive promises are a greater problem when motivated by generosity rather than self-interest, and requiring more elaborate formalities before such promises become enforceable may be justified in order to protect the interests of the donor or of other potential beneficiaries of his largesse who might fare better upon more thorough deliberation. Similarly, there is an obvious incentive for overreaching by potential beneficiaries, and requiring an additional formal protection such as an equitable trust, which imposes fiduciary duties on the trustee and in practice requires the participation of a legal professional to assure its effectiveness, probably serves to police such opportunism better than a mere writing requirement would.

**WRITING REQUIREMENTS.** While written documents are typically regarded as better guides to the intention of the parties than oral testimony or circumstantial evidence, the general common-law rule is that most contractual promises are enforceable without a writing, if they can be proven to a court's satisfaction. Certain categories of contracts, however, must be evidenced in writing to be

legally enforceable. In common-law systems, such contracts include contracts for the conveyance of interests in real property, contracts in consideration of a party's marriage, third-party guaranty contracts, contracts in which an executor agrees to pay the debts of an estate, contracts that cannot be completed within a year, and, as indicated above, contracts for the sale of goods above a monetary threshold (in the US, \$500). These various writing requirements are collectively referred to as the "Statute of Frauds" because they derive originally from a parliamentary statute of 1677 that was justified by its proponents as a safeguard against false claims of contractual agreement. Similar writing requirements have been extended by statute to a variety of other transactions, including contracts that designate property as collateral for a secured loan and contracts for the sale of any personal property with value over \$5000; and in some settings, such as option and guaranty contracts, a written promise suffices as a substitute for consideration.

These various requirements surely serve the purposes of *ex post* efficiency in the enforcement setting; for example, courts might well wish to screen out contractual disputes where the key evidence is not just oral but stale (the usual rationale for the one-year provision of the Statute of Frauds). Given that the interpretation of oral evidence is typically more uncertain, allowing high-value contracts to be alleged on purely oral evidence is particularly costly in terms of incentives for *ex post* rent seeking. And given the considerable private incentives for contracting parties to memorialize their agreement in writing even apart from the prospects of a dispute, the absence of a writing justifies, on Bayesian grounds, the inference that further proceedings are unlikely to be of much value.

Most of these requirements, however, also correspond to some standard type of market failure such as externality or bounded rationality. Guaranty contracts, secured lending, and contracts for the conveyance of land, for instance, all implicate the interests of third parties, and in former times, so did contracts in consideration of marriage. Requiring such transactions to be evidenced in writing reduces information costs for third parties trying to determine the status of assets in which they may also have a claim. Similarly, the expected costs of bounded rationality are especially great in high-value contracts, contracts likely to be entered into under circumstances of emotional stress, and contracts that reach far into the future. In such cases, the extra transaction costs of a formal writing may be justified because they deter impulsive or myopic decisions.

### 2.5.2 Substantive limitations on contracting

In addition to the practical restrictions imposed by formal requirements for contracting, some agreements are simply unenforceable as a matter of substance. Such limits on enforceability could serve to strike the entire contract, providing an affirmative defense to any liability for breach; or they could strike certain terms of the contract only, leaving the rest of the contract fully enforceable but with the offending terms replaced by default terms, or by terms the court deems fairer (see Craswell, 1993, for a discussion). The major defenses we consider here include fraud, lack of capacity, mistake, duress, undue influence,

unconscionability, and offense to public policy.

**FRAUD AND UNILATERAL MISTAKE.** All legal systems refuse to enforce contracts that are based on sufficiently incorrect or asymmetric information, at least in cases where the uninformed party is unaware of his informational disadvantage. Enforcement is especially disfavored when one party to the exchange has caused the other to become misinformed, for example by misrepresenting material facts relating to a proposed exchange. Under the common law, this disfavor is reflected in the defense of fraud.

The fraud defense is justified on efficiency grounds for two reasons. First, it deters inefficient exchanges that would not have taken place but for the fraud; and second, in cases where fraud takes the form of undertaking efforts to deceive others, it discourages rent-seeking. Determining what statements count as fraudulent, however, is not always easy. Vague or ambiguous statement may raise interpretation issues of the sort discussed *infra* in §4; and even statements that are literally true may be interpreted as making an implied fraudulent representation. For economic discussions of this issue, see Ayres and Klass (2005) and Craswell (2006).

Additionally, the defense is not limited to cases of affirmative deception: it can also be asserted when one party withholds critical information that the other reasonably expects to be disclosed. Such fraud by omission is controversial from a legal perspective because of the difficulty of determining when there is a duty to disclose, and from an economic perspective because the duty to disclose information will affect incentives to acquire it.

Similarly, the doctrine of unilateral mistake allows a party to escape a bargain if his assent was based on a significant misapprehension that the other party could have easily corrected (as when a contractor underbids a job because of a calculation error that is evident from the large discrepancy between the mistaken bid and all others received). Here the law takes the position that the small effort required to correct the mistake is justified by the dislocation imposed on the mistaken party and the significant chance that the bargain is inefficient.

In order for doctrines like fraud or unilateral mistake to be welfare-improving in practice, however, courts must have a reliable way to distinguish between asymmetric information on the one hand, and conscious risk-taking in the setting of incomplete but symmetric information on the other, else too many *ex ante* efficient transactions will be avoided. In this area, accordingly, there is potential for divergence between what the law provides and what would be most efficient from a second-best perspective. In general, the courts have been sensitive to such concerns and have limited the scope of the doctrines to fairly significant imbalances of information (and in the case of fraud, have imposed special requirements of proof and pleading).

**INCAPACITY.** When one of the parties to a contract is incapable of weighing costs and benefits, there is no longer any basis to presume that the contract

is efficient. Similarly, the law declines to enforce contracts entered into by children, intoxicated persons, and the mentally disabled; this result also deters rent seeking by those who would take advantage of the incompetent. In contrast to the situation of fraud and mistake, however, there are circumstances in which an incompetent person will be held to his bargain. One such circumstance is when the contract is for the sale of “necessities” (*e.g.*, clothing or shelter that are judged to be in the objective interest of the buyer). Another is when the competent party could not have known that the counterparty lacked capacity, although many courts will still supervise the bargain in such cases to ensure its substantive soundness.

The different treatment of the capacity and fraud defenses is justified by the differing costs of overcoming the specific transactional failure at issue. In the case of fraud, the failure can be overcome at no cost by abstaining from misrepresentation, and in the case of unilateral mistake, the failure can be overcome at low cost by correcting the mistake. In cases of incapacity, however, it may not be possible to correct the problem without losing the transaction entirely, so the law allows plainly efficient exchanges (or in the case where the incapacity is hidden, exchanges where it is plain that no rent-seeking took place.)

**MUTUAL MISTAKE.** In some cases, the law refuses to enforce bargains in which the parties’ *ex ante* beliefs were sufficiently different from the state of the world, even when those beliefs are the same and when neither party could easily have corrected the error. For example, in the celebrated case of *Sherwood v. Walker*, the court allowed the seller of a pregnant cow to void the deal on the grounds that, at the time of the bargain, both parties mistakenly thought that the cow was barren and set a price that corresponded to its slaughterhouse value. This doctrine, known as the defense of mutual mistake, is subject to much confusion and there is no canonical economic explanation of it. The legal confusion results because it is difficult in practice to know whether the parties were truly mistaken or just engaged in a hedging transaction. From the economic viewpoint, Posner and Rosenfield (1977) view the doctrine as substituting, in a situation of costly contracting, for a complete contingent contract in which it is efficient to call off the sale in the state of the world where the mistake is discovered. Ayres and Rasmusen (1993), on the other hand, argue that the mutual mistake doctrine, in contrast to unilateral mistake, undermines incentives for information acquisition, and conclude that it is dominated by a suitably limited unilateral mistake doctrine. Indeed, some legal commentators (*e.g.*, Chirelstein, 2006) suggest that in practice the mutual mistake doctrine is used when the court suspects fraud or unilateral mistake but cannot clearly make out the elements of those doctrines under the facts. The doctrine, accordingly, may afford a good opportunity for further theoretical investigation.

**DURESS AND RESTRICTIONS ON CONTRACTUAL MODIFICATIONS.** The doctrine of duress denies enforcement to contracts formed under conditions of unacceptable coercion; the standard hypothetical is a contract signed at gunpoint. The

key to the doctrine, of course, consists of which conditions are deemed unacceptable. Some courts and commentators have held that difficult economic circumstances can amount to duress; while others suggest that hard bargaining is permitted unless the beneficiary of the promise is responsible for the difficult situation (for example by having isolated the promisor from other potential contract partners) or owes some duty with regard to the difficult situation (for example, by having been granted a local monopoly). Whether a contract is void for duress, however, depends not just on external conditions but also on the content of the contract. For instance, a contract to salvage the cargo of a sinking ship at 10% of the value of the cargo might be enforced, while one that set a price of 90% of cargo value probably would not.

A similar functional problem arises in contracts that modify duties created by previous contracts, because many modifications are entered into under circumstances where one of the parties is dissatisfied with the original contract and is threatening to breach it. For example, in the case of *Alaska Packers v. Domenico*, the court held unenforceable a modification that provided a 66% increase in wages to a crew of sailors who threatened to strike on short notice in remote waters.

Much of the legal commentary on this doctrine justifies it on the norm of party autonomy, on the theory that a choice made under coercive circumstances is no choice at all. This justification is difficult to square with the economic viewpoint, however, unless one interprets it as claiming that the harsh circumstances somehow interfere with the exercise of rationality. But in most cases in which the doctrine is applied, rationality does not seem at issue—even in the gunpoint hypothetical, the victim acts rationally in choosing life over money. For this reason, some have argued (*e.g.*, Bar-Gill and Ben-Shahar, 2004) that the doctrine be made unavailable in circumstances where the coercive threat is credible—that is, where the party making it would actually find it in its interest to carry it out if the coerced party does not agree to the exchange in question.

A better economic justification of the doctrine is found in the phenomenon of rent seeking (*e.g.*, Tullock, 2005)—that we wish to discourage investments in coercion or against coercion. In the gunman hypothetical, the rent seeking is obvious, but as Cooter (1982) and Cooter et al. (1982) suggest, a similar argument can apply to ordinary hard bargaining in situations of bilateral monopoly. Even when an exchange is efficient, in the absence of a well-defined mechanism for dividing the gains from trade, the parties may destroy part of the surplus in attempting to influence its distribution.<sup>43</sup> Additionally, in dynamic settings, the prospect of earning such rents may lead to overinvestment or excess entry (*cf.* Mankiw and Whinston, 1986). Thus, by reducing the set of possible threats, the law can, in principle, limit the available rents and encourage a more efficient equilibrium.

The rent-seeking explanation is especially apposite in the case of contrac-

---

<sup>43</sup>Although when the bargaining is conducted under symmetric information it is difficult from a game-theoretic perspective to see why surplus would be destroyed (see Hermalin and Katz, 1993, and the discussion in §2.2.2).

tual modifications, because parties who make specific investments in an initial contract become vulnerable to holdup. In many circumstances, parties will be reluctant to make specific investments in settings in which contract renegotiation contract is possible; accordingly, it is, in principle, beneficial for them to commit not to renegotiate if they can credibly do so (Aghion et al., 1994).<sup>44</sup> The law of contracts has been unwilling to allow parties to rule out modification, however, perhaps because of concerns for imperfect information and bounded rationality (Jolls, 1997). It has, however, used various strategies over the years to limit modifications that appear motivated by opportunism. In particularly egregious cases, as Shavell (2006) points out, courts will void the attempted modification on grounds of duress or unconscionability. In less extreme situations, more flexible tools are available. Traditionally, such regulation operated under the formal rubric of the consideration doctrine, on the theory that a modification that unilaterally altered the contract in one party's favor was not a true exchange. The modern approach, however (*e.g.*, Hillman, 1979) requires the modification to pass a substantive test of fairness or good faith. For instance, a modification made in proportionate response to new circumstances, unanticipated at the time of contracting, would generally be enforced, while an outright attempt to rewrite the original terms of the bargain would not. Because it is difficult to identify objective criteria for which circumstances are unanticipated and what sort of adjustment would be proportionate, however, the standard of enforcement is uncertain, so that some incentives for rent seeking do remain.

**UNDUE INFLUENCE.** A fourth standard limitation on contractual freedom is undue influence. This defense applies when one party is deemed to have such influence over the other that the latter lacks the requisite autonomy for contracting. Classic examples include contracts between lawyer and client, physician and patient, caregiver and invalid, and, traditionally, husband and wife.

The undue influence doctrine overlaps in function with the other substantive doctrines listed above. It shares with incapacity the element of bounded rationality, in that the influential relationship may be thought to interfere with the vulnerable party's ability to exercise independent judgment (and as with incapacity, undue influence does not entirely bar the enforcement of a contract, but rather subjects its terms to *ex post* regulation on grounds of substantive rationality and distributional fairness). It also shares with fraud the element of asymmetric information (in that the relationship entails a reposition of trust), and with duress the element of rent seeking (in that assent may be given in response to an implicit threat to withdraw the relationship and its associated benefits).

Undue influence is less commonly invoked in the context of person-to-person contracts than the other doctrines discussed above, but it operates as a cornerstone of the law of fiduciary relationships, and thus of much law governing professional relationships and contracts between organizations and their governing agents. For example, the corporate law duty of loyalty, which subjects

---

<sup>44</sup>Shavell (2005) notes that there is no empirical evidence that parties are able to do so.

dealings between corporations and their officers or directors to special scrutiny, is usefully understood as an application of this basic contractual doctrine.

UNCONSCIONABILITY AND PUBLIC POLICY. Finally, the unconscionability doctrine serves as a sort of all-purpose limitation on contractual freedom applied *ex post* in cases where the court deems the parties' exchange to be sufficiently unfair in either substantive or procedural terms. Some commentators (*e.g.*, Epstein, 1975) have suggested that the doctrine can be justified in economic terms to the extent that it operates as a discretionary extension of the other major affirmative defenses in marginal cases. For instance, the doctrine has been used to protect consumers or unsophisticated small business people from contractual terms that imposed large *ex post* risks and that were buried in fine print or otherwise insufficiently disclosed, which is in some ways analogous to fraud or incapacity. (As noted in the earlier discussion of fraud and non-disclosure, it is not always easy to determine what kinds of statements or omissions should be treated as fraudulent.) The doctrine may also be partially justifiable on the economic grounds of externality (*e.g.*, E. Posner, 1995), to the extent that it prevents contractual creditors from driving their debtors into insolvency, and thus imposing financial obligations on the debtors' families or on the public. Similarly, when viewed in connection with a variety of legal rules that protect debtors in extreme situations, such as bankruptcy law and other statutory limitations on debt collection, the doctrine can be understood as responding to failures in the markets for credit and insurance.

These efficiency rationales cannot entirely account for the actual operation of the doctrine, which sometimes appears merely to reflect the courts' reluctance to hold a hapless party to a bad bargain, or a paternalist concern that a party has undertaken an excessively high risk. The doctrine's implications for freedom of contract, however, should not be overstated, despite its prominent treatment in academic discussions and introductory student texts. Its effect is limited by the requirement that it can only be applied by a judge as opposed to a jury, and by most judges' appreciation that, at least in commercial cases, parties have good reason to take risks and adequate opportunity to obtain insurance. As a practical restriction on the parties' ability to get their bargains enforced in court, accordingly, it is probably rather less important than either the costs of litigation or than the more general psychological tendency of both judges and juries to resolve ambiguous facts in favor of the party with whom they feel the strongest sympathy.

Similarly, courts sometimes decline to enforce contracts on the grounds of public policy, although this concept is less a specific doctrine than a label for a general practice of limiting contractual freedom on a variety of rationales. The main difference between unconscionability and public policy is that the former is based on unfairness, while the latter is based on the idea that enforcement runs counter to the general goals of the legal framework or judicial system, even though no specific statutory or common law rule contains an explicit ban on the contract in question. For example, contracts to engage in gambling or

in ostensibly immoral behavior were traditionally denied enforcement on such grounds, as were contracts in restraint of trade and contracts thought to promote litigation. A prominent modern case applying the concept is the Baby M case, in which the New Jersey Supreme Court refused to recognize a contract which purported to transfer parental rights in exchange for financial consideration. Because of the breadth of this category, probably the best way to summarize it is simply to say that while the principle of freedom of contract is generally respected by the legal system, it remains subject to a vague and implicit set of limits that operate at its margins, on analogy to the limits imposed by other fields such as tort and criminal law. A more extensive discussion of the concept can be found in Farnsworth (2004).

### 3 Formation of Contracts

In many standardized or spot market transactions, parties enter into contracts without any significant amount of bargaining. In more complicated situations, however, a contract typically follows an extended set of communications that can include offers, counter-offers, and other exchanges of information. Such communications are governed by an elaborate framework of legal rules that determine when, how, and on what terms contractual obligations are created.

Compared to other areas of contract law such as the rules governing remedies for breach, there has been relatively little discussion of this doctrinal subject from an economic point of view. In part, this neglect results from the fact that these doctrines are esoteric and little known outside the legal profession. Additionally, legal discussions on the topic tend to focus on the narrow problem of channeling the parties' communications into conventionally recognizable forms, rather than on more direct regulatory concerns.

From an economic viewpoint, however, the law of contract formation is relevant because it influences the outcome of exchange. By attaching consequences to the various acts and omissions that individual bargainers can choose from in a negotiation, legal rules affect the parties' incentives to make and to respond to offers, to exchange information, and to communicate with one another at all.

In this section, accordingly, we survey the various dimensions in which pre-contractual behavior can affect the efficiency of exchange, and the way in which the law of contract formation affects incentives along those dimensions.

#### 3.1 Pre-contractual behavior

Even before they meet, potential contracting parties make a number of economic decisions that influence the possible gains from future trade. Not all of these decisions, however, will necessarily be made optimally. Some decisions create positive or negative externalities, and others entail relational investments that are vulnerable to holdup. In this subsection, we identify and discuss three overlapping dimensions of pre-contractual behavior: searching for contractual

partners, investigating the value of exchange, and making pre-contractual investments.

### 3.1.1 Searching for contractual partners

In order for parties to be in position to enter into exchange, they must undertake the expense of searching for potential trading partners. Such expenses include advertising, correspondence, travel, and the parties' time. From the perspective of a social planner, one would want the parties to undertake such efforts up to the point where the marginal costs of additional search just outweigh its expected marginal value. While search can be modeled in various ways (see, *e.g.*, Diamond, 1987), under plausible assumptions, the value of an additional unit of search lessens as the quality of the bargain in hand increases. It follows that it is, in general, optimal for a party to search up to some cut-off level of satisfaction, and then stop. This cut-off level will depend on the cost of search, the distribution of information, the expected bargaining game to be played once search is completed, and so on.

The level of search that is privately profitable for an individual buyer or seller, however, is not necessarily the same as the level that would be socially optimal, for two reasons. First, in markets with bilateral search, each person's search efforts provide a positive externality that reduces the search costs of others. Second, to the extent that there are economic rents associated with trading (*i.e.*, if parties buy or sell at prices that diverge from their reservation prices), some amount of search is motivated by the desire to find a better distributional outcome. These effects work in opposite directions, so it is difficult to generalize about what public policies would be optimal in this regard, but it is possible in principle that enlightened regulation could improve social welfare.

For example, §3.3 below explains how various legal doctrines operate to promote the early formation of contractual liability. Such doctrines will thereby affect the level of search, but the direction of the effect is ambiguous. From the *ex post* perspective of parties who have already found each other, such rules should reduce the incentive for additional search, because a party who holds a binding commitment has less need to search; and a party who is bound faces reduced value from search because even if she finds a better bargain, she will still be liable for the first one. But from an *ex ante* perspective, parties may be more willing to undertake the cost of search if they are assured that any trading partners they find will stick with their deal. Which effect is more important depends on the details of the situation.

### 3.1.2 Acquiring and disclosing information about the value of exchange

In order to conduct exchange, the parties not only must find each other, but they must also determine whether trade is worthwhile. Both the acquisition and disclosure of such information can be costly, and the parties' willingness to incur these costs will depend on whether they can be recouped in subsequent

contract bargaining and performance.

As with informational investment more generally, imposing a duty to share information tends to undercut the incentive to acquire it, so there is a potential tradeoff between efficient production of information *ex ante* and efficient use of information *ex post*. In this regard, Kronman (1978a) argued that requiring commodity traders to disclose private information about market value would undercut their incentives to do market analyses and make forecasts; and Matthews and Postlethwaite (1985) showed how requiring manufacturers to disclose product quality could discourage product testing.

Whether information should be produced, of course, depends upon whether it is socially valuable, as opposed to simply having private redistributive value. For that reason, if information acquisition is, on balance, socially wasteful, it is best to impose a disclosure requirement in order to deter rent-seeking (or a tax on acquisition efforts or windfall profits if that is administratively cheaper).

Whether information is acquired or disclosed before exchange, however, will depend not just on regulatory requirements, but also on property rights, the relational nature of the information, and market structure. For example, in Shavell's (1994) model of information acquisition, a party with exclusive rights over a particular item of property has efficient incentives with regard to acquiring information that bears on the common value of that property, because he internalizes that common value when he sells or uses the property. One could imagine circumstances, however, in which a party without formal property rights had similarly good incentives because he held a monopoly position in bargaining and because the information increased his private value for the property, but not the owner's.

Disclosure may also be compelled by the structure of expectations in bargaining. For example, Grossman (1981) presents an influential model of product warranties in which buyers rationally assume the worst about any seller who does not disclose its private information. Sellers with relatively good information are thus led to disclose it in order to avoid the effect of buyer skepticism, and the result is a separating equilibrium in which the private information is revealed.<sup>45</sup> Note that disclosure induced by such game-theoretic incentives can discourage information acquisition every bit as much as disclosure required by legal regulation.

Even in cases where mandatory disclosure would not adversely affect the incentive to acquire information, it is still necessary to compare the benefits of disclosure against the costs. In many contexts, especially those involving ordinary consumers, the costs of disclosure will include communication costs—for example, the time that it takes buyers to read and process the information; or

---

<sup>45</sup>To be precise, information unravels to the point where the cost of credible disclosure equals the difference between the price charged by the pooled low-quality sellers and the price that the highest-quality member of this pool could obtain through disclosure. Observe the unraveling result depends on the disclosure being effective to convey information. See, *e.g.*, Fishman and Hagerty (2003), who show that if the fraction of customers who can understand a disclosure is too low, then mandatory disclosure benefits informed customers and harms the seller.

the potential cost of being distracted from other, more important information. These communication costs have rarely been incorporated into formal economic models, though Bebchuk and Shavell (1991) is an exception. Craswell (2006) provides an informal discussion of these costs.

Finally, it must be noted that in some instances *partial* disclosure of information is worse than no disclosure (Hermalin and Katz, in press). This is true when partial disclosure facilitates inefficient discrimination. For instance, disclosure of information correlated with on-the-job performance (*e.g.*, health status) could depress the wages of those disclosing the correlate in equilibrium, causing some talented individuals to exit the labor market.

### 3.1.3 Pre-contractual investment

The previous discussion emphasized that investigation of exchange value can operate as a sunk investment that is vulnerable to rent-seeking in later bargaining. The same is true of pre-contractual investments generally.

Much of the law of contracts is designed to protect investment incentives, and §4 and §5 below discuss the effect of interpretation rules and contract remedies on those incentives. But these rules come into play only after a contract has been formed. Some investments, however, must be undertaken before it is practical for contractual liability to attach. For example, search efforts inherently must precede contractual negotiation (else who would one negotiate with?), and negotiation must precede the formation of a contract. But both search and negotiation are costly, and their costs are at least in part relation-specific. Similarly, other investments can be delayed until negotiations are completed, but their value is reduced in doing so (Katz, 1996c). Suppliers of goods, for instance, can typically reduce their production costs by buying materials when prices are low or by doing advance work when business is slow; and buyers can increase their value from exchange by investing in complementary inputs. But if they wait until they are finished bargaining to begin preparations, many such opportunities will be lost.

It is not desirable to provide complete protection for pre-contractual investments, conversely, because such protection would lead to excessive reliance. As the parties negotiate, they may discover information that reveals that the intended exchange should not be pursued. If this happens, any relation-specific investments will be wasted. Optimally, the parties should take the risk of wasted investment into account before making them. The rules governing contract formation, accordingly, should ideally be designed to promote optimal reliance at the optimal time, balancing the benefits of productive investment against the costs of waste.<sup>46</sup>

---

<sup>46</sup>The mechanism for paying compensation for a taking (governmental exercise of its right of eminent domain) must strike a similar balance (Hermalin, 1995).

### 3.1.4 Strategic behavior in bargaining

In §3.1.2 above, we discussed incentives to produce and share information regarding the value of exchange. But information exchange is not the only dimension of efficiency in bargaining. Negotiating parties can dissipate resources through various types of rent-seeking: by excessive communications, by haggling and stalling, and by hard bidding that risks losing the bargain entirely. One way to deter such behavior is through substantive legal doctrines that limit the kinds of bargains that can be enforced, and thus lessen the temptation for overreaching; for example, Cooter (1982) justifies the doctrine of duress on such grounds. But another way is to create contract formation doctrines that impose the cost of lost bargains on parties who cause them through excessive rent-seeking.

In order to address the problem of excessively hard bargaining, however, it is necessary to develop a clearer understanding of incentives to engage in it. As we discussed in §2.2.2 above, in the *absence* of asymmetric information the parties should be able to reach an efficient bargaining outcome without outside intervention.

Symmetric information is likely the exception rather than rule, however. Complete symmetry assumes symmetry of knowledge about preferences, among other parameters, which is typically not true (see *e.g.*, Farrell, 1987b). If parties bargain under asymmetric information it is well-established (*e.g.*, Samuelson, 1985; Farrell, 1987b; Schweizer, 1988) that they will often fail to reach a first-best outcome. The theory of mechanism design implies that for any asymmetric information bargaining game, there is, in principle, some mechanism that maximizes the parties' expected bargaining surplus subject to their participation and truth-telling constraints; that is, achieves the second best.<sup>47</sup> But any particular bargaining game may not be the optimal mechanism, so the ultimate efficiency of the bargain could depend on limitations on bargaining imposed by the law on contract formation.

## 3.2 Avoiding miscommunication

Finally, before we move to a discussion of legal doctrine, we consider the important role played by formation rules in avoiding miscommunication. Negotiating parties need to know when bargaining has been completed and whether a binding obligation has been formed, since if they have inconsistent understandings in this regard, significant value can be wasted. If a party believes that he has entered into a binding obligation when in fact he has not, he may waste resources attempting to perform or may turn down other profitable opportunities. Conversely, if a party is unaware that a contract has been formed, she may fail to make adequate preparations or even incur multiple liability, resulting in breach of contract and associated losses. One of the most important functions of contract formation law, accordingly—and the function to which legal schol-

---

<sup>47</sup>See Laffont and Martimont (2002), Bolton and Dewatripont (2005), or Caillaud and Hermalin (2000b) for introductions to mechanism design.

ars have probably devoted the most attention—is to promote clear channels of communication so that the parties may know where they stand.

The law of contract formation pursues this task in two distinct ways. First, it establishes authoritative forms and terms of art that the parties can use when negotiating their agreements. Second, it allocates the risk of miscommunication so as to encourage parties to take optimal precautions to prevent and insure against misunderstandings, in the same way that the law of torts allocates liability for accidents to the least-cost avoider and least-cost insurer. For example, the doctrines of unilateral and mutual mistake, discussed in §2.5.2 above, illustrate how loss allocation rules can strengthen the parties' incentives to avoid misunderstandings and to allocate any remaining risk on their own.

This issue of clarity in communication is most usefully understood in terms of the problem of interpretation, which we discuss at length in §4 below. Determining whether a contract has been formed raises most of the same issues as determining the terms on which it has been formed. For example, one must identify the meaning of phrases that are used in negotiation, reconcile interpretative differences among the parties or between the parties and the court, establish default rules that apply when the parties have left their negotiations open or spoken ambiguously, and so on. We defer consideration of such issues until the next section. Keep in mind, however, the connection between the legal doctrines discussed here and their role in interpretation discussed in §4. For instance, as we discuss in §4.4.2 below, pre-contractual communications often bear on the interpretation of a contract, especially with regard to issues where its text is silent or unclear.

### 3.3 Legal doctrines addressing contract formation

The body of legal doctrine relating to contract formation is particularly elaborate, and we do not attempt to survey it here. Instead we discuss a number of key doctrines that help address the central economic aspects of pre-contractual behavior: disclosure, investment, and efficiency in negotiation.

#### 3.3.1 Offer and acceptance

The classic distinction between contract law and other sources of legal obligation is that contract is grounded in voluntary agreement. The fundamental principle of contract formation, it follows, is that there must be mutual assent in order to establish a binding contractual obligation. Assent refers not, however, to the parties' private mental states, which clearly cannot form the basis of a public system of enforcement, but to our inferences regarding those states as drawn from the parties' actions and statements.

To simplify this problem of inference, lawyers often say that to form an agreement there must be an offer (which evidences one party's assent) and also an acceptance (which evidences the counterparty's assent); and thus this body of doctrine is often referred to as the law of offer and acceptance. It should be recognized that this way of describing the matter is a formalism that allows

interpreters to focus on certain stereotypical features of negotiating a contract and abstract from others. In actual contexts it is often difficult or artificial to identify a particular communication that counts as an offer or an acceptance, and some doctrinal systems (such as the law of sales under UCC Article 2) dispense with the effort. In general, however, the traditional common law has followed a highly formalistic approach in this regard, with the result that the treatises and casebooks are filled with a variety of arcane and colorfully named mechanical rules. Such rules include, among others, the “mailbox rule” (*i.e.*, when parties communicate a distance, a contract is formed at the moment an acceptance is dispatched, not when it is received), the “mirror-image rule” (in order to constitute an acceptance, a responding communication must mirror the offer in all relevant respects, else it is deemed a rejection and counter-offer), and the “last-shot rule” (if parties exchange a series of non-matching offers followed by one of the parties commencing performance, a contract is deemed to be formed with the beginning of the performance constituting the acceptance, and with the terms of the contract supplied by the final offer outstanding prior to acceptance). Some of the formalistic rules have received substantial economic attention (see, *e.g.*, Baird and Weisberg, 1982, and Ben-Shahar, in press, on the last-shot and mirror-image rules), but many others have not.

Over the last half century, such formalistic rules have come under increasing criticism, and more recent developments tend to de-emphasize formality, making it easier to establish liability without complying with prevailing formal conventions. This movement away from formality has had its most important effect on communicative efficiency; and is discussed at greater length in §4.4.1 below. Similarly, and simultaneously, the trend has been for the courts to establish that liability is incurred earlier in the bargaining process (with the mailbox rule providing an early precursor in this regard). This trend is probably most important with regard to incentives for pre-contractual investment, and we take it up in the subsection immediately following this one.

In addition, however, the particular rules of offer and acceptance also influence the transactions costs of negotiation in potentially significant ways. As an illustration, consider the doctrine of silent acceptance (Katz, 1990a, 1993), under which an offeree must respond affirmatively in order to create a binding obligation; silence operates as an acceptance only in special circumstances. The rule can serve to conserve on message costs. Under it, two messages are required to form a contract, while only one is needed in negotiations that do not result in exchange. A converse rule under which an offeree must respond in order to avoid being bound, would require one communication for a contract and two for a rejection. In the usual context where the majority of offers are rejected and rejections and acceptances cost the same to transmit, the common-law rule is efficient. But in settings where acceptance is likely or where acceptances are costlier to transmit than rejections (for instance, because of a mirror image requirement), the rule is inefficient.

Additionally, the common law rule reduces the costs of rent-seeking through opportunistic offers. Under a silent acceptance regime, offerors will have an incentive to propose inefficient exchanges to offerees with high response costs, in

the hopes that the offeree will choose to accept a mildly unprofitable contract in preference to incurring the costs of sending rejections. Such incentives explain why negative option plans of the sort offered by mail-order book clubs and telephone companies typically provide buyers with substantial up-front benefits as an inducement to enter the plan.

Finally, based on an extensive survey of the case law, Craswell (1996b) argues that offer and acceptance doctrines are widely used by courts to promote efficient relational investment. In his view, courts are often in position to evaluate *ex post* whether reliance has been efficient, and in such cases, they can provide good incentives by finding a contract when a party relied reasonably in response to a pre-contractual communication, but finding no contract when the party relied either excessively or not at all. Whether Craswell's argument provides a normative justification for courts practices in this regard, however, as opposed to an accurate positive account of what they do in fact, is open to dispute. In the first place, his survey of cases suggest that the fact of reliance is decisive to a finding of liability only in relatively close cases, not in all cases, so the incentive effects of this practice may be limited. Additionally, in order for courts to employ this tool, they must be able to judge the efficiency of reliance in hindsight. The practical limitations of the legal process, the fact that reliance decisions frequently entail nonverifiable actions, and the cognitive biases associated with hindsight (Rachlinski, 1998) may call this assumption into question. Instead, courts may do better to follow the simpler approach of identifying and holding liable the party who is the least-cost avoider for wasted reliance. A model based on this possibility is presented in the following subsection.

### 3.3.2 Promissory estoppel and analogous doctrines

The general principle of estoppel was developed in the English equity courts as a corrective device to preclude the operation of legal rules that were thought to yield an unjust result in specific cases. The more specific doctrine of promissory estoppel gained influence in the 19th century as a way to soften the formalities of the consideration doctrine, but in the latter half of the 20th century in the United States, it became increasingly used to loosen the formal requirements of offer and acceptance.

The key element necessary to invoke promissory estoppel is relation-specific investment, which lawyers call reliance. Under it, a promisor is precluded from asserting a lack of mutual assent if she made a promise that reasonably can be expected to induce the promisee's reliance, which actually does induce such reliance, and which will result in injustice if not enforced. "Injustice" is admittedly subjective, but in effect it serves to protect reliance that most people would consider reasonable. Excessive or unreasonable reliance is not protected by the doctrine, although there are obvious difficulties in determining reasonableness in this setting. For example, in the case of *Drennan v. Star Paving*, a general contractor used a subcontractor's offer in calculating his bid on a construction job. After the general contractor won the construction contract, the subcontractor attempted to withdraw the original offer on grounds of miscalculation, arguing

that under prevailing rules of offer and acceptance, the offer was revocable until the moment of acceptance. The court held that the subcontractor's bid was non-retractable, notwithstanding that there was never any explicit promise to hold it open and that the general contractor could have explicitly bargained for a binding option.

The efficiency of this result, and of the estoppel doctrine generally, depends on whether such liability is necessary to protect investment incentives. Katz (1996c) offers a model of pre-contractual investment in which one party makes an offer and the counterparty chooses to rely; and either of these decisions can be taken over a period of time leading up to the deadline for ultimate performance. Excessively early reliance is inefficient in this model because there is too high a chance that the investment will be wasted; excessively late reliance is inefficient because the productive surplus from investment cannot be enjoyed. The main conclusion of the model is that estoppel liability is desirable if the relying party cannot protect the value of its reliance investment in post-reliance bargaining, but undesirable if it can. The underlying intuition is that if the offering party holds the *ex post* bargaining power, the relying party will refuse to invest unless protected by liability; while if the relying party holds the *ex post* bargaining power, the offering party will refuse to specify its needs for fear of being held liable in circumstances where it does not pay to complete the contract, and being vulnerable to rent-seeking through a last-minute counter-offer in circumstances where it does pay to complete the contract.

Katz's model focuses on the timing of reliance investment, but its underlying intuition also applies to situations in which the extent and type of reliance is in question. As a general matter, reliance investments might be protected through liability rules or by *ex post* bargaining; and in cases where bargaining power is sufficient to provide incentives, legal liability is superfluous and may even provide excessive protection. How these considerations play out in individual bargaining contexts, however, depends on the precise nature and sequence of play. For instance, Bebchuk and Ben-Shahar (2001) consider the possibility of legal rules under which courts can condition liability on the level of reliance investment, on its reasonableness, or on the reasonableness of positions taken in *ex post* bargaining, and show that such rules can be used to promote efficient bilateral incentives. They also discuss the effects of liability rules on the parties' incentives to enter into negotiations initially, and show that prospect of liability need not deter initial negotiation (although under some conditions, it might). The informational requirements of such rules are of course significant, which limits their practical applicability in many situations.

The idea of conditioning liability on the reasonableness of bargaining behavior, however, underlies a related legal doctrine: the duty to bargain in good faith. This doctrine is much less widely used than promissory estoppel; it is applicable when the parties enter into an agreement that is too indefinite for a court to enforce as written, but that could be enforced if the parties were to undertake an additional round of bargaining in which they filled in enough gaps. For example, in *Teachers Insurance & Annuity Co. v. Tribune Co.*, 670 F. Supp. 491 (SDNY 1987), the defendant refused to complete a complicated real

estate deal, following significant negotiations resulting in a commitment letter that referred to a “binding agreement,” after learning that interest rates had significantly dropped and that the deal would probably not qualify for favorable tax treatment. The court held that even though the defendant had reserved a further right of approval on the part of its board of directors, it was obliged to continue negotiations in good faith, which in the specific context meant that it could not condition its approval on a contingency (*i.e.*, favorable tax treatment) that was deliberately not included in the commitment letter, and could not withdraw from the deal simply because interest rate changes had rendered it unprofitable.

The precise contours of the duty to bargain in good faith are not entirely clear. For instance, it is difficult to predict whether a court would hold it bad faith for a party to insist that all open terms be resolved in its favor or to ask that some previously settled term be re-opened, on the grounds that the deal had turned sufficiently sour that a fair distribution of rents justified such an adjustment. (Cases applying the modern doctrine of modification, which also turns on a good faith standard, have often held such adjustments to be acceptable.) The good faith doctrine has been criticized for its uncertainty and apparent subjectiveness in application. Nonetheless, the bulk of commentators continue to approve of the doctrine as a means of protecting relational investments in negotiation. In *TIAA v. Tribune*, for instance, the parties had already incurred significant costs and, more importantly, a deal of such complexity could not have been concluded without sinking such costs. Absent some form of pre-contractual liability, it would be difficult to enter into complicated contracts without subjecting reliance investments to the risk of holdup.

Other legal systems have developed analogous rules to deal with this functional problem. For example, the civil law regimes of continental Europe do not make use of the doctrine of estoppel, but they do include an invigorated version of the duty to bargain in good faith, known as *culpa in contrahendo*. Under this doctrine, parties entering into contract negotiations are held to a mutual duty of care that is intended to protect the reasonable expectations with which they enter bargaining. This duty has traditionally been regarded by comparative lawyers as being more demanding than the implicit duties imposed by the doctrine of promissory estoppel, but one can argue that modern development in the common law have effectively brought the two systems closer together in this regard.

### 3.3.3 Duty to disclose

Finally, the common law also imposes an ill-defined duty to disclose especially important information when negotiating with contractual partners. The duty does not extend to all information a counterparty might find relevant; rather, it is limited to situations in which nondisclosure would be regarded as effectively equivalent to a representation that the information does not exist. For example, in the classic case of *Laidlaw v. Organ*, a trader with advance knowledge of a peace treaty that would shortly result in the lifting of naval blockade bought

up tobacco that skyrocketed in value when the news became public. When the seller sought to avoid the contract on the grounds of fraud, the court suggested that, under the circumstances, the parties were entitled to hold each other at arms' length and to retain the benefits of such information for their private use. On the other hand, a homeowner who sells a termite-infested house to a buyer who is known to be unaware of any problem may well be under a duty to disclose it, especially if the buyer would not be able to discover the infestation though the exercise of ordinary diligence.

The duty to disclose has received substantial attention in the scholarly literature; and it was recognized early on that nondisclosure might have effects on the fairness and efficiency of exchange. In the law and economics literature, it has been recognized, at least since Kronman (1978b), that imposing a duty to disclose could deter information acquisition. Though later scholarly efforts (see, *e.g.*, the discussion in §3.1.2 above) have refined economic understanding of the issue, doctrinal application of these efforts has been limited. Kronman argued that, at minimum, there should be a duty to disclose information that was acquired without effort or expenditure, on the grounds that such a duty would not reduce the availability of information and would improve the *ex post* efficiency of exchange, but even this suggestion has not made its way into the case law, which has tended in this area to focus more on the question of rights than on efficiency.<sup>48</sup> Still, the trend in legal decisions seems to run in the direction of increased disclosure requirements, especially in settings where the interests of consumers, workers, or small businesses are involved.

### 3.3.4 Overall assessment of the law of contract formation

The division between the law of contract formation and other parts of contract law is not clear-cut. For instance, some of the rules discussed above in §2.5 on freedom of contract could alternatively be interpreted and analyzed as formation rules. Take for example the Statute of Frauds (see §2.5.1). From a freedom-of-contract perspective, the Statute helps to deter boundedly rational parties from entering into contracts without adequate deliberation, and may help internalize costs that are imposed on the public legal system when parties enter into agreements without adequate evidentiary backing. But it also serves some of the purposes discussed above, such as the disclosure of information or the reduction of miscommunication. Similarly, the mistake doctrine allows inefficient contracts to be avoided, but it also encourages parties to share information that would prevent misunderstandings and the wasteful investments that can result from them.

Additionally, as elaborated in §5 below, the rules governing contract formation significantly interact with the rules governing the remedies for breach. For instance, as we will see, promissory estoppel may substitute for a formal offer and acceptance in the appropriate case, but the damages theoretically available

---

<sup>48</sup>While *complete* disclosure will generally enhance efficiency, *partial* disclosure can reduce welfare (Hermalin and Katz, in press).

to the disappointed promisee may be rather less than the damages that would be awarded if the proper formalities have been observed (although there is scholarly controversy regarding whether courts actually follow this theoretical distinction in practice). Again, these doctrinal complications are likely to provide fruitful opportunity for additional economic research in future years, as is the law of contract formation generally.

## 4 Interpretation of Contracts: Contractual Incompleteness

Offer and acceptance only begins the process of contractual exchange. In order for the transaction to be completed, the contract must be performed, and if performance is not forthcoming, enforced. Probably the most common source of contractual disputes is differences in interpretation, if only because the parties have limited incentive to pursue a dispute if they can foresee and agree upon its likely outcome. The problem of contract interpretation thus provides a central backdrop for the law of contracts, which contains many rules and principles that are designed to address it.

The legal issue of interpretation corresponds to the economic issue of contractual incompleteness—a topic that has been a central focus of research in microeconomic theory in recent years (*e.g.*, Grossman and Hart, 1986; Hart, 1987; Hart and Moore, 1999; Tirole, 1999). Contractual incompleteness captures the idea that real-life contracting can fail to produce contracts that are as precise and detailed as traditional—albeit possibly naïve—economic theory predicts. Economists typically attribute this failure to an informational asymmetry: the parties to the contract anticipate observing events that they might wish to be contingencies, but which cannot serve as contingencies because they are not verifiable (*i.e.*, observable by a third-party adjudicator of any contract dispute); in other words, the parties will be better informed about payoff-relevant information than any third party. For their part, lawyers often attribute the failure to complete contracts to the inevitable ambiguities in ordinary language or to some bounded rationality on the part of the parties (perhaps arising from their decision not to employ lawyers in the drafting of the contract).

In this section, accordingly, we discuss the problem of contractual incompleteness and relate it to the question of interpretation and to associated legal doctrines. We begin in the next subsection by considering various definitions of contractual incompleteness. Subsection 4.2 then discusses the sources of contractual incompleteness, including *ex ante* determinants present at the outset of contracting, such as bounded rationality and asymmetric information, as well as *ex post* factors such as verification costs and dynamic incentives arising from the prospect of renegotiation. Subsection 4.3 addresses the consequences of contractual incompleteness for the efficiency of exchange; and subsection 4.4 outlines and analyzes the main legal doctrines that govern in this area.

## 4.1 Modeling incomplete contracts

From a theoretical perspective, it is useful to model a contract as a mapping from *verifiable* events to outcomes. For instance, an insurance contract could contain a provision that related damage to one's car (a verifiable event) to a payment to the insured (an outcome).

In this context, "verifiable" means an event is observable not only by the parties to the contract, but also by any third party (*e.g.*, a judge) who might be called upon to adjudicate a dispute. The focus on verifiable events is motivated as follows. Were an outcome contingent on an unverifiable event (*i.e.*, one not observable to the third party), then there would be no way for the third party to judge the extent of breach of contract (if any) or even who breached (if anyone did). Hence, a contractual obligation that is contingent on an unverifiable event cannot be effectively enforced by a third party.

In the incomplete-contracts literature, it is standard to assume that the parties to a contract can observe events that cannot be verified by any judge. In the parlance of the literature, such events are described as observable, but not verifiable. As we will see, the parties could ideally wish to base their contract on such observable, but unverifiable events.

Enforcement would also be difficult if one of the parties to the contract couldn't observe an event on which an outcome was contingent (for example, as in the case of a consumer who cannot tell whether a mechanic has properly repaired an automobile). That party would not know the extent to which the other party was out of compliance with the contract (if he was). Such ignorance would, thus, make a contract impossible to enforce even through private sanctions of the sort discussed in §5.4.2 and §5.4.3 below. For this reason *observability* is considered a minimal informational requirement for an event to define a contractual contingency.

To be more formal, if we take  $\Omega$  to be the set of verifiable events (with  $\omega$  a representative element) and  $\mathcal{A}$  to be the set of outcomes, then a contract can be seen as a mapping,  $C : \Omega_C \rightarrow \mathcal{A}$ , where  $\Omega_C \subseteq \Omega$ . Contractual incompleteness can, then, be seen as situation in which the parties to the contract would or should ideally wish to base their contract on some set other than  $\Omega_C$  (for example,  $\Omega$  if  $\Omega_C \subset \Omega$ ; or the set  $\mathcal{O}$  of observable events).

Within the economics literature, the terms verifiable and observable are typically employed without any explicit consideration of the costs associated with investigation, measurement, documentation, or monitoring; that is, activities necessary to make information useful contractually. To an extent, one implicit set of assumptions typically employed is that observable events are observable at no cost, similarly for verifiable events, while *unverifiable* events would be verifiable only at a prohibitive cost (possibly infinite). As Hermalin and Katz (1991) point out, events can be "partially" verifiable in the sense that although a third party cannot observe them with the precision of the parties to the contract, a third party can still observe some information about the event that causes him or her to update his or her beliefs about what happened in a way useful for adjudication (see discussion below in §4.1.3 and §4.3.1). In any case,

whatever the costs of observation and verification, they are almost always taken to be exogenous to the model. Endogenizing these costs through the design of measurement or monitoring systems remains, for the most part, an important area for future research.

#### 4.1.1 Literal incompleteness and unmapped contingencies

Following Hermalin and Katz (1993), we make a distinction between literal incompleteness, which we consider now, and economic incompleteness, which we address in §4.1.3.<sup>49</sup>

A contract is literally incomplete if an event or contingency can arise that is not anticipated by the contract; hence, the contract is silent with respect to what should happen given this event or contingency. Literal incompleteness corresponds to a situation in which there are elements of  $\Omega$  not in  $\Omega_C$ . Ayres and Gertner (1989) refer to such as elements as “gaps.” Other scholars have referred to them as *unforeseen contingencies*. Let  $\bar{\Omega}_C$  denote the set of gaps or unforeseen contingencies (*i.e.*,  $\bar{\Omega}_C = \Omega - \Omega_C$ ).

The assumption of literally incomplete contracts has played an important role in law & economics (both implicitly, as in Shavell, 1980, and Rogerson, 1984, and explicitly, as in Goetz and Scott, 1981, Ayres and Gertner, 1989, and Hadfield, 1994). Nonetheless, as Hermalin and Katz (1993) observe, it is a potentially problematic assumption, because it is so easy to complete contracts by adding a stereotypical residual (“none-of-the-above”) clause to a contract. That is, literal *completeness* can be achieved simply by adding a clause that states, “if an event (contingency) other than those listed above occurs, then the outcome shall be . . .”

In order to explain literally incomplete contracts, consequently, it is not enough to assume that the parties fail to foresee some contingencies. It would seem necessary to assume also that the parties fail to foresee that they could fail to have foreseen some contingencies.

Alternatively, one might assume the gap is rational, insofar as there is some affirmative reason why the parties deliberately left contingencies unmapped. One reason could be that the parties are content to let the courts apply a particular outcome, and this outcome is no worse than what the parties might have provided on their own. Under traditional common law, for instance, courts often responded to contractual gaps by treating the contract as a nullity, on the grounds that there had been no “meeting of the minds.” The result would be that the parties would be left where they lay when the contingency arose (although parties who had partially or fully performed might be entitled to a refund or to restitution of any value conferred on the counterparty). Modern-day courts, in contrast, are more likely to react to a gap by trying to fill it with either an objectively reasonable term, or with their best guess as to what the parties would have wished had they negotiated over the contingency in question.

---

<sup>49</sup>Ayres and Gertner (1992) refer to the first type of incompleteness as “obligational” and second as “insufficiently state contingent.”

(Although if the gaps are too significant or the parties' hypothetical wishes too unclear, the traditional approach of declaring the contract to be at an end is still a real possibility.)

To the extent that courts apply a predictable default rule in such situations (or even an unpredictable one), one could say that there is no such thing as a literally incomplete contract—rather, the implicit residual clause is just that the parties agree to go to court—or follow existing legal requirements that they pursue private arbitration or other method of dispute resolution—and abide by any result that is reached there. At a semantic level, such logic is unassailable, if perhaps unappealing to those worried about truly naïve parties who indeed failed to foresee that might have failed to have foreseen a contingency.

#### 4.1.2 Linguistic under- and overdetermination

In many situations in which a contract does not point to a clear outcome, the problem is not that the parties have said nothing; rather, it is that they have said too little or too much. For example, the parties might provide multiple and inconsistent provisions dealing with the same event. Alternatively, the parties might use terms that admit multiple meanings or that depend on other terms of the contract.

One could treat such cases as equivalent to contractual gaps, arguing that, if the parties have not settled a term definitively, they have not settled it at all, but this interpretation does not appear to comport with the behavior of either legal institutions or actual bargainers. An alternate approach, accordingly, would be to model contracts not as single-valued functions, but as multi-valued correspondences. Formally, instead of representing a contract as a mapping of the form  $C : \Omega_C \rightarrow \mathcal{A}$ , we could represent it as a mapping of the form  $C : \Omega_C \rightarrow P(\mathcal{A})$  where  $P(\mathcal{A})$  denotes the power set of  $\mathcal{A}$ .<sup>50</sup> Under this interpretation, incompleteness would arise whenever the contract mapped an event to a set with more than one outcome, so that the person applying the contract would have to choose among those outcomes based on extra-contractual factors.

This type of incompleteness has received less attention in the literature, but some recent work has begun to explore its implications. For instance, a recent paper by Hart and Moore (2004) models a contract as a list of outcomes to which the parties are restricted. The determination of which outcome from that list is implemented occurs through *ex post* bargaining. Through this approach, they show that a relatively loose contract (*i.e.*, a relatively long list) preserves flexibility, allowing the parties to make use of new information that arises, but at the expense of distorting *ex ante* investments due to the increased danger of holdup. Similarly, Ben-Shahar (2004) has argued that courts should respond to such incompleteness by granting both parties the option to enforce the agreement on whatever terms are most favorable to the counterparty; such a response would preserve the benefits of whatever bargain the parties had already reached, while allowing them to enjoy further gains from trade through

---

<sup>50</sup>A sigma field should  $\mathcal{A}$  be non-denumerable.

further negotiation.

A criticism of Hart and Moore (2004) is that relies heavily on some strong assumptions. First, they rely heavily on the observable but unverifiable distinction. As Hermalin and Katz (1991) and Maskin and Tirole (1999) note, there are reasons to doubt this distinction in many contexts. More critically, they also rely on the ability of parties to carry out what would normally be seen as incredible threats. Hart and Moore justify these by appealing to the psychological tendency of people to sacrifice resources in order to get revenge against those who have, in their eyes, mistreated them (see Rabin, 1993, for a discussion of such psychological effects in economic contexts). While such tendencies might exist among individuals, one is hesitant to assume that sophisticated parties (*e.g.*, firms) would be prey to such base emotions.

### 4.1.3 Economic incompleteness and coarse mappings

Much of the *economic* literature on incomplete contracts has focused on a different standard of incompleteness: A contract is incomplete when the set of verifiable events is not the same as the set of observable events.<sup>51</sup> That is, even if the contract is literally complete (*i.e.*,  $\bar{\Omega}_C = \emptyset$ ) it would be judged *economically* incomplete if  $\Omega \neq \mathcal{O}$ , where  $\mathcal{O}$  is the set of observable events. In this case, the parties would benefit from being able to condition their contractual obligations more finely (for instance, by allowing an excuse when performance is commercially impractical), but they cannot do so because the condition cannot be effectively enforced (for instance, because the court cannot tell the difference between commercial impracticability and mere seller recalcitrance).

As noted earlier, it is natural to assume  $\mathcal{O}$  is “larger” than  $\Omega$ . In the literature, there are essentially two ways this assumption is modeled.<sup>52</sup> One is to define  $\Upsilon$  as the set of observable, but unverifiable events (with representative element  $v$ ) and take

$$\mathcal{O} = \Upsilon \times \Omega.$$

In other words, an observable event is a vector consisting of events that the parties can observe, but not verify, and events that the parties can both observe and verify.<sup>53</sup> Under this formulation, a verifiable event  $\omega$  reveals not only that

<sup>51</sup>Obviously, both sets must be defined in some way over *relevant* events; that is, we don’t want to say the sets differ if the observable, but unverifiable events are irrelevant to the contracting situation. Defining relevance is, however, difficult insofar as one strain of the literature (*e.g.*, Hermalin and Katz, 1991; Maskin and Tirole, 1999; Edlin and Hermalin, 2001) has been devoted to showing when the observable, but unverifiable distinction is unimportant; that is, situations in which the parties can achieve the same outcomes with incomplete contracts as they could with complete contracts. In such situations, it would be misleading to say that the observable and verifiable sets are the same over relevant events, because equilibrium play, but not outcomes, would be different were it feasible to base contracts on the set of observable events. A definition of relevant events must, therefore, account for potential differences in the ensuing game, as well as outcomes.

<sup>52</sup>An example of the first way is Grossman and Hart (1986). An example of the second way is Anderlini and Felli (1994).

<sup>53</sup>It is worth noting that both  $\Upsilon$  and  $\Omega$  could, themselves, be vector spaces.

$\omega$  has occurred, but also that

$$v \in \mathcal{O}_\omega \equiv \{v | (v, \omega) \in \mathcal{O}\}$$

(the set  $\mathcal{O}_\omega$  is called the *section of  $\mathcal{O}$  at  $\omega$* ). Because, absent additional assumptions,  $\mathcal{O}_\omega = \Upsilon$  for all  $\omega \in \Omega$ , it might at first seem that knowing that  $v$  is in  $\mathcal{O}_\omega$  is not useful information. But if some  $(v, \omega)$  pairs are impossible, then some sections satisfy  $\mathcal{O}_\omega \subset \Upsilon$  and, therefore, learning  $\omega$  can lead to inferences about which  $v$  occurred.

In such situations, a so-called “forcing” contract (as in Harris and Raviv, 1978) could be useful. Suppose the parties wish to induce one party, “the actor,” to choose a specific  $\hat{v}$ . If there are  $\omega$  such that  $\hat{v} \notin \mathcal{O}_\omega$ , then those  $\omega$  constitute proof that the actor has not chosen the desired action and the contract can, therefore, threaten the actor with sufficient punishment should such  $\omega$  occur that he would never choose an  $v \in \mathcal{O}_\omega$ . Hence, some undesirable actions can be avoided. Ideally, if there is a subset  $\Omega_0 \subset \Omega$  such that, for all  $v \neq \hat{v}$ ,  $v \in \mathcal{O}_\omega$  for some  $\omega \in \Omega_0$ , but  $\hat{v} \notin \mathcal{O}_\omega$  for any  $\omega \in \Omega_0$ , then  $\hat{v}$  can be achieved by a contract that sufficiently punishes the actor for any  $\omega \in \Omega_0$  and rewards him only for  $\omega \in \Omega \setminus \Omega_0$ .<sup>54</sup>

Even if all  $(v, \omega)$  pairs are possible—so  $\mathcal{O}_\omega = \Upsilon$ —the *conditional distribution* of  $v$  given  $v \in \mathcal{O}_\omega$  may differ from the distribution given  $v \in \mathcal{O}_{\omega'}$ ,  $\omega \neq \omega'$ . Such differences in distributions can be very powerful, often allowing contractual solutions that replicate the outcomes that would have occurred were the parties able to contract on  $\mathcal{O}$  directly (see discussion in §4.3.1; also Hermalin and Katz, 1991).

The second way economic incompleteness can be understood is to take  $\Omega$  to be a *partition* of  $\mathcal{O}$ ; that is, each  $\omega$  is a subset of  $\mathcal{O}$ , any given element of  $\mathcal{O}$  can be in only one  $\omega$ , and the  $\omega$ s, as a class, contain all the elements in  $\mathcal{O}$ .<sup>55</sup> Of course, to make this interesting, there must be at least one  $\omega$  with two or more elements of  $\mathcal{O}$  in it; that is, for this  $\omega$  at least, learning  $\omega$  does not perfectly reveal the observable information. Conversely, economic *completeness* would correspond to a situation in which each  $\omega$  contained a single element of  $\mathcal{O}$ .

The standard interpretation of this second representation of economic incompleteness is that there is inherently no difference between what is observable and what is verifiable, except that it is impossible or prohibitively expensive to describe the observable events with sufficient precision in a contract.<sup>56</sup> Instead, the observable events can only be described coarsely; so that different observable

<sup>54</sup>The notation  $S \setminus T$  denotes the set whose elements are in  $S$  but not  $T$  (i.e.,  $S \setminus T = S \cap T^c$ ).

<sup>55</sup>Formally, the  $\omega$ s satisfy

$$\omega \cap \omega' = \emptyset, \forall \omega \neq \omega'; \text{ and } \bigcup_{\omega \in \Omega} \omega = \mathcal{O}.$$

This second way of representing economic incompleteness can be seen as a special case of the first in which  $\mathcal{O} \subset \Upsilon \times \Omega$  such that  $\mathcal{O}_\omega \cap \mathcal{O}_{\omega'} = \emptyset$  for any two  $\omega, \omega'$  and  $\bigcup_{\omega \in \Omega} \mathcal{O}_\omega = \Upsilon$ .

<sup>56</sup>Typically these costs are taken to be “writing costs.” They could also, as Rasmusen (2001) notes, be due to “reading” costs.

events get lumped together in a single  $\omega$ . For example, consider the “contract” between us, the authors of this chapter, and the editors of this volume. Without effectively writing the chapter themselves, it is impossible for the editors to stipulate fully what the chapter should look like; the best they can do is stipulate the topics to be covered and the overall length of the chapter.

#### 4.1.4 Discussion

Each of these models of contractual incompleteness has some appeal, though none is entirely satisfactory as a complete account of the phenomenon. To illustrate, consider the case of *Spaulding v. Morse*, in which a divorcing couple entered into a maintenance agreement under which the husband agreed to pay child support of \$1200 per year until the couple’s son entered some “college, university, or higher institution of learning . . .,” and \$2200 per year for four years thereafter. The dispute arose when the son completed high school and was immediately drafted into the US Army, and the husband took the position that his contractual obligation was suspended while the son remained in military service. One can view this dispute as arising from literal incompleteness if we say that the contract simply did not provide for the case in which the son was drafted into the army. In order to take this view, however, we must read the term in question in a non-literal way, since a literal reading makes it appear to be a residual clause. (That is, it appears to say that so long as the son never started college—*e.g.*, if he chose to pursue a military career or even if he were killed in combat—the husband would be required to make yearly payments in perpetuity.)<sup>57</sup>

One can alternatively view the dispute as arising from linguistic underdetermination if we focus on the phrase “child support.” Does the phrase simply refer to any payments made for the benefit of the son, or is it instead limited to payments that are objectively necessary for the son’s adequate maintenance? Ordinary language is probably not precise enough to provide a definitive answer to this question, and the best response may be to say that the phrase can mean either or both.

Finally, we could view this dispute as a case of economic incompleteness if we assume that what the parties actually wished to do was to condition payment on the son’s need for maintenance (an observable but perhaps not verifiable event) but that they instead conditioned on the son’s not yet having completed four years of college (a coarser but perhaps more easily verifiable proxy). Had the parties been able to describe the former condition at reasonable cost in their divorce settlement, the whole dispute could have been avoided. On the other hand, it is not clear that the couple had such a clear purpose in arriving at the clause in question. Another possibility is that the husband simply wished to

---

<sup>57</sup>Nobody in the case made this argument, however, not even the wife. Instead, the court found in favor of the husband, on the grounds that the unstated purpose of the clause was to provide financial support for the son in his minor years, and once the Army had taken responsibility in this regard, the purpose had been served.

pay as little as possible, the wife wished to receive as much as possible, and the clause was a rough compromise.

As the example indicates, modeling contracts as a formal mapping from events to outcomes is useful, but it could fail to capture two important real-world features of the situation—namely, that parties do not always possess a well-specified understanding of the domain of possible events, either *ex ante* or *ex post* and that they write contracts in ordinary language with all its ambiguities, rather than in the formal but restricted language of mathematics. This is not to say, however, that such failures necessarily invalidate or diminish the insights of formal modeling. Nor is it to say that formal modeling can't deal with these features should they be important. Certainly, the first could be readily modeled by incorporating some uncertainty over what the domain of events might be. The second is potentially trickier, but techniques such as assuming asymmetric information with regard to the meaning of the terms, uncertainty over their interpretation, or even suspension of the usual common-priors assumption could be useful. With all that in mind, the next subsection discusses possible economic causes of contractual incompleteness.

## 4.2 The sources of contractual incompleteness

### 4.2.1 Bounded rationality

A common explanation for incomplete contracts, especially literally incomplete contracts, is bounded rationality.

The simplest “model” of bounded rationality is that people make mistakes. They fail to foresee all possible contingencies and, thus, their contracts suffer from unforeseen contingencies. However, as noted above (§4.1.1), failure to foresee certain contingencies need not generate incomplete contracts unless the parties also fail to foresee that they may fail to foresee certain contingencies. If the parties recognize their imperfect foresight, then they can complete any contract with a residual (“none-of-the-above”) clause.

Of course, the fact that the parties *could* complete any contract with a none-of-the-above clause does not mean that they would wish to do so. After all, they may fear that the ideal response to different unforeseen contingencies varies with those contingencies. By its nature, a none-of-the-above clause is a one-size-fits-all solution. Hence, the parties may wish for flexibility should an unforeseen contingency occur; legal proceedings can provide such flexibility insofar as the court's ruling will typically depend on the contingencies that have occurred. How the courts should respond to these “intentional” gaps has become an issue in the literature (see, *e.g.*, Ayres and Gertner, 1989, 1992). We discuss this literature in §4.4 below.

To an extent, deciding how the courts should respond depends on the reason for unforeseen contingencies in the first place. As, however, E. Posner (2003a) points out, economic theory does not offer particularly good or widely accepted means of modeling unforeseen contingencies (or bounded rationality more gen-

erally).<sup>58</sup> To the extent models of unforeseen contingencies exist, they are too abstract to be readily utilized in the study of contracts. Admittedly, one could simply appeal to the existence of gaps in actual contracts to justify assuming literal incompleteness. Such an atheoretic rationale, however, makes it difficult to analyze important issues such as how the courts would know whether gaps are truly mistakes or have arisen because one side or the other is being strategic. This, in turn, affects how the courts should respond to gaps (see Ayres and Gertner, 1989, 1992, and Shavell, in press).

Although some parts of the literature appeal to unforeseen contingencies as a possible motivation for *economic* incompleteness (see, *e.g.*, Maskin and Tirole, 1999), it is difficult to reconcile such bounded rationality with the “dynamic-programming” rationality (to use Maskin and Tirole’s term) that the incomplete-contracts literature typically assumes. Other, more rational, explanations for economic incompleteness strike us as more plausible. We explore some of these explanations in the next few sections.

#### 4.2.2 Describability and contracting costs

As we observed, a complete contract between us and the book’s editors about what this chapter should be is rather impractical. If the editors spelled out in great detail what they wanted, they would, in effect, be writing the chapter themselves. Their costs of doing presumably outweigh whatever benefits such detailed contracting would generate. More generally, there is a cost to writing fine-tuned contracts. In some instances, such as our chapter example, it is simply impractical to describe the relevant contingencies finely enough.

DESCRIBABILITY. Because describing something can be viewed as an algorithm, it follows from the mathematics of computability (see, *e.g.*, Boolos et al., 2002) that it is conceivable that some events or contingencies cannot be described, in the sense that no algorithm for describing them would ever finish. This point is made by Anderlini and Felli (1994). An intuitive sense of this idea can be had by recalling the well-known math fact that the constant  $\pi$  cannot be given a decimal representation; no algorithm could ever complete the task of fully describing  $\pi$  as a decimal number.

Indescribability *per se*, however, seems a doubtful explanation for economically incomplete contracts for at least two reasons. First, the parties might be able to approximate their desired description arbitrarily well (*e.g.*, 3.1415 might be a good enough approximation of  $\pi$ ); therefore, they suffer arbitrarily little from the indescribability of a contingency. In fact, Anderlini and Felli prove just such an approximation theorem. Second, the parties to the contract need to have internal descriptions of the relevant events or contingencies to know if they have occurred; hence, if they cannot write descriptions, it is hard to under-

---

<sup>58</sup>For a general survey of recent modeling on bounded rationality see Rubinstein (1998). For a more specific survey on unforeseen contingencies see Dekel et al. (1998) (also see Dekel et al., 2001).

stand how they know the relevant descriptions. In other words, if we take the feasible contractual contingencies to be subsets of some partition of the relevant state space, then why isn't the set of observable contingencies this same partition?<sup>59</sup> And, if it is, then there is no practical difference between observable and verifiable; that is, the contract is *not* economically incomplete.

DESCRIPTION COSTS. A better way to proceed is to assume that the relevant states can be described (or described arbitrarily finely), but to recognize that the costs of doing so can sometimes be so large as to render detailed descriptions impractical (see, *e.g.*, Dye, 1985). The chapter-writing example given earlier is an example of such a situation.

Having recognized that contract costs are increasing in the details of the contract,<sup>60</sup> the next question is how to balance the marginal gain from more details against these marginal costs? As it turns out, however, in many settings the marginal *benefit* of adding details is zero; as Hermalin and Katz (1991) and Maskin and Tirole (1999) show, even rough descriptions can be sufficient for the parties to do as well as they could were it practical to write very detailed contracts. We return to this point later.

Another reason the marginal benefits to the parties could be small is that the courts will fill in or interpret the missing or vague terms (see, *e.g.*, Ayres and Gertner, 1989, 1992, or Shavell, in press). Hence a model of how the courts enforce incomplete contracts is essential to an analysis of the effort that parties should make to make their contracts more precise.

EVALUATION AND EXPERTISE. It is possible that even if a judge, or other adjudicator of contractual dispute, can observe a relevant variable, he or she may not properly evaluate it. His or her evaluation could differ from that of the parties to the contract, perhaps because of a lack of expertise in the relevant field. This idea can be captured by assuming that, when the parties observe  $x$ , the judge observes (concludes)  $x + \varepsilon$ , where  $\varepsilon$  is the judge's error in evaluation. The error,  $\varepsilon$ , could also be introduced because the judge reads or interprets the contract differently than the parties intended. (The model, which we consider in §4.4.1 *infra*, on form versus substance also touches on this point.)

While such errors might, at first, seem to render the relevant variable essentially unverifiable, the parties can, in some circumstances, do as well as they would were the judge to observe the relevant variable without error (Hermalin and Katz, 1991). We return to this point later in §4.3.1.

---

<sup>59</sup>Let  $\Sigma$  be the state space. What can be described, the verifiable events, are subsets of  $\Sigma$  (*i.e.*,  $\omega \subset \Sigma$  for all  $\omega \in \Omega$ , where  $\bigcup \omega = \Sigma$  and  $\omega \cap \omega' = \emptyset$ ). But if, as seems reasonable, what the parties can describe in writing is the same as what they can describe to themselves, then each  $\omega$  is, then, an element of  $\mathcal{O}$  and  $\mathcal{O} = \Omega$ . [In a sense, this is reminiscent of Wittgenstein's (1922) Proposition 7 that what we can't think about, we can't verbalize.]

<sup>60</sup>One can model contracts being more detailed by assuming that there is an order,  $\succ$ , on partitions of  $\mathcal{O}$ , such that  $\Omega_i \succ \Omega_j$  (*i.e.*, a contract based on  $\Omega_i$  is more detailed than one based on  $\Omega_j$ ) if for each  $\omega_i \in \Omega_i$  there exists an  $\omega_j \in \Omega_j$  such that  $\omega_i \subseteq \omega_j$  and for at least one  $\omega_j \in \Omega_j$  there exist  $\omega_i$  and  $\omega'_i$  in  $\Omega_i$  such that  $\omega_i \cup \omega'_i \subseteq \omega_j$ .

### 4.2.3 Complex environments

Segal (1999a) suggests that economically incomplete contracts can arise when the contracting environment is complicated. In his model, the parties wish to trade one “widget” from a set of different widgets. As the number of potential contingencies (*i.e.*, widgets in the set) rises, the optimal second-best contract does increasingly worse. In the limit, the optimal second-best contract does no better than the simple incomplete contract (equivalently, in his model, no contract). The reason for this worsening performance is that, as the number of contingencies rises, the number of incentive constraints increases. As in all optimization problems, as the number of constraints increases, the optimality of the solution decreases (at least weakly). Given that writing complex contracts is costly, there is a cutoff point at which the environment is so complex that the simple, less costly contract is superior to a complex contract.

Segal’s model, while elegant, rests on a number of strong assumptions. Among these is the assumption that almost no variable of interest is verifiable, despite many of them being observable (his Assumption 2, page 60). If either the optimal widget to trade is known *ex ante* or which is the optimal widget to trade can be *verified ex post*, the model collapses. As we have observed, the assumption of observable but unverifiable can be a questionable one, especially in many of the contexts to which this model is intended to apply.

### 4.2.4 Asymmetric information

As noted in §2.3.2, informational asymmetry at contracting can lead to distortions in the contract that is written. Spier (1992) builds on this idea by suggesting that one consequent distortion could be contractual incompleteness.

A simple model can serve to illustrate her idea. Suppose a manufacturer needs a part for one of its manufacturing machines. The manufacturer can have the part sent by a default carrier, in which case the part will certainly arrive in two days. Normalize the default carrier’s price to be 0. Alternatively, the manufacturer can contract with an express carrier. If the express carrier makes no special efforts, the part will arrive in one day with probability  $1 - q \in (0, 1)$ ; with probability  $q$  it arrives the day after that. For simplicity, take the express carrier’s cost when it makes no special efforts to be 0.<sup>61</sup> Alternatively, by spending  $c > 0$ , the express carrier can ensure the part arrives in one day. Let  $L > 0$  be the amount per day that the manufacturer loses from the idle machine. If the manufacturer selects the default carrier, its payoff is  $-L$ . If it selects the express carrier, but no special efforts are made, its expected gross payoff is  $-qL$ . If it selects the express carrier and special efforts are made, its expected gross payoff is 0; but, in this case, the express carrier’s gross payoff is  $-c$ . Assume that there are two types of manufacturer, 0 and 1, with

$$qL_1 > c > qL_0.$$

---

<sup>61</sup>Making it a positive amount does not alter the analysis materially, while setting it to zero simplifies the notation.

Observe that it would be welfare maximizing to have the express carrier make special efforts for a high-value manufacturer (type 1), but not for a low-value manufacturer (type 0). Let  $\theta \in (0, 1)$  be the probability that the manufacturer is type 1. Hence,

$$\mathbb{E}L = \theta L_1 + (1 - \theta)L_0.$$

Assume that the manufacturer's type is its private information.

To close the model, we need to make some assumptions about bargaining between the manufacturer and the express carrier. Assume that if the express carrier can't infer the manufacturer's type, then the express carrier and manufacturer set the price through Nash (1950) bargaining over the expected surplus,  $\mathbb{E}L - q\mathbb{E}L$ . In that case, the price is

$$p_u = \frac{1}{2}(1 - q)\mathbb{E}L.$$

Assume the parameters are such that

$$-qL_0 - p_u > -L_0$$

(this clearly holds for low values of  $\theta$ ). Because it is welfare reducing for a low-value manufacturer to request special efforts, we can assume that the express carrier interprets any request for special efforts to be from a high-value manufacturer. In this case, total surplus is  $L_1 - c$ . If bargaining is again Nash bargaining, then the resulting price is

$$p_i = \frac{L_1 + c}{2}.$$

Finally, consider a high-value manufacturer who is deciding between asking for special efforts—and thus revealing its identity—or pooling with low-value manufacturers and not asking. Its expected net surplus in the first case is  $-p_i$ . Its expected net surplus in the second case is  $-qL_1 - p_u$ . It is readily shown that there exist parameter values such that latter exceeds the former (*e.g.*,  $c = 2$ ,  $L_1 = 5$ ,  $L_0 = 3$ ,  $q = 1/2$ , and  $\theta = 1/4$  would work). Hence, it is possible that asymmetric information leads to a form of contractual incompleteness: Although it would be welfare improving for a high-value manufacturer and the express carrier to have a contingency in their contract requiring the carrier to make special efforts, they don't have such a contingency in equilibrium.

#### 4.2.5 Verification costs

In much of the contracting literature, it is assumed that, if information is verifiable, it can be verified costlessly. In real life, however, information is made verifiable at a cost (*e.g.*, surveillance monitoring, auditing, and record keeping). If such costs are large relative to the benefits that complete contracting affords, the consequence will be incomplete contracts.

There is a strand of the literature on “costly state verification,” starting with Townsend's (1979) seminal article, that considers, in part, the consequences of

an inability to verify payoff-relevant variables due to the cost of verification. The principal concern of this literature has been on financial contracting; particularly the decision to use debt rather than equity.<sup>62</sup>

#### 4.2.6 Dynamic inconsistency with respect to renegotiation

A common, although not universal, view in the economics literature is that parties never “leave money on the table” even on out-of-equilibrium paths. More precisely, should the parties ever reach a point at which it is common knowledge that the allocation dictated by the contract is Pareto inferior, the parties will renegotiate the contract. This ability to renegotiate undermines the value of contracts that rely on Pareto-inferior allocations as threats should one or both parties deviate from their contractual obligations. In the lingo of the literature, such contracts fail to be *renegotiation proof*. That contracts be renegotiation proof is, therefore, often a requirement imposed in the literature.

To illustrate the concept of renegotiation-proofness and how it can lead to incomplete contracts, we briefly review the Fudenberg and Tirole (1990) model of renegotiation in agency. In a standard agency model, to induce the agent to work hard, the principal and agent agree to a contract that makes the agent’s compensation contingent on the outcome (*e.g.*, salespeople’s compensation is frequently tied to their sales). Such a contract is, however, generally not first-best optimal because it causes a risk-averse agent to bear risk. Fudenberg and Tirole consider what happens in the standard model if there is a period after the agent has committed his effort but before its outcome is known (*e.g.*, after he is back from his sales calls, but before the orders arrive) during which the principal and agent can renegotiate. Observe that if, as is typically assumed, the principal is risk neutral, then it is common knowledge that leaving the agent with risky compensation is inefficient; the parties should renegotiate to a fixed, non-contingent wage for the agent. However, a rational agent would forecast such renegotiation and, thus, that his compensation will ultimately be unrelated to his effort. But if it is unrelated to his effort, then he has no incentives to work hard. While Fudenberg and Tirole show that there are ways to restore *some* incentives, the potential for renegotiation limits their effectiveness. Moreover, a consequence could well be that the parties forgo using an incentive contract even though they would use one were renegotiation not possible.

This last point shows how the threat of renegotiation can lead the parties to use less complete contracts than they would were renegotiation infeasible. That is, while they might wish to write a contract contingent on the outcome, they can’t. Schwartz and Watson (2003) consider the relation between contractual complexity and the feasibility of renegotiation in greater depth, with a particular emphasis on law & economics issues.

---

<sup>62</sup>See, for instance, Webb (1992) or Hart and Moore (1998).

### 4.3 Consequences of contractual incompleteness

#### 4.3.1 Does incompleteness matter?

A debate has emerged in economics as to whether incomplete contracts matter. Specifically, even if it is accepted that contracts are incomplete, are there ways for the parties to effectively contract around the incompleteness?

Hermalin and Katz (1991) offer one model in which incompleteness doesn't matter. Let  $\Upsilon$  denote the set of observable, but unverifiable variables. Assume that  $\Upsilon$  is finite. Let  $\Omega$  denote the set of verifiable variables. Take it to be finite as well, with  $N$  elements indexed by  $n$ . Let there be two parties to the contract: the agent and the principal.

The agent chooses  $v \in \Upsilon$ . The principal has preferences over the  $v$ s. Suppose that the agent is willing to sign a contract with the principal if his equilibrium (expected) utility is at least some reservation level, which we can normalize to be zero. Let  $w$  denote the agent's gross utility and let  $w - d(v)$  denote his net utility ( $d : \Upsilon \rightarrow \mathbb{R}_+$ ). Assume it costs the principal  $c(w)$  if the agent receives gross utility  $w$ , where  $c(\cdot)$  is increasing and strictly convex.<sup>63</sup> Finally, assume that the conditional distribution of the verifiable variable,  $\omega$ , depends on the  $v$  chosen; specifically, let  $\pi_n(v) = \Pr\{\omega_n|v\}$  and let

$$\boldsymbol{\pi}(v) \equiv (\pi_1(v), \dots, \pi_N(v))$$

be the density over  $\Omega$  induced by  $v$ .

Suppose that the principal wants the agent to choose a specific  $v^*$ . If we assume that the principal has all the bargaining power, her ideal would be a contract in which the agent agrees to choose  $v^*$  in exchange for a net utility of 0, which costs the principal  $c(d(v^*))$ . Except in the special case where  $v^*$  minimizes  $d(v)$ , this ideal contract is infeasible because  $v$  is not verifiable and the agent prefers an  $v$  other than  $v^*$ .

A feasible alternative is a contingent-compensation contract; the principal agrees to provide utility level  $w_n$  should verifiable outcome  $\omega_n$  occur. Let  $\mathbf{w} = (w_1, \dots, w_N)$ . Such a contingent contract will be agreeable to the agent and induce him to select  $v^*$  if

$$\boldsymbol{\pi}(v^*) \cdot \mathbf{w} - d(v^*) \geq 0 \text{ and} \tag{7}$$

$$\boldsymbol{\pi}(v^*) \cdot \mathbf{w} - d(v^*) \geq \boldsymbol{\pi}(v) \cdot \mathbf{w} - d(v) \text{ for all } v \in \Upsilon \setminus \{v^*\}. \tag{8}$$

Condition (7), the individual rationality constraint, is the condition that the agent agree to the contract. Condition (8), the set of incentive compatibility constraints, is the condition that the agent prefer choosing  $v^*$  to any other  $v$ .

Suppose the principal has all the bargaining power. It is a well-known result (see, *e.g.*, Grossman and Hart, 1983) that if a contract (a vector  $\mathbf{w}$ ) exists satisfying expressions (7–8), then there exists one such that (7) holds as an equality. Because that contract minimizes the principal's expected cost, she

<sup>63</sup>A natural interpretation is that the agent is risk averse, while the principal is risk neutral, in which case  $c(\cdot)$  is the inverse function of the agent's utility function.

will offer that contract. Under general conditions (see, *e.g.*, Grossman and Hart, 1983),  $\text{Var}(w|v^*) > 0$ . Hence, by Jensen's inequality,  $\mathbb{E}\{c(w)|v^*\} > c(d(v^*))$ ; that is, even if the principal can devise a contract to implement  $v^*$ , it will cost her more than the ideal contract would were  $v$  verifiable. It would seem, therefore, that there is an adverse consequence to  $v$  not being verifiable.

This last conclusion, however, need not follow if the parties have time between the realizations of  $v$  and  $\omega$  to renegotiate the contract. Suppose that the principal retains all the bargaining power in renegotiation.<sup>64</sup> Because, here, the principal observes  $v$ , the problem identified by Fudenberg and Tirole (1990) is not triggered by renegotiation (see discussion in §4.2.6). Because the principal's cost increases with the variability of  $w$ , the principal would offer, in renegotiation, a non-contingent  $w$ ,  $\bar{w}$ . Because the agent's cost from selecting  $v$  is sunk at the point of renegotiation, he will accept  $\bar{w}$  if and only if  $\bar{w} \geq \pi(v) \cdot \mathbf{w}$ . As she has all the bargaining power, the principal will choose the lowest such  $\bar{w}$ , namely the one that just equals  $\pi(v) \cdot \mathbf{w}$ . Hence  $\bar{w}$  is a function of  $v$ , which we indicate by writing  $\bar{w}(v)$ . By Jensen's inequality,  $c(\bar{w}(v)) \leq \mathbb{E}\{c(w)|v\}$ ; that is, renegotiation lowers the principal's cost. Moreover, because  $\bar{w}(v) = \pi(v) \cdot \mathbf{w}$ , it is readily seen that anticipation of such renegotiation does *not* change the individual rationality nor incentive compatibility constraints (7–8). Moreover, from (7), we see that  $\bar{w}(v^*) = d(v^*)$ . We therefore have

**Proposition 6** *Suppose that renegotiation is possible. If a contract  $\mathbf{w}$  exists that satisfies individual rationality and incentive compatibility (i.e., expressions 7 & 8), then even though  $v$  is unverifiable, the principal can achieve the same outcome as she could if it were verifiable.*

In other words, in this context, the distinction between observable and verifiable is not relevant to the outcome.

Proposition 6 rests on the existence of a  $\mathbf{w}$  satisfying expressions (7–8). Fortunately, the conditions for such a  $\mathbf{w}$  to exist are rather minimal. Somewhat loosely, it is sufficient that the distribution over  $\omega$  induced by  $v^*$  be distinct from the distributions induced by the other  $v$ .<sup>65</sup>

Proposition 6 also rests on the supposition that renegotiation is possible; that is, that there be a lag between when  $v$  is chosen and when  $\omega$  is realized. Renegotiation would seem possible in many contexts of interest. For instance, if  $\omega$  is a judge's observation and, thus, a ruling as to the value of  $v$ , then the parties presumably have time to renegotiate prior to the judge's verdict.

A provocative way to summarize this is as follows: Even when critical variables are not verifiable, parties can often present some relevant evidence that

<sup>64</sup>Edlin and Hermalin (2001) extend the analysis to a broader set of bargaining games in renegotiation.

<sup>65</sup>Formally, it is sufficient that  $\pi(v^*)$  not be an element of the convex hull of  $\{\pi(v)|v \in \Upsilon \setminus \{v^*\}\}$ . As Hermalin and Katz (1991) discuss, this condition is readily met in most situations of interest. As Edlin and Hermalin (2001) show, for more general bargaining games in renegotiation, a slightly stronger condition could be required: there exist a subset  $\Omega^*$  of  $\Omega$  such that  $\Pr\{\omega \in \Omega^*|v^*\} > \Pr\{\omega \in \Omega^*|v\}$  for all  $v \in \Upsilon \setminus \{v^*\}$ .

is informative, in a noisy way, about the critical variables. If renegotiation is feasible, then, by contracting on this relevant evidence, the parties can typically write contracts *as if* the critical variables were verifiable.

Renegotiation *à la* Hermalin and Katz is a particular type of game or mechanism. If a larger set of mechanisms is considered, one would find that the observability-verifiability distinction is irrelevant under an even wider set of circumstances. In a series of articles, Maskin and Tirole do just that, considering mechanisms more generally (see, *e.g.*, Maskin and Tirole, 1999; see, also, Tirole, 1999, for an overview).

In a mechanism, the parties send messages to the mechanism arbitrator (possibly the judge or other dispute adjudicator) about the variables they can observe. As Farrell (1987b) notes, perhaps the earliest recorded use of a “mechanism” is the one King Solomon used to determine which of two women claiming to be the mother of a child was the true mother.

A review of mechanism design is beyond the scope of this chapter. We can, however, give a sense of its application in this context. Suppose that there is some task that either of two parties, 1 or 2, could do. *After* contracting, the parties observe their costs  $(c_1, c_2) = v$  of doing the task, but cannot verify them. Suppose that  $c_i \in \{L, H\}$ , where  $L$  (low) is smaller than  $H$  (high).

Efficiency requires that the party with the lower cost do the task. The parties might further wish to have the one who doesn’t do the task partially compensate the other (*i.e.*, pay some portion of the cost). But, by assumption, the parties cannot write such a contract directly because it would be contingent on the realizations of their costs. Fortunately, however, a mechanism can be constructed that replicates the desired outcome: Both parties simultaneously announce their individual costs (*i.e.*, 1 announces  $c_1$  and 2 announces  $c_2$ ). The following rules govern what happens next:

- If they announce the same cost, then a coin is flipped. The loser of the coin flip has to do the task, but is paid  $\hat{c}/2$  by the winner, where  $\hat{c}$  is the commonly announced cost.
- If they announce different costs, then the lower-cost party must do the task, but she is paid  $H/2$  by the high-cost party.

If the parties announce truthfully, then the allocation of the task will be efficient. It remains, however, to check that the parties will announce truthfully in equilibrium (*i.e.*, that the mechanism is incentive compatible). There are three cases to consider. In each, we are assessing whether truth telling is a best response to truth telling; and, therefore, does a truth-telling equilibrium exist.

1.  $v = (L, L)$ . If a party announces truthfully, she expects to get

$$\frac{1}{2} \left( \frac{L}{2} - L \right) + \frac{1}{2} \left( -\frac{L}{2} \right) = -\frac{L}{2}.$$

If she lies, she expect to get  $-H/2$ ; hence, she does better to tell the truth than to lie.

2.  $v = (H, H)$ . If a party announces truthfully, she expects to get  $-H/2$  (see calculation above). If she lies, she must do the task, but is paid  $H/2$ ; hence, her net is  $-H/2$ . She does least as well to tell the truth; that is, truth telling is a best response.
3.  $v = (L, H)$  or  $(H, L)$ . Consider the higher-cost party. If she tells the truth, she expects to get  $-H/2$ . If she lies, she expects to get

$$\frac{1}{2} \left( \frac{L}{2} - H \right) + \frac{1}{2} \left( -\frac{L}{2} \right) = -\frac{H}{2};$$

so truth telling is a best response. Consider the lower-cost party. If she tells the truth, she expects to get  $H/2 - L$ . If she lies, she can expect

$$\frac{1}{2} \left( \frac{H}{2} - L \right) + \frac{1}{2} \left( -\frac{H}{2} \right) = -\frac{L}{2}.$$

Because

$$\frac{H}{2} - L > \frac{L}{2} - L = -\frac{L}{2},$$

she does better to tell the truth.

Although such mechanisms have an artificial feel to them, they can often be converted into more realistic looking contracts. For instance, the above mechanism can be converted into an *option contract*: Party 1 is obligated to do the task and Party 2 is obligated to pay Party 1  $H/2$ . However, Party 2 has the right (the option) to assume responsibility for the task if he wishes and, if he assumes responsibility, then Party 1 is obligated to pay him  $L/2$ . It is readily verified that Party 2 never gains from exercising his option if  $c_2 = H$  and always gains if  $c_2 = L$ . Hence, Party 2 assumes responsibility only if he is low cost, which ensures that the task is always undertaken by the lower-cost party.

The above mechanisms are *balanced*, in the sense that any payment made by one of the parties is received by the other. The literature generally restricts itself to balanced mechanisms; in part, this is done because *unbalanced* mechanisms seem at odds with what we observe in real life and, in part, because unbalanced mechanisms are too strong. The entire observable but unverifiable notion could be rendered completely meaningless, for instance, by using an unbalanced “shoot-them-both” mechanism: The judge asks the parties to simultaneously state what the observable variables are. If their reports agree, then the contract is enforced as required given the commonly reported realization of the variables. If their reports differ, then they are both shot (punished severely enough that neither side would willingly make a report that he or she knew differed from the other side’s). Truth telling is an equilibrium of such a mechanism and, among the many equilibria, is arguably focal.<sup>66</sup>

<sup>66</sup>Although truth telling is a focal equilibrium given no pre-mechanism communication, it is not clear that this mechanism would work if pre-mechanism communication were permitted. That is, if you hear the other party say on the way up the court steps, “I will announce  $X$ ,” then you would also have to announce  $X$  even if that is neither the truth nor an advantageous statement.

### 4.3.2 The holdup problem and renegotiation

As discussed in 2.5.2 above, the possibility of holdup helps explain duress and related limitations on contractual freedom. The holdup problem has also been the subject of intense research in the area of incomplete contracts.<sup>67</sup> This recent work has focused on the extent to which contract design can provide an alternative mechanism for addressing policing holdup in situations where judicial intervention is unavailable or impractical. As such, this research can be seen as a particular application of the question, considered in the previous section, “does incompleteness matter?”

An example illustrates the problem. One firm,  $A$ , is to develop a product that a second firm,  $B$ , will market ( $A$ , *e.g.*, is a studio and  $B$  is a film distribution company). Because of its expertise,  $B$  can observe the quality of the product  $A$  has produced; yet quality could be sufficiently amorphous that it would be difficult, if not impossible, to describe in a contract or verify in court unambiguously. Setting incentives correctly would, thus, seem problematic. If  $B$  commits to buy the product at a fixed price, then  $A$ , acting opportunistically, has no incentive to invest enough in producing a quality product. If the contract is a revenue-sharing contract, then  $A$  and  $B$ 's incentives could be too weak because of the teams problem (Holmstrom, 1982). Finally, if no contract is signed in advance, but the parties bargain after  $A$  has developed the product, then  $B$  could holdup  $A$ —that is, capture some of the value  $A$  creates by bargaining opportunistically—thereby adversely affecting  $A$ 's incentives.

Demski and Sappington (1991) propose option contracts as a solution to this problem. Specifically, let  $I \in \mathcal{I} \subset \mathbb{R}_+$  be  $A$ 's investment (in dollars, say) and  $\beta(I)$  be  $B$ 's value of the product as a function of  $A$ 's investment.<sup>68</sup> Assume for  $I$  and  $I'$  in  $\mathcal{I}$  that if  $I' > I$ , then  $\beta(I') \geq \beta(I)$ . Suppose that  $I^*$  is the optimal investment for  $A$  to make. Demski and Sappington's solution is for the parties to sign an option contract whereby  $B$  has the right to acquire the product at a strike price of  $\beta(I^*)$ . If  $A$  underinvests, then  $B$  won't exercise the option, so  $A$  loses. It is a waste for  $A$  to overinvest. Hence,  $A$  should invest the appropriate amount exactly and  $B$ , being indifferent between exercising or letting the option lapse, can be assumed to exercise in equilibrium. Any desired sharing of expected surplus can be achieved through non-contingent transfers.

The Demski and Sappington solution is, unfortunately, vulnerable to renegotiation (Edlin and Hermalin, 2000). Recall that  $B$  is indifferent between exercising its option and letting it lapse, but exercises in equilibrium. Recognizing that, at the exercise date, its payoff is thus independent of its decision,  $B$  should pursue the following strategy: Let its option lapse, but then approach  $A$

<sup>67</sup>A partial list of articles in this area is Bernheim and Whinston (1998), Che and Hausch (1999), Chung (1991), Demski and Sappington (1991), Edlin and Hermalin (2000), Edlin and Reichelstein (1996), Grossman and Hart (1986), Hermalin and Katz (1993), MacLeod and Malcolmsen (1993), Nöldeke and Schmidt (1995, 1998), and Rogerson (1992).

<sup>68</sup> $\beta(I)$  could be an expected value and it subsumes  $B$  having made the optimal marketing decisions conditional on  $I$ .

about renegotiating a deal at a lower price.<sup>69</sup> Because there would otherwise be no rationale for trade, it must be that the product is worth less than  $\beta(I^*)$  to  $A$ . Hence, there is a gain to the parties from negotiating a new price and, by the logic of the Coase theorem (see §2.2.2), they presumably will. Because the new price will be less than  $\beta(I^*)$ , this strategy of  $B$ 's dominates a strategy of simply exercising the original option.<sup>70</sup> As a rational player,  $A$  will anticipate this; that is, that it will ultimately be held up despite the option contract. Anticipation of holdup will adversely affect  $A$ 's incentives.

Some authors (*e.g.*, Bernheim and Whinston, 1998; Nöldeke and Schmidt, 1998), have sought to restore the capacity of option contracts to solve the holdup problem. Their approach has been primarily to allow the parties to structure the option contracts in such a manner that the contracts are robust to the sort of renegotiation described above.

Edlin and Hermalin (2000) are less sanguine than these other authors about the parties' ability to make their contracts robust to renegotiation. Accordingly, they ask whether option contracts could still work, even given the threat of renegotiation, if the strike price is lowered from  $\beta(I^*)$  to a more appropriate level? If the derivative at  $I^*$  of the product's value should it remain in  $A$ 's hands is greater than  $d\beta(I^*)/dI$ , then it is possible to achieve the efficient outcome using an option contract (with a strike price below  $\beta(I^*)$ ). If, however,  $d\beta(I^*)/dI$  is greater, then Edlin and Hermalin prove that no contract (option or otherwise) can achieve the efficient outcome. To summarize: Recognizing the possibility of renegotiation, if any contract can achieve an efficient outcome, then an option contract can; but there are circumstances in which no contract achieves an efficient outcome.<sup>71</sup>

The literature on holdup problems yields, therefore, a mixed message with regard to the importance of the distinction between observable and verifiable. Under a variety of conditions, the distinction is ultimately meaningless, but there are circumstances in which the distinction is important. It is worth noting that, in terms of information, this literature has considered a rather stark world—if a variable is unverifiable, then it is (often implicitly) assumed that there is no correlated signal that is verifiable. As discussed in the previous

---

<sup>69</sup>In a technical sense, this is *not* renegotiation insofar as the original contract has been honored— $B$  had the right *not* to exercise the option.  $B$  is, in a technical sense, engaging in new negotiations following the expiration of the original contract.

<sup>70</sup>This argument doesn't hold if the bargaining game in renegotiation gives all the bargaining power to  $A$ ; see Bernheim and Whinston (1998) and Edlin and Hermalin (2000) for differing views on the feasibility of (essentially) assigning the bargaining power to  $A$  *ex ante*. In addition, it could be welfare improving for the law to adopt rules that prevent renegotiation (Shavell, 2005); although, as noted in footnote 69 *supra*, technically this bargaining between  $A$  and  $B$  is not renegotiation, but new negotiations following the expiration of the original contract.

<sup>71</sup>When first-best efficiency is not attainable, Edlin and Hermalin (2000) show that a simple sales contract is the second-best efficient contract. Because a simple sales contract can be replicated by an option contract with a very high strike price, there is a sense in which the optimal contract is always an option contract.

subsection, were such signals to exist, then the circumstances in which this distinction mattered would be even further circumscribed.

#### 4.4 Legal doctrines addressing contractual incompleteness

Given the complexity of the foregoing theoretical discussion, it should not be surprising that a significant portion of the law of contracts deals in one respect or another with problems of interpretation. Such problems include: how to determine whether the parties have entered into a contract, how to determine the terms of any contract that is formed, and what to do if the contract does not explicitly or implicitly cover the situation that has arisen *ex post*. It is not feasible to review this entire body of doctrine here (for such reviews, see the sources listed in §1.3.2), but we can survey the main regulatory strategies that the law uses to deal with the economic problem of incomplete contracts. We begin by discussing three major themes that pervade the law in this area—the role of default terms, the dichotomy of form and substance, and the tension between objective and subjective modes of interpretation—and then move to a series of specific doctrinal rules that illustrate these themes in operation.

##### 4.4.1 Contract interpretation generally

THE FUNCTION OF DEFAULT RULES. In §2.1 above, we discussed the role played by contractual default rules in regulating market failures. The primary function of such rules, however, is to provide guidance for those interpreting contract terms on which the parties do not otherwise clearly agree. For example, in an ordinary sale of goods, the default rule is that the buyer must pay in cash at the time of delivery; if the parties wish to provide for credit or some other medium of payment, they must specify so in their agreement. Such default rules extend to virtually all aspects of contractual agreements, including remedial terms such as damages; and no legal system could operate without them.

The law and economics literature has taken three main approaches to the question of how default terms should be set. These approaches should be understood as reflecting different means towards achieving the goal of an efficient system for completing incomplete contracts, each of which is premised on a different assumption regarding what types of transaction costs are most empirically common.

One approach, following the work of Goetz and Scott (1985), argues that default rules should be chosen to provide terms that would minimize the cumulative transaction costs incurred by parties contracting around them.<sup>72</sup> In the special case where all parties face similar contracting costs and it is equally costly to contract around all terms, for instance, this implies terms that would be favored by a majority. If, conversely, some terms are costlier to contract out

---

<sup>72</sup>This argument is typically presented in intuitive form, but it could easily be formalized along the lines of §2.2.2 *supra*.

of or into than others (for example, if it is harder for the parties to specify multi-factor standards than simple rules), then other things being equal the default should be set to the terms that are easiest for the parties to escape (Schwartz and Scott, 2004). One might call this the transaction-cost-reducing approach to default rules, with the caveat that the underlying concept of transaction costs may be ambiguously defined.

A second approach recognizes that contracting costs will lead many parties to stick with a default rule even if it is not the one they would have chosen. Hence, this approach advocates choosing default terms for their substantive efficiency. Such an approach is most appealing in settings where there are substantial network externalities, adverse selection, or bounded rationality in contracting, so that individual parties are reluctant to depart from familiar and widely-used terms. For instance, Korobkin (1998a,b) presents evidence suggesting that cognitive and social-psychological factors lead parties to retain default (or standardized) contract terms that they would be better off contracting around; and Kahan and Klausner (1997) have argued that the network externalities are widespread in the corporate and business area due to the regular use of standardized forms, the value of which depends on a population of users who have invested in expertise in its application. This second approach might be called the central-command approach to default rules.

A third approach, following Ayres and Gertner (1989, 1992), argues that default rules should be used to encourage informationally advantaged parties to reveal their types before contracting, even if this means choosing terms that in equilibrium no one wants to use. They offer as an example the doctrine of *contra proferentem*, under which ambiguous contract terms are interpreted against the interest of the drafting party. Such an interpretation is unlikely to be what the drafter intended, they argue, but it serves as an incentive for the drafter to choose clear language that will convey to the counterparty the meaning of relevant terms. Ayres and Gertner, and much of the literature that follows, refer to such rules as penalty defaults, because they choose terms that operate as penalty for nondisclosure; an alternative and perhaps more precise terminology is to call this the information-forcing approach.

The concept of penalty default has been influential in the literature for its lucid illustration of the value of screening and signaling models in the analysis of contract law. Its general applicability in practice, however, is unclear from both positive and normative perspectives. From a normative standpoint, information forcing is only desirable under some circumstances; as we discuss in §3.3.3 and §5.2.7, it may be necessary to allow parties to keep the fruits of their informational advantage in order to induce them to acquire information initially. And even when it would be desirable to promote information revelation, a penalty default rule will fail to provide sufficient incentives in this regard if the value of the rents associated with informational advantage exceed the penalty. This can be the case if the informed party lacks bargaining power in a situation of bilateral monopoly (Johnston, 1990), if the information would reveal trade secrets valuable beyond the instant contract (Ben-Shahar and Bernstein, 2000), or if the asymmetric information is multidimensional and revelation along one

dimension would allow the recipient to draw inferences about another (Adler, 1999). From a positive standpoint, conversely, E. Posner (2005) has recently argued that most actual legal rules do not fit the penalty default model, basing his argument both on a survey of Ayres and Gertner's original examples, and on case law generally.

FORM VERSUS SUBSTANCE IN CONTRACT INTERPRETATION. The key policy question underlying contract interpretation is how thorough the interpretive process should be; and this question is commonly articulated in terms of the dichotomy of form and substance. Specifically, many legal rules have the effect of privileging certain interpretive materials and discounting others. Such rules are often termed formal or formalistic in that they confine attention to a subset of materials that may or may not give rise to the same inferences as would the universe of materials as a whole. A more substantive approach to contract interpretation, in contrast, would attempt to come to a more all-things-considered understanding, based on all of the materials reasonably available. For example, under the common law, the parol evidence rule (though subject to many exceptions) provides that a written document that integrates the parties' agreement may not be contradicted or varied by evidence of prior or contemporaneous oral understandings (E. Posner, 1998). In the absence of a writing, however, a party may introduce and a court can consider any information it wishes in order to determine the content of a contractual agreement.

From the economic viewpoint (Katz, 2004), the question of form versus substance can be assimilated to the problem of optimal information acquisition. A fuller or broader context can always be purchased, but at a cost of time, trouble, and interference with incentives, so it pays to stop at some optimal point. The problem can be formalized as follows: let the decision rule  $D(\cdot)$  denote a function that maps from a set of facts to a legal outcome, for example, an award of damages. A court's *ex post* interpretation of the facts will depend upon the information available to it, which we can denote by information set  $\mathcal{I}$ . (For example, a court that is aware of commercial trade usage will interpret the words of a contract differently from one that is not.) This information set depends on three potential sources: the information embedded by the parties in the contract itself (denoted as  $\mathcal{I}_0$ ), the information presented by the parties *ex post* at trial (denoted as  $\mathcal{I}_1$ ), and the information available to the court based on its general knowledge and experience ( $\mathcal{I}_a$ ). The information introduced by the parties in the contract or at trial is a variable that is either chosen in a cooperative game to maximize their returns from contracting or litigation, or the equilibrium outcome of a non-cooperative game in which they individually decide what information to introduce. Introducing information is costly, but the costs of *ex post* and *ex ante* information can interact, and more information can help the court reach a more efficient decision.

Within this framework, a legal interpretation regime consists of a function  $R(\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_a)$  that specifies how the interpreting agent combines the three possible sources of information (for example, the most anti-formalist would be that

the agent can consider all information available, in which case  $R$  would be the union function). A more formal regime would restrict  $R$  by limiting the influence of certain categories of information; for example, under the parol evidence rule, all information in  $\mathcal{I}_1$  that was inconsistent with  $\mathcal{I}_0$  would be thrown out. Conversely, a regime that disfavored standardized contracts on the theory that no one reads them would throw out any information in  $\mathcal{I}_0$  that was inconsistent with  $\mathcal{I}_1$ . The framework is quite general and can incorporate many types of evidentiary systems; for instance, in a pure adversarial system, taking to an extreme the practices of common-law regimes, the agent could not take account of any information in  $\mathcal{I}_a$  that was not confirmed by information introduced by the parties in  $\mathcal{I}_1$ . (Of course, the rule will affect the content of the information presented; if extrinsic information contradicting the written contract is ignored, no one will want to incur resources to present it.)

Given this framework, the expected legal outcome aggregating over all interpreting agents will be  $V = \mathbb{E}_a\{D(R(\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_a))\}$ . The parties will choose what information to embed in the contract, and to introduce at the time of a dispute, in order to maximize their individual returns in this outcome, and the lawmaker (who could be the parties themselves specifying in the contract what interpretation rule they wish to be followed) will choose the function  $R$  in order to maximize expected contractual surplus. From the standpoint of the lawmaker, this is just a constrained principal-agent problem, the solution of which depends on the particular costs associated with  $\mathcal{I}_0$  and  $\mathcal{I}_1$ , and the incentive properties of the decision rule  $D$  with regard to the primary behavior governed by the contract (here suppressed in the notation, and assumed to occur at an intermediate stage between the writing of a contract and the potential emergence of a dispute).

The considerations that determine the optimal approach to contract interpretation are thus quite broad-ranging. The regime of contract interpretation will influence contracting parties' behavior in many respects: with regard to decisions to breach, to take advance precautions, to mitigate damages, to gather and communicate information, to allocate risk, to make reliance investments, to behave opportunistically, and to spend resources in litigation, and so on. Accordingly, it is difficult to draw strong general conclusions regarding how interpretation should proceed.

If one is willing to make restrictive assumptions about the costs and benefits of information acquisition, however, it is possible to reach more specific conclusions. For instance, Schwartz and Scott (2004) develop a model in which parties benefit when their contract terms are interpreted correctly on average, but are relatively indifferent (due to risk-neutrality) to interpretive variance around the mean. In this case it would be optimal for courts to make interpretations on a minimum evidentiary base, because additional interpretative efforts are costly but provide no incentive benefits from an *ex ante* perspective. Conversely, Shavell (in press) assumes that *ex ante* contract-writing costs are positive and that it is possible to write at least some interpretative rules so that they can be applied costlessly (or more realistically, that the marginal cost of applying the rule is zero given that the parties are going to court anyway). In this case, it

will be optimal for the parties to leave at least some open contract terms to be filled in *ex post* by courts, and if contract-writing costs are sufficiently high, for courts actually to override the written terms in favor of another interpretation.

Obviously, such assumptions are special; in their absence, perhaps the best that can be said is that private parties should be allowed the leeway to choose their favored interpretative regime—a leeway not always recognized by the legal system. For public lawmakers, who likely lack detailed contextual knowledge about the costs and benefits of information, however, it is possible to offer some general rules of thumb. For example, formal interpretation is more efficient, other things being equal, when (1) *ex ante* negotiation costs are low relative to renegotiation costs; (2) either the parties or the tribunal are likely to be biased in interpreting the contract; (3) the chance of a dispute (or the *ex post* stakes in the event of a dispute) are relatively high; (4) legal outcomes are relatively sensitive to litigation expenditure; (5) specific investments must be undertaken by persons distant from the transaction, or have value that is relatively context-independent; (6) the parties have relatively good control over their contract-drafting agents relative to their negotiating agents; and (7) the parties have relatively good access to non-legal sanctions. In all these situations, the expected costs of filling contractual gaps *ex post* are low relative to the costs of filling them in *ex ante* so that it pays to undertake more effort in doing so gaps up front. Conversely, when renegotiation costs are high, the extent of bias is significant, and so on, the expected costs of *ex post* gap-filling are relatively low and it pays to defer more of this effort until an interpretative dispute arises. The logic here is analogous to Kaplow's analyses of the choice between rules and standards (1992) and the optimal complexity of rules (1995).

It follows from these heuristic principles that substantive interpretation is relatively more valuable to small and infrequent traders, who are less well-placed to undertake the fixed cost of detailed *ex ante* negotiation; who have relatively poor access to reputational networks *ex post*; who are likely to do their own contract negotiating but to contract out when acquiring legal services; who are less likely to be able to recover specific investments in substitute exchanges; and who possibly are less likely to face bias in *ex post* judicial tribunals. Conversely, large and experienced traders should prefer their contracts to be governed by relatively formalistic rules of interpretation. This distinction is consistent with casual empiricism—in general, it is experienced traders whom we observe contracting into relatively formal enforcement regimes through devices such as arbitration and forum selection clauses—and is also generally accepted in the case law and commentary.

**OBJECTIVE VERSUS SUBJECTIVE INTERPRETATION.** The default rule concept and the form-substance dichotomy both reflect the underlying problem of information acquisition and transmission, but focus on different dimensions of this problem. Specifically, the doctrines relating to form and substance are primarily concerned with the relationship between the contracting parties and the *ex post* interpreting tribunal; they govern the extent to which the burden of information

acquisition is allocated between the parties and the tribunal, and they affect the parties' incentives to transmit information about contractual meaning to the tribunal. Doctrines relating to default rules also implicate the relationship between parties and tribunal (especially when viewed from the transaction-cost-reducing approach of Goetz and Scott, 1985), but they additionally concern the contracting parties' informational relationship with each other (as emphasized by the penalty-default approach of Ayres and Gertner). These two dimensions of informational incentives interact in a third major theme of the law of interpretative doctrine: the tension between objective and subjective perspectives.

In this context, objective interpretation refers to the meaning that would be recognized by a reasonable outside observer, while subjective interpretation refers to the meaning as actually understood by the parties. For example, the case of *Lucy v. Zehmer* involved a dispute over whether the parties had agreed to the sale of the defendant's farm when, after a night of drinking, they signed a napkin that contained words of sale, referred to the farm, and stated a price. The defendant claimed that he was joking when he signed the napkin, and that the plaintiff either knew or should have known this. The plaintiff for his part insisted that the deal was a serious one. In the end, the court sided with the plaintiff, but rather than emphasizing the plaintiff's state of mind as its justification, it referred to publicly observable factors such as the duration of the negotiations, the existence of previous negotiations between the parties, and the reasonableness of the contract price.

In general, however, official doctrine does not give clear indication of when objective meaning controls and when subjective meaning controls. The Second Restatement of Contracts (1981), generally recognized as the most influential modern summary of American doctrine in the area, states that subjective meaning controls if the parties attach the same meaning to an agreement, or if the parties attach different meanings to the agreement and one (and only one) of them is or should be aware of the other's understanding. (For instance, if the drafter of a form knows that the form contains an unusual term and also knows that the counterparty is unaware of the presence of the term, and the counterparty is not otherwise charged with knowledge of the term, then the term is not part of the contract.) On the other hand, more traditional authorities reject these principles and insist that unreasonable meanings will not be enforced; and the Restatement itself elsewhere provides that when the parties attach different meanings to the agreement and there is no asymmetric information between them, various objective factors will be used to determine the contract's meaning.

The practical complexity of the doctrine arises from the fact that it is being used to regulate information transmission both between the parties (*i.e.*, the problem of observability), and between the parties and the court (*i.e.*, the problem of verifiability). For example, the Restatement's rule that contracts are to be interpreted according to the meaning of the asymmetrically uninformed party plainly operates as a penalty default rule, and can be justified either as forcing disclosure or as protecting the distributional interests of the uninformed. And conversely, the rule that charges parties with knowledge of meanings of

which they should be aware operates as an incentive for investigation.<sup>73</sup> But the objective-subjective distinction can also be understood in terms of the dichotomy of form and substance, in that an objective standard of interpretation directs the interpreter to limit its attention to factors that would be accessible to an objective observer, while a subjective interpretive standard directs the interpreter also to consider factors that might be accessible only to the parties to the contract. Thus, all the factors that bear on the optimal choice between formal and substantive interpretation (for example, the difficulty of litigating the parties' states of mind *ex post* as opposed to a counterparty's contrary meaning *ex ante* or the likely variance of error when making such assessments also bear on the choice between objective and subjective interpretation.

#### 4.4.2 Specific interpretative doctrines

PAROL EVIDENCE. In many cases, especially when substantial value is at stake, parties will memorialize their agreement in the form of a final written document. When they do, the intended effect with regard to pre-contractual communications can be ambiguous. For instance, suppose that, in oral negotiations, the buyer requests a clarification with respect to some technical feature of the goods and the seller provides clarification, but the final written document makes no mention of the issue. Should we regard the clarified term as implicitly incorporated into the final writing, or should we instead regard it as a rejected offer that the parties meant to exclude?

In common law systems, this question is addressed by a network of doctrines that are collectively referred to as the parol evidence rule. Under this rule, if the parties adopt a particular document as an authoritative statement of their contract—or in doctrinal language, as an integrated agreement—then the presumption is that they meant to reject all prior inconsistent communications. As a procedural matter, furthermore, the rule operates to exclude all evidence of antecedent understandings or negotiations that would vary or contradict the words of the document.

The parol evidence rule is subject to a number of exceptions and counterdoctrines (so much so that some commentators have suggested that it is misleading to refer to it as a rule); and the overall trend during the twentieth century has been toward decreased stringency in its application (typically, by finding, based on extrinsic evidence, that an apparently complete document was, despite superficial appearances, not intended by the parties to be authoritative). More recently, however, a number of commentators writing from an economically influenced perspective (*e.g.*, Scott, 2000; Schwartz and Scott, 2004) have argued for its re-invigoration.

Formal economic analysis of the rule, however, has been limited. Eric Posner

---

<sup>73</sup>Although, from a neoclassical economic perspective, it is unclear why it should be necessary to *provide* (strengthen) such an incentive. Presumably, private incentives should be sufficient (or even socially excessive) in this regard.

(1998) models the rule as posing a trade-off between Type 1 error (enforcing a term that is not part of the contract) and Type 2 error (failing to enforce a term that is part of the contract), as well as between transaction costs *ex ante* (e.g., contract-writing costs) and *ex post* (e.g., litigation costs). Posner's model should be understood as a special and more doctrinally detailed case of the general analysis of form and substance discussed *supra*. On his analysis, the rule makes little sense when courts process *ex post* information accurately and cheaply; it is justified when courts have poor ability to process *ex post* information and contract-writing costs are low. The implications for the more problematic situation where both contracting and litigation costs are high, however, are unclear and must presumably await future empirical investigation.

It is worth noting that the parol evidence rule does not apply to post-contractual communications, even though their use raises many of the same costs and benefits as pre-contractual communications do. Instead, such communications are analyzed using more flexible interpretative standards, and can operate as modifications or as waivers of contractual duty even when made informally.<sup>74</sup> It is possible that timing issues relating to renegotiation, holdup, and the increased value of *ex post* information could justify this disparate treatment, but at present this doctrinal distinction remains undertheorized from both legal and economic viewpoints.

TRADE USAGE AND COURSE OF DEALING. In contrast to the disfavored category of parol evidence, other categories of material are given specially privileged status when interpreting individual contracts. The most important of these is trade usage, which refers to any practice of dealing that is sufficiently regular and widespread in the relevant area or line of trade as to justify an expectation that it will be observed in the particular case.<sup>75</sup> For example, a contract for the sale of two-by-four wooden planks will ordinarily be read to refer not to planks that are literally two inches by four inches, but to the smaller  $1\frac{3}{4} \times 3\frac{1}{2}$  planks that generally pass under that description in the lumber business; a buyer who wants planks that are literally two by four must state so explicitly.

This doctrine can be justified in terms of both transaction cost reduction and information forcing. From the viewpoint of transaction cost reduction, if most contracting parties wish to follow a trade usage, it saves transactional resources for interpreters to read it into all contracts, so that only those who wish to disclaim it need undertake the extra costs of negotiating explicitly. From the viewpoint of information forcing, to the extent that one party subjectively intends to contract with regard to trade usage and the other does not, these differing understandings may lead them to enter into an inefficient exchange. Given that the trade usage is well evidenced and other usages are less so, re-

---

<sup>74</sup>In some settings (e.g., in sales contracts governed by Article 2 of the Uniform Commercial Code) it is possible for the parties to exclude oral modifications by using an appropriate clause in their original written contract, but it is not possible for them to exclude all possibility of *ex post* waiver.

<sup>75</sup>Restatement (Second) of Contract §222, Uniform Commercial Code §1-205.

quiring everyone to learn it offers a cheaper way of putting the parties on a similar informational footing.<sup>76</sup>

Such justifications are only persuasive, however, if trade usages can be identified cheaply and reliably, and if parties who wish to opt out of standard usage can do so at low cost. A number of recent commentators, however, have called both these assumptions into question. With regard to the first of these assumptions, Bernstein (1996, 1999) suggests that in many instances trade usages—in the sense of unwritten terms that are understood by the parties to impose actual legal obligations—do not actually exist, and that in any event, courts cannot tell the difference between behavior that evidences the existence of unwritten legal entitlements and behavior that instead represents settlement in the shadow of such entitlements. Also, Craswell (2000) has argued that courts often identify (or purport to identify) the actual content of any contested custom by consulting their own judgment, or the judgment of outside witnesses, about what custom would be most efficient. With regard to the second assumption, various commentators (*e.g.*, Scott, 2002) have argued that although contracting around trade usage is permissible under official doctrine, courts in practice imply so heavy a presumption against it so as to make trade usage (or more properly, what courts consider as trade usage) effectively mandatory.

The related doctrine of course of dealing, which refers to a sequence of conduct between the parties to an individual contract, raises similar problems. Whether a pattern of bilateral behavior amounts to a legal understanding, however, is even more difficult to determine than with group behavior because such patterns tend to be less well evidenced and because the participants may wish to recharacterize their past behavior with an eye toward influencing the outcome of a current dispute. In addition, as Ben-Shahar (1999) has argued, allowing course of dealing to influence the interpretation of future contracts can lead parties to be excessively rigid when demanding contractual performance, for fear that greater flexibility will prejudice their future rights.

Such functional problems can be understood as stemming from the basic problem of contractual incompleteness, with the extra twist that it is not just the individual contract terms that are unverifiable, but also the alleged trade usage or course of dealing. Accordingly, whether courts should read trade usage and course of dealing into contracts as a default rule, or whether parties should make greater efforts to disclaim such a reading, remain open questions.

**CHANGE OF CIRCUMSTANCE.** The doctrines of trade usage and course of dealing are used to fill out incomplete contracts in routine circumstances. In contrast, another set of doctrines are used to imply missing terms in unusual circumstances—specifically, when unforeseen events intervene to reduce substantially one or both parties' gains from trade under the original contract. For example, in *Taylor v. Caldwell*, a landlord whose building was destroyed in a fire was excused from liability to a promoter who had leased the premises for

---

<sup>76</sup>See Hermalin (2001) for a brief survey of research on the efficiency gains to be had from terminological conventions.

a theatrical event, even though the lease contract contained no explicit excuse of this sort; and in *Krell v. Henry*, a royalist socialite who had leased, at inflated short-term rates, a flat overlooking a public square in which the incoming British king was to be crowned was excused from the obligation to pay rent after the coronation was postponed due to the king's emergency appendectomy.

When this principle is used to excuse a producer or seller, it is often referred to as the doctrine of impossibility, although this term is a misnomer in that (1) the doctrine is often applied to situations in which it is highly burdensome but not impossible for the seller to perform, and (2) even if performance is not physically possible, the payment of damages is. Other labels include impracticability, commercial impracticability, and (especially when the buyer is excused, as in *Krell v. Henry*) frustration of purpose.

An analogous principle is sometimes used to excuse parties from complying with conditions that turn out to be unexpectedly burdensome, so long as the excuse does not unduly disadvantage the counterparty. (For example, a condition in an insurance contract requiring all claims to be made within two weeks of the occurrence of a casualty would typically be excused if the delay was caused by the insured party's being incapacitated in the underlying accident.) Finally, as indicated in §1.3.2 above, parties are sometimes excused from liability under the doctrine of mistake when they contract without the benefit of information that, once revealed, substantially alters the contract's efficiency or distribution of surplus. As observed above, from an information-theoretic perspective this situation is functionally equivalent to the case of *ex post* changed circumstances.

Economically influenced commentators have most commonly justified such doctrines on transaction-cost-reduction grounds. Specifically, Posner and Rosenfield (1977) suggest that, insofar as the doctrines apply only to excuse a party who neither has hidden information nor takes a hidden action with regard to the excusing event, they mimic the risk allocation terms that would most likely be provided in a complete contingent contract. (For instance, in *Taylor v. Caldwell*, given that the theater owner had already suffered the loss of a building, he was unlikely to be the least-cost insurer of the promoter's losses, at least in the absence of any specific facts evidencing moral hazard or adverse selection.)

The changed-circumstances doctrines can also be understood as special applications of the penalty-default and form-versus-substance frameworks outlined above. From a penalty default perspective, a contingency that could reasonably have been anticipated *ex ante* by one of the parties should not count as an excuse, because denying the excuse will encourage the informed party to bring the issue into the contractual negotiation. Conversely, from a form-substance perspective, contingencies that are easier to specify and ascertain *ex post* either because information is better after the fact or because their probability is too low to warrant up-front attention, are appropriately left off to be dealt with by a court in the unlikely event that they arise.

It should be noted that the changed-circumstance doctrines can also be understood as special applications of the theory of optimal contract remedies along the lines of the discussion in §5 below. For instance, White (1988) points out that excusing a party for changed circumstances has consequences not just

for risk allocation, but also for various countervailing considerations such as incentives for precaution and relational investment. She argues that it is only in rare cases that zero damages are second best, and accordingly that changed-circumstance cases should be treated like any other contractual breach. On the other hand, once the costs of legal enforcement are taken into account, excuse from liability may be warranted as a litigation-saving device to be applied in extreme situations where it is sufficiently clear that the contract should not be performed and that breach is not due to any failure of precaution or disclosure (in the same way that negligence liability in tort saves administrative costs by requiring litigation only in case of a breach of duty).

**STANDARDIZED CONTRACTS.** Most contracting parties do not negotiate individual contracts for each occasion on which they enter into exchange; instead they use standardized forms that incorporate customary or boilerplate terms, and negotiate only over those few terms (such as price and quantity) that are essential for their transactional purposes. From an economic viewpoint, such behavior is a straightforward response to scale economies in the production of contractual terms. In legal circles, however, standardized contracts are regarded with more ambivalence. While most legal commentators accept that such contracts are a practical necessity, many continue to regard them as problematic on grounds of autonomy (because most people do not read or understand other parties' standard forms and hence cannot knowingly assent to their terms) or distributive justice (because the party who drafts the form is deemed to hold an inequitable degree of bargaining power).

At an abstract level, standardized contracts are interpreted using the same methods as contracts more generally. For instance, if the drafting party knows that a particular term in the contract is unusual, that the non-drafting party is unaware of the term, and that the non-drafting party would not enter into the contract if aware of the term, then the term is not part of the contract; as observed in the discussion of objective versus subjective interpretation above, this result follows from ordinary principles regarding subjective interpretation. At a practical level, however, most lawyers understand form contracts to constitute a special category subject to specialized rules of interpretation, such as the doctrine that contracts are to be construed against the interests of the drafter. Conventional legal wisdom holds that form-contract terms are less likely to be enforced than individually negotiated terms, especially when they are unusual or thought to be especially burdensome to the non-drafting party.

There is relatively less reason to be concerned about standard terms on market power grounds, notwithstanding many legal commentators focus on this issue. Because standardization lowers the per-unit cost of contracting, both competitive firms and those with market power will find it of value. While firms with market power might distort non-price terms from the efficient level, the direction of distortion is *a priori* ambiguous (Spence, 1975). If all consumers have the same willingness to pay for contractual terms, a monopolist will do best to provide optimal terms and to extract available profits through a high

price; and if consumers differ in their willingness to pay for contractual terms, super-optimal terms are as likely as suboptimal ones. While oligopolists could use standard forms to collude on non-price terms, it is similarly unclear in which direction their incentives cut. Accordingly, to the extent standard form contracts raise competition concerns, these are best policed by antitrust law, rather than by common-law courts applying the law of contract.

Standard form contracts do raise informational concerns, however, since they can vary substantially in their terms and the drafting party knows much more about the terms than does the non-drafting party. If contract reading is costly, parties may rationally assume that a particular contract contains the average terms available on the market, with the result being adverse selection along the lines of Akerlof's market for lemons (see also the discussion in §2.3.4). In addition, as Katz (1990b) has shown, if contract reading is a relation-specific investment (as it will be if different drafters offer different terms), drafters will be tempted to expropriate this investment by choosing terms that are just favorable enough for the non-drafter to accept after having read them, thus undermining any incentive to read. Interpretative rules that make it harder for drafters to enforce unusual standard terms without calling them to the other party's attention, such as the doctrine of *contra proferentem* or the rules requiring warranty disclaimers to be conspicuous, help to address this informational asymmetry, by putting the burden of communication on the party who can undertake it most cheaply. Such rules provide perhaps the clearest illustration of the penalty default approach discussed above.

**IMPLIED DUTY OF GOOD FAITH.** Finally, most modern-day courts will require the parties to a contract to exercise good faith in the performance and enforcement of their contractual duties. Commentators disagree on exactly what this obligation requires, but as a general matter it prohibits opportunistic actions that, while complying with the bare letter of the contract, depart from it in spirit by operating to deprive a counterparty of the reasonably expected benefit of its bargain. For example, it has been held a breach of good faith for a buyer of goods to seize on an otherwise trivial defect in delivery in order to escape an unfavorable contract in a falling market, for an employer to exercise its termination rights on an at-will labor contract just before the employee's accumulated sales commissions were due to be paid, or for a realty purchaser to prevent his seller-broker from acquiring the property to be delivered under the contract (and thus forcing the seller into breach) by showing up at an initial land auction and outbidding him.

In theory, the parties could guard against opportunistic exercises of discretion by providing specific limits in their original contract. For all the reasons discussed above in §4.2, however, they often do not do so, leaving courts with a disagreeable choice between imposing additional *ex post* limits on an *ad hoc* basis, and countenancing an unanticipated and perhaps unfair result. Traditional common-law courts often responded to this dilemma by refusing to enforce such agreements entirely, on the theory that such agreements were so one-sided as to

lack formal consideration. Modern courts, in contrast, are more willing to treat such contracts as genuine bargains, and to police party opportunism directly and substantively after the fact. As our earlier discussion of form versus substance indicated, such an approach could serve the parties' economic interest if the courts' costs and error rate in filling contractual gaps *ex post* are sufficiently low, or if the parties' costs of dealing with them *ex ante* are sufficiently high.

Whether the duty of good faith is best viewed as a rule of interpretation, or alternatively as a limitation on contractual freedom, has not always been clear from the case law or commentary. The fact that the scope of the duty depends on other terms of the agreement, as well as on the overall commercial context, suggest that it is an interpretative rule; but the fact that the parties are not permitted to disclaim it (though they are permitted to stipulate its content so long as they do not do so in a "manifestly unreasonable" way) suggests that it is a regulatory intervention. The uncertainty is reinforced by the murky boundaries of the duty, which in limiting cases appears to shade into other mandatory doctrines such as unconscionability and public policy.

What counts as opportunistic behavior, however, typically lies in the eye of the beholder, and critics of the modern duty of good faith (*e.g.*, Schwartz, 1992; Bernstein, 1999; Scott, 2000) argue that the behavior that it seeks to induce is typically unverifiable, and thus not effectively subject to judicial oversight. Courts' attempts to enforce good faith, accordingly, are as likely to depart from contractual expectations as to enforce them, in the end just resulting in an increase in the cost of litigation and in the variance of judicial outcomes. On this view, enforcement of such "soft" contractual expectations should be left to private enforcement mechanisms such as reputation and repeat dealing. For the most part, however, the courts have not yet accepted this critique.

#### 4.5 Overall assessment of the law of contract interpretation

In general, the economics of contract interpretation is a relatively unmined field compared to the economic analysis of contract law generally. While the general contours of legal understanding in this area are more or less consonant with the main insights of the economic theory of contract, the rapid and recent development of the economic literature has not yet been matched by a corresponding growth in legal scholarship. It is likely, accordingly, that this set of topics will draw increasing attention from law and economics scholars in upcoming years.

### 5 Enforcement of contracts

As we observed at the outset of this chapter, the rationale for contracting is to lock in a commitment *ex ante* that one or both parties would otherwise not wish to honor *ex post*. The use of a contract to establish such commitment is undermined, of course, if the contract will not be enforced in the way the parties anticipate.

The enforcement of contracts is often the province of courts, who impose sanctions on parties who violate or breach a contractual obligation. As we discuss below, contracts can also be enforced by the parties themselves, through the use of various self-help remedies. We begin, however, with a discussion of judicial enforcement, partly because this has been the focus of most of the existing literature, and partly because an understanding of judicial enforcement is useful in order to understand when and why private enforcement might be more attractive.

## 5.1 General issues in enforcement

### 5.1.1 Remedies and contractual incompleteness

In one sense, every legal dispute is about remedies. Even when a court rules that no contract was ever formed, that ruling can be seen as saying that the complaining party is not entitled to any remedy for breach of contract. Moreover, such a ruling might still entitle the parties to some other legal remedy—for example, in some circumstances, one or both might be required to return any advance payments that had been made. From an economic standpoint, what ultimately matters is not the doctrinal question, “was a valid contract formed?” but rather the remedial question, “how much money can I now collect?” (or “how much money will I now be required to pay?”)

In another sense, though, the identification of certain actions as a breach of contract, and the identification of certain payments as remedies, is an artifact of contractual incompleteness. After all, in a hypothetical complete contract, the contract itself would specify the payments and other actions that would be required in every possible state of the world. With such a contract in place, there would be no reason to label some actions in some states of the world as complying with the contract, while labeling other actions in other states as non-complying or breaching. There would also be no reason to label some (but not all) of the associated payoffs as “remedies for breach.”

As discussed in §4.2 above, however, complete contingent contracts are rarely observed in the real world. In particular, real contracts are often incomplete in two ways that are relevant here. First, rather than specifying every possible action that a party might take, they may specify only those actions by each party that will suffice to comply with the contract—for example, “seller to deliver fifty widgets by July 1; buyer to pay \$1,000 by July 10.” Second, they may fail to specify the steps that should be taken if either party fails to take a complying action—for example, if the seller delivers only forty widgets, or delivers them on July 15; or if the buyer refuses to pay.

Parties sometimes do provide explicit terms that specify the remedy for some forms of noncompliance; these terms are known as “liquidated damage clauses” and they will be discussed in §5.3.4. In other cases, however, where the parties’ contract fails to specify the consequences of noncompliance, the legal system will supply a *default* remedy, just as it supplies default rules for other questions that a contract might not address (see §4.4). As most of the literature has focused

on these default legal remedies, that is where we begin.

### 5.1.2 Overview of default remedies

In some cases, parties who fail to comply with their contracts may be ordered by courts to perform—in legal terms, specific performance (see §5.3.3 below), with the order backed up by threats of more severe sanctions for contempt of court. In most cases, though, the default remedy for breach has the breaching party pay money to the aggrieved party.

Following Fuller and Perdue (1936, 1937), it has become customary to identify the most common remedies as expectation damages, reliance damages, or restitution damages. *Expectation damages* attempt to put the injured party in as good a position as he would have been in if the contract had been performed. For example, if a buyer contracts to pay a price  $p$  for a good that has a gross value of  $v$  to him, but whose value can be realized only if the buyer spends  $r$  to make use of the good, full performance of the contract would leave the buyer with a net value of  $v - r - p$ . If the seller then breaches the contract, by failing to deliver the good, expectation damages would require the seller to pay enough to leave the buyer in that same net position:  $v - r - p$ . For example, if the buyer had already spent both  $p$  and  $r$ , expectation damages would require a payment of  $v$ .

By contrast, *reliance damages* require the seller to pay only enough to leave the buyer with the level of utility he would have enjoyed if the contract had not been signed. For example, if the buyer in this case would have had a net gain of zero had he not signed this contract, reliance damages would require the seller to pay enough to bring the buyer back to that position. So, for example, if the buyer had already spent both  $p$  and  $r$ , reliance damages would then equal  $p + r$ , enough to bring the buyer back to zero.

Another standard remedy, *restitution*, allows both parties to recover the reasonable value of whatever they gave to the other party. In this example, restitution would allow the buyer to recover damages of  $p$ .

As this example shows, these three remedies can often be ranked by size, with expectation damages usually providing the largest remedy and restitution giving the smallest. (We may assume  $v > p + r$  because otherwise the buyer would not sign the contract.) Similarly, reliance usually exceeds restitution, because the buyer's reliance typically provides little benefit to the seller. Of these three remedies, expectation damages are conventionally regarded as the predominant measure in both theory and practice, with reliance and restitution reserved for cases in which there is some defect in the bargain or in the plaintiff's ability to prove it, or when equitable considerations justify a departure from the ordinary rule. In practice, however, this generality does not always hold, due to measurement difficulties and to other aspects of the legal rules governing the availability of each remedy.

For example, expectation damages entitle a plaintiff to compensation for profits that would have been earned on the breached contract, but calculating hypothetical revenues and expenses on a cancelled project is often a speculative

endeavor. Accordingly, courts often presume that the future profits to be earned on the contract approximate the costs so far incurred, thus using reliance damages as a proxy for expectation damages. Conversely, reliance damages should in principle compensate a plaintiff for the loss of forgone alternatives, but such opportunity costs are often difficult to verify *ex post*. As a result, courts often presume that the forgone opportunity would have yielded the same profits as the breached contract—in which case expectation serves as a proxy for reliance. Similarly, it is often difficult to verify the value of restitution, especially when it takes the form of services not traded on any market. Thus, courts sometimes measure restitution in terms of the reliance costs incurred by the non-breacher, though their willingness to do so may depend on their perceptions of relative fault. There are even situations in which the non-breacher is permitted to choose among expectation and restitution, or among expectation and reliance; obviously, in those cases the more generous remedy will be chosen.

In addition, as we discuss below, each of these remedies can be modified or adjusted in various ways: for example, if the victim could have prevented some or all of the losses by appropriate mitigating behavior. As we also discuss below, there are a few additional remedies that are occasionally imposed following a contract breach, including punitive damages in rare cases.

In part for these reasons, it is often more helpful to think of a monetary damage remedy as a continuous choice variable  $d$ , whose value can in principle be set by the legal system to any level. In some cases, the damages awarded may more appropriately be written  $d(\mathbf{x})$ , where  $\mathbf{x}$  is a vector representing any number of actions taken by one or both of the parties. Put less formally, remedies for breach can affect the parties' incentives both indirectly, just by threatening a party with having to pay some amount  $d$  in certain states of the world; and more directly, by conditioning the size of  $d$  on some particular behavior  $x$  that the law seeks to influence. In general, though, damage awards that are conditioned on particular actions will often be more difficult for courts to implement, depending on the ease with which those actions can be verified by courts.

## 5.2 Monetary damages for breach of contract

We now turn to the effects of monetary remedies of various sizes and descriptions. Obviously, larger monetary remedies increase the non-breacher's payoff (and reduce the breacher's payoff) in the event of a breach, while smaller remedies produce the opposite effect. However, this effect by itself is merely distributional, and will not by itself change the transaction's expected value. As long as both parties correctly estimate the probability of a breach, the prospect of liability for higher (or lower) damages can be offset by charging a higher (or lower) price, leaving both parties with the same expected return.

Instead, changes in the expected value of the transaction must come from some other cause. For example, if one or both parties are risk averse, there can be welfare gains from shifting more of the variance in payoffs to the less risk-averse party. We discuss this possibility in §5.2.9 below.

More important, in most cases the expected value of the transaction depends

in part on various actions the parties might take to increase (or reduce) that value. Moreover, the remedy for breach will usually affect the parties' incentive to take these actions—for example, their incentive to perform the contract rather than breaching, or to take precautions against contingencies that might lead them to breach, or to make transaction-specific investments whose value will be lost if there is a breach. Each of these incentive effects is discussed below.

### 5.2.1 The breacher's decision to perform or breach

Suppose that a seller must decide whether to perform a contract, at a cost  $c$ , when performance will confer on the buyer a value  $v$ . Total welfare is maximized if the seller performs when and only when  $v \geq c$ . Otherwise, it would be more efficient for the seller not to perform, an outcome often referred to as “efficient breach.”<sup>77</sup>

To be sure, if  $v$  and  $c$  are known at the time of contracting, the parties would have no reason to contract unless  $v$  exceeded  $c$ . In many situations, however,  $c$  or  $v$  (or both) are stochastic variables whose values will not be realized until just before performance is to take place, long after the contract was signed. Accordingly, one problem of interest is to design remedial rules that will give the seller an incentive to perform the contract if and only if  $v \geq c$ .

As the earliest contributions to this literature recognized, a remedy of expectation damages creates exactly this incentive, as long as those damages are accurately measured.<sup>78</sup> As defined above, expectation damages force the seller to internalize whatever losses the buyer suffers from the seller's breach. Consequently, the seller has an incentive to breach only if her gains from breach exceed both parties' losses. (A symmetric analysis, which we omit here for the sake of brevity, could be applied to the buyer's incentives to perform or breach.)

Before turning to other relevant incentives, four points should be made concerning even this simple result. First, while the expectation remedy forces the seller to internalize the buyer's losses from breach, it does not necessarily give the seller any incentive to consider effects that her performance or breach might have on third parties. As with most of the literature, we restrict attention to cases where third-party effects are absent (but see §2.3.1 and §5.3.4).

Second, like any other remedy, expectation damages will affect the seller's incentives only to the extent that the seller actually expects to pay those damages. If, instead, there is a significant chance that the seller could escape having to pay, the seller's incentives would be correspondingly reduced. As we discuss

---

<sup>77</sup>The term “efficient breach,” though widespread in the literature, is somewhat of a misnomer, and has had the unfortunate effect of leading legal scholars untrained in economics to suppose that there is a tradeoff in this regard between efficiency and the deontological value of promise-keeping. From the perspective of a hypothetical complete contract, calling off performance when  $v < c$  (possibly combined with some side payment) is precisely what the parties would have wished to specify, so a legal default rule that results in that outcome is better understood as promoting the parties' promissory intentions than subverting them. The term “efficient implied cancellation option” would be more accurate, if less arresting.

<sup>78</sup>See Birmingham (1970) and Barton (1972). The first formal model is Shavell (1980).

below (§5.4), the costs and other difficulties of pursuing a lawsuit may sometimes make this legal remedy less effective.

Third, if courts instead award some higher or lower measure of damages—or, equivalently, if they attempt to award expectation damages but predictably err in measuring those damages—the seller’s incentives to breach will be altered. If courts tend to award damages that are higher than the true measure of expectation damages, the seller will have an even stronger incentive to avoid breaching: too strong an incentive, relative to the “efficient breach” condition described above. If courts instead tend to award lower damage measures, as is widely believed to be the case, the seller will have a weaker incentive to avoid breaching. The possibility of judicial error is particularly likely if some or all of the buyer’s benefits from performance are either unobservable or nonverifiable (see §4.2 above).

Indeed, if courts instead could observe every relevant variable without error, it would be trivial to create efficient performance incentives simply by having the courts evaluate the efficiency of every breach *ex post*. If the court decided that a breach was efficient, the breacher could be excused from any remedy (or even rewarded with a bonus), while if the court decided a breach was inefficient, the breacher could be hit with huge sanctions. Viewed from this standpoint, the potential benefit of using expectation damages to create efficient incentives for performance or breach is that this remedy does not require courts to be able to evaluate the efficiency of any particular breach. The expectation remedy does, however, require courts to be able to calculate the amount the buyer would have gained from performance. How often courts are able to make such calculations is a matter of some dispute.

Fourth, and most important, the incentives to perform or breach may not even matter as long as the parties can renegotiate after the values of  $v$  and  $c$  have been realized. If this *ex post* renegotiation is possible, then—regardless of the legal remedy for breach—there should always be an agreement that will maximize both parties’ gains by performing the contract if but only if performance is efficient (Shavell, 1980). To be sure, renegotiation may entail positive transaction costs, but most legal remedies (including expectation damages) also entail positive transaction costs. While some of the early literature attempted to identify the legal remedy that would satisfy the “efficient breach” condition with the lowest possible transaction costs,<sup>79</sup> empirical evidence on actual transaction costs is largely nonexistent, and that literature is best described as inconclusive.

As discussed earlier, the possibility of *ex post* renegotiation also plays an important role in the literature on incomplete contracts (see §4.2.6 *supra*). This similarity should not be surprising—for as noted earlier—remedies for breach are really just one aspect of contractual incompleteness. As we discuss below, the possibility of *ex post* renegotiation also complicates the economic analysis of other incentives created by legal remedies.

---

<sup>79</sup>See, *e.g.*, Schwartz (1979), Macneil (1982), and Bishop (1985).

### 5.2.2 The breacher's decision to take precautions

When the seller's costs  $c$  are stochastic, the distribution from which  $c$  is drawn might be given by nature, entirely independent of any action by the seller. More commonly, though, that distribution is itself a function of the seller's investment in the transaction. For example, a seller's costs might depend in part on how often her assembly line malfunctions; and the probability of such a malfunction might depend on how much the seller spent on maintenance. Typically, the seller must choose her expenditure on maintenance at one point in time; only then, after that expenditure has been made, the actual value of  $c$  will be realized.

Expenditures such as these are often referred to as "precautions" (see generally Cooter, 1985). The optimal expenditure on precautions can be defined straightforwardly in terms of (a) the cost of each possible precaution, and (b) its marginal effect on the distribution from which  $c$  is drawn, together with the welfare associated with each possible realization of  $c$ , taking account of the fact that some of those realizations will result in the contract being performed, while other realizations will result in the contract being breached. Given the usual assumption of diminishing marginal returns, there will be a unique expenditure on precautions that maximizes the total expected value of the transaction.

Interestingly, expectation damages (if measured accurately) can give sellers an incentive to choose this optimal expenditure on precautions (Kornhauser, 1983). As discussed above, expectation damages force the seller to internalize all of the buyer's losses from any realizations of  $c$  that result in a breach. As a consequence, expectation damages can in principle optimize both of the incentives discussed so far: (a) by giving the seller an incentive to choose the optimal expenditure on precautions; and (b) once the actual value of  $c$  is realized, by giving the seller an incentive to choose between performing and breaking the contract. Again, any remedies that are systematically less than expectation damages—whether by design, or because of measurement error—should reduce both of these incentives, while any remedies that systematically exceed expectation damages should increase both incentives.

To be sure, if courts could observe the seller's actual expenditures on precautions, and if they could also observe all the factors necessary to calculate the optimal expenditure, there would then be other ways to give sellers an incentive to choose the optimal level. For example, courts could excuse from liability any seller whose precautions were optimal, in much the same way that negligence standards in tort law excuse defendants whose precautions were optimal. Alternatively, if the contracting parties could specify in their contract the optimal expenditure on precautions, that requirement could itself be enforced by a court, as long as the court could verify the seller's actual expenditure. Thus, just as in the case of incentives for performance (§5.2.1), the case for using expectation damages to optimize a seller's incentives to take precautions rests in part on the assumption that it is easier for courts to observe only those factors necessary for the calculation of expectation damages (*i.e.*, only those factors that go to the value that the buyer would have received if the contract had been performed)

than it is for courts to observe any of these other factors.<sup>80</sup>

Alternatively, if buyers can observe a seller's precautions (before contracting), competition among sellers could give them a market incentive to choose the optimal expenditure on precautions (Kornhauser, 1983). Even if sellers do not choose their precautions until after contracting, optimal market incentives could be possible if buyers can observe sellers' reputation for taking precaution. In §5.4 below, we consider reputations and competitive markets in connection with enforcement by means other than legal remedies.

### 5.2.3 The non-breacher's reliance decision

The value the buyer places on performance by the seller is often a function of the buyer's investment in the transaction. For example, a business that is buying a new machine may get more value from that machine if it spends money training its employees to use it. Expenditures such as these are often referred to as reliance on the contract.

By definition, reliance expenditures are at least partially transaction-specific—for example, training designed for one particular machine may be worthless if that machine is not delivered. Although there can often be varying degrees of performance, for concreteness assume that there are only two discrete possibilities—either the contract is performed, or it is breached. Let  $v(r)$  represent the value that the buyer will receive from performance conditional on his reliance investment  $r$ , while  $w(r)$  represents his value if the contract is not performed. If performance of the contract is certain, the optimal reliance expenditure is simply the value that maximizes  $v(r) - r$ . In the more usual case, where the contract should be performed in some states of the world only, the optimal reliance expenditure is the value that maximizes

$$qv(r) + (1 - q)w(r) - r, \quad (9)$$

where  $q$  is the probability that contract should be performed (here taken to be independent of  $r$ ; *e.g.*, with probability  $1 - q$  some “act of God” makes performance prohibitively expensive for the seller).

The remedy of expectation damages ( $d = v(r) - w(r) - p$ , in this example) will generally not give the buyer the right incentive to choose the optimal value of  $r$ . Expectation damages are a “full insurance” remedy that gives the buyer the same net return in every state of the world; that is, the buyer's net return is  $v(r) - p$  under performance and  $d + w(r) = v(r) - p$  under non-performance. Unless  $v'(r^*) = w'(r^*)$ , where  $r^*$  maximizes expression (9), the buyer will choose a non-optimal level of reliance. If one assumes, as is often true, that the marginal return from reliance investment is greater given performance than non-performance (*i.e.*,  $v'(r) > w'(r)$ ), then buyers have excessive incentives to rely under expectation damages.

Reliance damages ( $d = r - w(r)$ , in this example) also will fail to provide the buyer the proper incentives. If  $v'(r) > w'(r)$ , then  $w'(r^*) < 1$ ; in this case,

<sup>80</sup>For a discussion of this aspect of expectation damages, see Cooter (1985).

reliance damages will encourage the buyer to rely excessively because increased expenditures on  $r$  will increase the damages they can collect.

More generally, the probability  $q$  is endogenous insofar as the level of damages (either expectation or reliance) affects the probability that the contract will in fact be performed (see §5.2.1 and §5.2.2). Even with an endogenous  $q$ , however, neither reliance nor expectation damages will generally provide the proper incentives. To see this, observe that the first-best solution requires performance whenever  $v(r) - c \geq w(r)$  and, therefore, that investments maximize

$$\int_0^{v(r)-w(r)} (v(r) - c)f(c|s)dc + \int_{v(r)-w(r)}^{\infty} w(r)f(c|s)dc - r - s, \quad (10)$$

where  $s$  is the seller's investment in precaution and  $f(c|s)$  is the density of cost given such investment. To replicate the first-best solution with regard to whether the seller performs, it must be that  $p - c \geq -d$  if and only if  $v(r) - c \geq w(r)$ ; hence,  $d = v(r) - w(r) - p$ , which accords with expectation damages, but not reliance damages. However, as we saw above, expectation damages causes the buyer to face the maximization problem  $v(r) - r$  with respect to his choice of reliance. As discussed, the  $r$  that solves the buyer's problem will differ from the  $r$  that maximizes social surplus (*i.e.*, expression (10)).<sup>81</sup>

If and only if *ex post* renegotiation is impossible, a remedy of no damages at all—*i.e.*, a regime of no liability for breach—can optimize the buyer's reliance incentives in a second-best sense (given the probability of performance), because such a regime forces the buyer to bear all of the downside as well as all of the upside of any reliance expenditure (Shavell, 1980). To be sure, such a remedy does little to optimize the seller's incentives, as discussed earlier in §5.2.1 and §5.2.2. But if the breach probability (*i.e.*,  $q$  above) is exogenous or the seller's incentives can be optimized independently (*e.g.*, through market-reputation effects), a zero-damage remedy could be optimal. This is not wholly surprising, as this forces the buyer to face the social planner's problem as given by (9) above.

Mathematically, zero is just a constant and, more generally, any constant-damage measure—that is, any measure whose value does not change with the buyer's level of reliance—will also optimize the buyer's reliance incentives (for any given probability of performance), as long as *ex post* renegotiation is still impossible.<sup>82</sup> Under a constant-damage measure, the buyer will still capture all of the marginal benefits, and bear all of the marginal costs, of higher or lower reliance expenditures. Moreover, a constant-damage measure could also optimize the seller's incentives, if the constant is set equal to the value that performance would have to a buyer who in fact relied optimally:  $d = v(r^*) - w(r^*) - p$ , in this example. In addition to optimizing the buyer's reliance expenditures, this

<sup>81</sup>See Shavell (1980) for greater details. Consider also Shavell (1984), Kornhauser (1983), and Rogerson (1984).

<sup>82</sup>Shavell (1980) considers a model in which restitution is a constant-damage measure.

will also give the seller optimal incentives, both to perform or breach and to choose an optimal level of precautions.<sup>83</sup>

However, as noted, this result holds only when *ex post* renegotiation is impossible. When renegotiation is possible, reliance by the buyer can have the additional effect of altering the terms likely to be reached in any renegotiation, thus further distorting the buyer's reliance incentives (recall the discussion of holdup above in §4.3.2). Under certain conditions, holdup concerns may bias the buyer's incentives toward choosing too little reliance (Rogerson, 1984).

Edlin and Reichelstein (1996) show that, in some circumstances, the parties may be able to choose contract terms that balance the risk of holdup against the risk of moral hazard, thus implementing a first-best solution (recall, too, the discussion in §4.3.2). For example, suppose the parties enter into a fixed-price contract in which they can freely adjust the quantity term, and again, suppose that the buyer must rely before he knows whether the seller will perform. If the contract quantity is set at zero (*i.e.*, if there is no enforceable contract at all), the buyer will under-rely because part of the value created by reliance will be expropriated through *ex post* holdup. On the other hand, if the contract quantity is set at the level  $Q^*$  that the buyer wishes to purchase, he will be fully insured against non-performance, and will over-rely. By continuity, there exists some contract quantity  $\hat{Q} \in (0, Q^*)$  that will provide optimal reliance incentives. In the event that the seller can perform, the parties can then make up the difference between  $\hat{Q}$  and  $Q^*$  *ex post*, entering into a spot transaction for an additional amount  $Q^* - \hat{Q}$ .

Alternatively, buyers' reliance incentives can always be optimized if courts are capable of evaluating directly the efficiency of any reliance expenditure. If courts can evaluate reliance expenditures directly, then it is easy to create optimal incentives simply by rewarding any buyer that relies optimally or by penalizing any buyer who fails to rely optimally. Indeed, there are various legal doctrines that might be interpreted as having this effect by, for example, awarding damages only for "reasonable" reliance expenditures (Goetz and Scott, 1980), or by calculating recovery based on the reliance that would have been "reasonably foreseeable" (Cooter, 1985). Obviously, the effect of any such mechanisms depends on how courts define a "reasonable" (or "reasonably foreseeable") level of reliance. This, in turn, depends heavily on the verifiability of the factors that define the optimal level of reliance (the  $v(\cdot)$  and  $w(\cdot)$  functions, in the model sketched here).

Of course, if the parties themselves can write a complete contract, reasonable expenditures on reliance could be spelled out in the contract itself. As noted earlier, though, this much contractual completeness is sometimes difficult or impossible. Alternatively, if parties can acquire reputations for relying optimally (or for relying excessively), then the competitive advantages of such a reputation could also induce the parties to make optimal reliance investments. We defer discussion of these possibilities to the section on non-legal enforcement (§5.4).

---

<sup>83</sup>See Cooter (1985) and Cooter and Eisenberg (1985).

### 5.2.4 The non-breacher's precautions

A closely-related issue concerns precautions that the buyer may be able to take to reduce the loss he will suffer if the seller breaches. For example, if there is a significant chance that the seller may not deliver the machines she promised, it might be efficient for the buyer to keep his old machines as back-ups rather than getting rid of them as soon as the contract is signed, even if keeping the old machines around is costly. If the seller does perform the contract, the expenses involved in keeping the old machines will have been unnecessary—but if the seller fails to perform, those expenses will have been well spent. Conceptually, then, *failing* to take such a precaution can be thought of as another form of reliance on the promise of performance (Cooter, 1985).

As a consequence, the legal remedy for breach can also affect buyer's incentives to take these sort of precautions. As discussed above, simple expectation or reliance remedies create a moral hazard problem that can leave buyers with incentives to rely too heavily—or, in this context, to take too few precautions. There are, however, legal doctrines that might correct that problem by limiting buyers' ability to recover for losses that could or should have been avoided by appropriate pre-breach precautions. The doctrine of *Hadley v. Baxendale*, for example, may sometimes be used to deny recovery of losses that the buyer could have prevented, by ruling that such losses were not “reasonably foreseeable” to the seller (*e.g.*, R. Posner, 2003b, p. 127). Similarly, the implied excuse doctrines—impracticability, frustration, and mistake—could in principle be used to optimize buyer's incentives by releasing the seller from liability entirely, thereby making the buyer bear all losses suffered from the seller's nonperformance in just those cases where the buyer should have taken more precautions than he did (Posner and Rosenfield, 1977). If courts release sellers from liability only when buyers have behaved suboptimally, in a manner similar to a contributory negligence rule in tort law, that could give buyers an incentive to optimize their own behavior. And as long as buyers optimize their own behavior, courts could then return to holding sellers fully liable, thereby (in principle) optimizing both parties' incentives.

However, these doctrines can have this effect only if courts are able to evaluate fully the precautions available to buyers *ex ante*, in order to recognize cases when the buyer should indeed have taken additional precautions. The extent of courts' ability to do this—and to do it reliably enough so that buyers will know when they will be held liable if they do not take precautions—is a matter of some dispute.<sup>84</sup>

### 5.2.5 The non-breacher's mitigation of losses

Another common form of precaution involves steps the buyer takes after the seller's breach becomes final. For example, if the buyer has bought special equipment to use with a machine that is never delivered, that equipment can

---

<sup>84</sup>For a skeptical view, see Kull (1991).

sometimes be salvaged or re-used, though perhaps not for its full original value. In the notation used in the preceding section, this salvage value was implicit in the  $w(r)$  function, but that notation suppresses the fact that the buyer (the non-breacher) may have to take various steps in order to realize that salvage value. Moreover, some such steps are likely to be efficient while others are not—for example, in some cases the cost that must be incurred to salvage unused equipment may exceed the equipment's salvage value.

Courts are rather more willing to consider such *ex post* opportunities for mitigation of losses and to require buyers to take advantage of them than they are to consider the precautions that buyers might have taken before breach. In part this is because the value of mitigation is especially salient after the fact of breach (or from an incomplete contracts perspective, more easily verifiable). As a matter of law, this issue is most explicitly addressed by the mitigation doctrine, which limits buyers' remedies by denying compensation for any losses that could have been avoided by reasonable mitigation.

As with many legal concepts, what counts as "reasonable" mitigation is a matter of some dispute. To the extent that courts' definition of reasonable mitigation corresponds with efficient mitigation, this limit on remedies can give buyers efficient incentives on this dimension.<sup>85</sup> Obviously, though, the efficacy of this mechanism depends on courts' ability to verify the costs and benefits of various mitigation activities.

### 5.2.6 The decision to terminate a project

In many cases, performance of a contract requires a sequence of many choices or events. For example, the construction of a building involves hundreds of separate steps whose performance can extend over weeks or months. Moreover, sometimes the earliest steps in the sequence must be taken at a time when it is still uncertain whether it will be worthwhile to perform all of the later steps. For example, if future construction costs are uncertain, it might (or might not) prove uneconomic to finish the building, but the early stages of construction may have to start before anyone can be certain what the eventual costs will be.

In such cases, each step in the sequence can be thought of as involving a choice between (a) continuing to perform, or (b) giving up the attempt to perform by terminating the project. Terminating the project early can allow the parties to take steps to reduce their losses, as discussed in the previous section on mitigation. On the other hand, continuing to perform preserves the option value of the project, by deferring to each later stage the decision to continue or terminate. The optimal decision, therefore, is to terminate early only when the savings from doing so exceed the project's option value (Triantis and Triantis, 1998).

The doctrine of anticipatory repudiation allows the potential breacher—the seller, in the examples above—to terminate a contract early simply by repudi-

---

<sup>85</sup>For discussions of this possibility, see Wittman (1981) and Goetz and Scott (1983).

ating it.<sup>86</sup> While this renders the seller liable for damages for breach, it also triggers the mitigation doctrine discussed above, thus requiring the buyer to begin taking any available steps to reduce those damages. As a result, if the seller faces a high probability of being unable to perform, she may be able to reduce her eventual liability by terminating the contract early. Viewed from this perspective, a termination decision by the seller can be thought of as just another form of breach, which will be efficient (or not) depending on whether early termination is optimal.

Indeed, if the seller can be held fully liable for all the losses from breach—either all the current losses, if the contract is terminated early, or all the subsequent losses if the contract is not terminated early—then she will have an incentive to make the optimal choice between continuing or terminating early, for much the same reasons discussed in the section on efficient breach (§5.2.1). However, holding the repudiating party liable for all losses may be difficult. The losses from an early termination include the loss of the option value of the contract, which may be difficult for courts to measure; omitting this loss from the measure of damages will bias the seller's incentives toward terminating too early (Triantis and Triantis, 1998). At the same time, the losses from any eventual non-performance may also be difficult to measure; or they may be difficult to collect, if the seller is by then insolvent. Omitting these losses from the damage measure will bias the seller toward terminating too late (Craswell, 1990).

Perhaps because of these difficulties, various legal doctrines also allow the non-breaching party—the buyer, in the example used here—to force an early termination (while still holding the other party liable for damages) subject to *ex post* judicial review. From the buyer's standpoint, a decision to terminate early can be thought of as another form of precaution (albeit an extreme one), which reduces the losses that will be suffered in the event of a breach. Equivalently, a decision not to terminate early can be thought of as another form of reliance by the buyer, since such a decision increases the gains if the project is eventually successful, but also increases the losses if the project eventually fails. In keeping with the earlier analysis of reliance and precautions (§5.2.3 and §5.2.4), the buyer will not have an incentive to make this decision optimally if he is fully insured by being guaranteed full compensation for his losses.

Perhaps for this reason, the law does not give the buyer the unrestrained power to elect an early termination of the contract, but (using various legal doctrines) subjects that decision to judicial review. For example, if the seller has already breached the contract during early stages of performance, the buyer can elect to force an early termination as long as those breaches are judged by a court to be “material” or “total.” Alternatively, even if the seller has not yet violated any contractual requirement, the buyer may still elect to force an early termination if he has “reasonable grounds for insecurity” about the seller's performance, and if the seller is unable to provide “adequate assurances” that she will be able to perform.

---

<sup>86</sup>for discussions of the measurement of damages in cases involving anticipatory repudiation, see Jackson (1978) and Goetz and Scott (1983).

Of course, the exact effects of these doctrines depend on how courts interpret such vague standards as “material,” “reasonable,” or “adequate.” (If courts could observe all the factors necessary to determine whether early termination was optimal, they could allow termination in exactly those cases; but their ability to do this is doubtful.) The issue is further complicated by the fact that courts do not usually issue advance opinions on pending disputes, but instead render their judgments in hindsight. An insecure buyer must thus decide whether to force an early termination before he is certain whether a court will agree that the other party’s earlier breach was “material,” or that the other party’s assurances were not “adequate.” If the buyer tries to terminate early, and a court later rules that grounds for early termination were lacking, the buyer may himself be held liable for breaching the contract prematurely. Given the substantial uncertainty associated with such vague legal standards, a risk-averse buyer may prefer not to force the issue (and a less risk-averse seller may take advantage of the situation to force a modification or a release of its obligation to perform).

Finally, as with other perform-or-breach decisions, the decision to terminate early should be made optimally whenever *ex post* renegotiation is possible, as there should always be some combination of side-payments that makes it in both parties’ interests to terminate early whenever early termination is optimal. However, the legal doctrines discussed above will affect each party’s share of any surplus, depending (for instance) on whether the buyer has a unilateral right to force an early termination without the seller’s consent. As usual, though, such renegotiation may itself be costly; and the prospect of higher or lower side-payments (in the event of renegotiation) can also alter many of the other incentives discussed above, such as the incentive to rely or the incentive to take precautions.<sup>87</sup>

### 5.2.7 Decisions to gather and disclose information

Much of the literature described in the preceding subsections presupposes that both parties already know all of the relevant parameters, such as the distribution from which the seller’s cost,  $c$ , will be drawn, or the extent of a buyer’s reliance investments,  $r$ . In many cases, though, this (and other) information cannot be discovered without costly investments in information-gathering. To be sure, a number of legal doctrines affect the incentives to gather or disclose information, including some that are not normally classified as “remedies” doctrines (for example, see the discussions of fraud and mistake earlier in §2.5.2 and duty to disclose in §3.3.3). However, the remedies for breach can also affect various information-related incentives. While there are many different kinds of information that might be gathered or disclosed, and the effects of remedies on these incentives have not been studied as extensively as have the other incentives discussed earlier, we discuss some representative examples.

---

<sup>87</sup>For further discussion of all of these issues, see Goetz and Scott (1983), Craswell (1990), and Triantis and Triantis (1998).

DECIDING WHETHER TO ENTER INTO CONTRACTS. The greater the penalties for breach, the greater will be sellers' incentive to gather information about potential risks that might leave them unable to perform. If a seller who fails to perform is held fully liable for all of a buyer's losses, that will make her internalize the full costs of any failure on her part to gather sufficient information. However, this will not be sufficient to make sellers' incentives optimal unless sellers can also internalize the full benefits of any additional investments in information-gathering. A perfect price-discriminating monopolist may be able to internalize all of those benefits, by charging higher prices; and in perfectly competitive markets, competitive pressures may force sellers to account for those benefits as well. In markets that fall between these extremes, however, it is more difficult to design a remedy that gives sellers the optimal incentive to gather this information (Craswell, 1988).

SEARCHING FOR CONTRACTUAL PARTNERS. In many markets sellers (and buyers) may have a choice as to how much effort they devote to finding potential trading partners. Here, too, the parties' incentives depend partly on other legal doctrines—in this case, the rules governing initial contract formation (see §3 above). However, the parties' incentives on this point, too, will also depend on the general remedies for breach. For example, the steeper the penalties for breach, the harder it will be to get out of any contract once a contract has been formed, and so the more it will pay parties to invest in searching longer and harder to make sure they have found the best deal. Here, too, some of the benefits of increased search may be felt by the party who is found, rather than by the party who is doing the searching, so efficient search incentives can be created only if the searching party is able to internalize these benefits. For formal models incorporating these effects, see Diamond and Maskin (1979, 1981).

OPTIMAL PRECAUTION AND RELIANCE. Even after one party has gathered information, in some cases the other party is the one who needs to be given that information in order to make an optimal decision. For example, if the efficient level of seller precautions depends on how much the buyer would gain from performance, and if the buyer already knows that information, the seller's precaution incentives can more easily be optimized if the buyer can communicate his information to the seller (Bebchuk and Shavell, 1991). Similarly, if the optimal level of reliance by the buyer depends in part on the probability that the seller will fail to perform, and if the seller already knows that probability, the buyer may be better able to choose an optimal level of reliance if the seller tells him what that probability is (Craswell, 1989b). While most of the conventional remedies for breach do not create any incentives for a buyer to convey this information to the seller, some remedies affect this incentive by conditioning the measure of damages on the information that has been disclosed. Under the rule of *Hadley v. Baxendale*, for example, a buyer facing large losses from breach is more likely to be allowed to recover those losses if she has told the seller about

them in advance.<sup>88</sup>

### 5.2.8 The effects of party heterogeneity

Information also matters if sellers (or buyers) must deal with a heterogeneous population of trading partners. In many settings, for example, sellers may differ in the probability that they will be able to perform, or buyers may differ in the amount they would lose if the contract were breached. If each party's type is fully known to the other party, those differences can be fully reflected in each side's prices and other behavior, allowing each interaction to be analyzed as if it involved entirely homogeneous parties. When differences in characteristics cannot be fully observed, however, much of the analysis described above must be altered in some way.

For example, if buyers differ in the amount they will lose from a breach—and if the law's damage measure reflects these differences, by holding sellers liable for different damages depending on the buyer's type—sellers' expected liability costs will be lower when dealing with low-damage buyers than when dealing with high-damage buyers. If sellers can recognize the high-risk buyers, they can adjust their prices or their level of precautions or both to reflect those greater risks. However, this gives high-damage buyers an incentive to conceal their greater riskiness, if they can. In that case, the equilibrium will depend on the ability of the low-damage buyer to signal his type or the seller's ability to screen buyer types. As discussed in §2.3.2, signaling and screening can result in a loss of welfare. If screening or signaling are ineffective at separating the types, then a pooling equilibrium could result in which low-damage buyers effectively subsidize high-damage buyers. Alternatively, to avoid this subsidy, low-damage buyers could exit; that is, an adverse selection ("lemons") problem ensues. Limits on the recovery of unusually high damages—including the doctrine of *Hadley v. Baxendale*, discussed in the preceding subsection—could conceivably be used as a screen to prevent subsidization or drop out (see Quillen, 1988).

Buyer heterogeneity is also important if a seller has market power, insofar as it introduces the possibility of price discrimination. To be sure, if the seller can fully observe each buyer's type, she can engage in first-degree price discrimination, a practice that general contract law does not restrict. But if the seller cannot observe each buyer's type directly, she may be able to use some of the terms of her contract (including the remedies for breach) to separate different classes of buyers—screen—thereby increasing her profits through second-degree price discrimination (see, *e.g.*, Matthews and Moore, 1987). This possibility could also be relevant to the analysis of liquidated damage clauses, discussed later in §5.3.4. As discussed in §2.3.3, especially footnote 29, the welfare consequences of such price discrimination are *a priori* ambiguous.

---

<sup>88</sup>For economic analyses of this aspect of the Hadley doctrine see, *e.g.*, Ayres and Gertner (1989), Wolcher (1989), Johnston (1990), Bebchuk and Shavell (1991), Adler (1999), and Ben-Shahar and Bernstein (2000). Because of possible adverse price discrimination *vis-à-vis* the shipping price, however, a buyer may be reluctant to reveal that she faces large losses (see §4.2.4 and §5.2.8).

### 5.2.9 Risk allocation and insurance

Finally, in addition to all of the incentive effects discussed above, remedies for breach also have the effect of allocating various risks between the two parties. For example, the expectation remedy defined in §5.1.2 will, if it is measured perfectly, leave the buyer (the non-breaching party) fully insured. Thus, if buyers are risk averse, while sellers are risk neutral, this remedy will be efficient in terms of its effects on the parties' attitudes toward risk. For other combinations of attitudes toward risk, remedies other than expectation damages will be superior (see generally Polinsky, 1983). Unless one of the parties is actually risk loving, however, remedies that exceed the non-breacher's actual losses (*e.g.*, punitive damages) will almost always be undesirable as far as risk sharing is concerned (Craswell, 1996a).<sup>89</sup>

One case of interest concerns breaches that inflict non-pecuniary losses that do not increase the buyer's marginal utility of money (see Cook and Graham, 1977). For example, a photographer's failure to take wedding pictures might reduce the welfare of the marrying couple, but that does not mean they would prefer to buy an insurance policy that would pay them additional money if their wedding pictures were lost.<sup>90</sup> Viewed purely from the standpoint of the parties' taste for insurance, then, it might be better if contract remedies did not compensate for this sort of loss (as, indeed, they generally do not). At the same time, though, excluding these losses from contract remedies could also distort some of the other incentives discussed above, such as the photographer's incentive to take adequate precautions. For an analysis incorporating both of these effects, see Rea (1982).

## 5.3 Complications in determining monetary damages

In the discussion so far, we have dealt with the incentive and insurance effects induced by monetary remedies for breach. To be sure, much of the literature on this topic developed by analyzing particular legal remedies: either particular measures of monetary recovery, or non-monetary remedies such as specific performance. Reflecting that literature, §5.3.1 considers various legal and practical limits on contract damages; and section §5.3.2 considers some alternative damage measures that are imposed in certain cases. Then, §5.3.3 and §5.3.4 discuss specific performance (injunctive relief) and liquidated damage clauses (remedies stipulated in the contract itself).

---

<sup>89</sup>Risk aversion also plays a role in why limits on damages could be Pareto improving when the damage terms signal information about the potential breacher. See §2.3.2 and the discussion of Aghion and Hermalin (1990).

<sup>90</sup>Such preferences are, admittedly, inconsistent with neoclassical economic theory unless the couple exhibits lexicographic preferences in photos-money (other goods) space.

### 5.3.1 Limits on the measure of damages

In theory, expectation damages leaves the non-breacher just as well off in every state of the world (see §5.1.2); but in practice, a number of legal doctrines limit the losses that expectation damages will compensate. For instance, the non-breacher must prove the amount of his loss with “reasonable certainty”; often this will exclude the recovery of “speculative” losses whose amount was uncertain. Also, as noted earlier, contract law only rarely allows compensation for emotional losses (see §5.2.9). Perhaps more significant in this regard than any rule of contract law, however, is the general default rule of US legal procedure that requires each party to bear his own costs in litigation. In most other Western legal systems, in contrast, prevailing litigants are entitled to recover attorneys’ fees as well as other out-of-pocket litigation costs. As a result, they come closer to being made whole than do winning litigants in the US.

By reducing the effective amount of the remedy, doctrines such as these weaken many of the seller’s incentives discussed earlier (for example, the seller’s incentive to take precautions, see §5.2.2). Of course, by shifting more of the loss to the buyer, the same doctrines may also strengthen the buyer’s incentive to take various precautions (see, *e.g.*, §5.2.4). Also, when these exclusions are conditioned on particular behavior by the buyer—*e.g.*, if recovery for certain losses is excluded unless the possibility of such losses is disclosed in advance, under the doctrine of *Hadley v. Baxendale*—that could strengthen the buyer’s incentives in more focused ways (see §5.2.7). Finally, if buyers differ in the extent to which they suffer non-recoverable losses, excluding those losses from the damage measure may reduce the cross-subsidization that could otherwise result (see §5.2.8).

### 5.3.2 Other measures of monetary damages

The damage measures identified above in section §5.1.2 above have received the most attention in the law and economics literature, but other measures are also sometimes used. As we show, though, most of the economic effects of these alternate measures can be decomposed into one or more of the effects already considered above.

**COST OF COMPLETION.** One recurring issue involves sellers who breach by leaving work unfinished, or by performing work incorrectly, when it would be extremely costly to finish or correct the work. For example, suppose that the buyer would have realized net gains of  $v$  from complete performance, while leaving the work in its current state would leave the buyer with net gains of  $u$ , and it would now cost  $f > v - u$  to finish the work as it should have been performed. If  $v$  and  $u$  have been measured correctly, this implies that it would be inefficient to finish the work. However, courts vary in their treatment of this case, sometimes allowing the buyer to recover the completion cost  $f$  in damages, while sometimes limiting the buyer to recovery of  $v - u$ .

Awarding  $f$  in damages would be inefficient if it led to the work actually

being completed. If, however, completion would in fact be inefficient, then the buyer will pocket the damage payment rather than use it to finish the work. The principal effects of awarding  $f$ , rather than awarding  $v - u$ , will, thus, be those discussed previously. The larger remedy will (in general) strengthen sellers' incentives to avoid committing this sort of breach; perhaps strengthening them excessively, if the actual losses caused by the breach are only  $v - u$ . On the other hand, if  $v - u$  actually understates the buyer's losses—say, because some of the benefits from performance are hard to measure, and have therefore been omitted from the court's measure of  $v$ —then increasing the remedy to  $f$  could improve the seller's incentives in some respects (see Muris, 1983, for a discussion of both effects). Even then, much would depend on the exact nature of the benefits that were excluded, as was also discussed in preceding subsections. For example, if buyers differ in the extent to which they would realize certain benefits, excluding those benefits from the damage measure could reduce any cross-subsidization that might otherwise result (see §5.2.8).

**DISGORGEMENT.** Similarly, courts occasionally require a breaching party to disgorge any profits he may have earned as a result of the breach, even if those profits exceed the non-breacher's expectation loss, often citing a general principle that no one should profit from his own wrong. This result is not the norm in contract cases, but is reserved for situations in which the breacher is thought to have engaged in bad faith or "willful" breach (an ill-defined notion in the case law) or when the non-breacher is considered to have a property or quasi-property interest in the subject of the contract and is thus entitled to the proceeds of resale, even if he could not have earned such proceeds on his own (see Farnsworth, 1985, for a general discussion).

Disgorgement damages, if assessed with certainty, leave the breaching party indifferent between performance and breach. As such, they entirely eliminate any incentive to breach, which from the viewpoint of efficient breach, is an excessive deterrent. In turn, the absence of efficient breach could also generate excessive reliance by the promisee.

A rationale sometimes articulated in favor of disgorgement damages is that, by removing any incentive for unilateral breach, they encourage a party who would like to escape performance to approach the counterparty and negotiate a modification or release (see, *e.g.*, Friedmann, 1989). Such negotiation may be desirable if the potential breacher would otherwise be uninformed about the size of the counterparty's expectation loss (thus leading him mistakenly to breach), or if the transaction costs of renegotiation are less than the costs that would be occasioned by a lawsuit. The frequent association of disgorgement with the elements of bad faith or willfulness, however, suggests an underlying punitive component to this remedy, which raises the topic of punitive damages generally.

**PUNITIVE DAMAGES.** In general, explicitly punitive damages are rarely imposed in contracts cases. This is in part because breach of contract, though a legal wrong, is typically judged more leniently than would breaches of duty in a tort

or property setting, for a number of related reasons. First, because liability is based on a voluntary exchange relationship, the expected costs of breach and of paying damages are likely to be reflected in the contract price, so that efficient breaches work to the *ex ante* advantage of both parties. Second, to the extent that punitive damages are justified by negative externalities imposed on the general community, the concern is less pressing in the ordinary contract. Third, to the extent that punitive damages are justified by a high likelihood of non-detection, this concern is lessened when the parties know each other and are likely to be watchful of proper performance.

Punitive damages are very occasionally imposed, however, in response to breaches of contract that also involve a tort, gross unfairness, or a violation of some public policy. For example, punitive damages have regularly been imposed on insurance companies that refuse without justification to pay valid claims. This result is often defended in terms of the imbalance in the parties' economic power, the particularly difficult circumstances in which such refusal places the insured, and the likelihood that most victims of such opportunism are likely never to seek redress in court—all factors that go to the general case for punitive damages, as laid out in succeeding chapters of this handbook.

### 5.3.3 Specific performance

In some cases courts order specific performance rather than awarding monetary damages for breach. In effect, this remedy requires the seller (or other breaching party) to perform the contract in full, backed by the threat of fines or even imprisonment if she fails to do so.

If *ex post* renegotiation is impossible, such an order would lead to inefficient performance of the contract in any case where breach was more efficient. As long as the parties can renegotiate, however, they should always be able to avoid this loss by agreeing not to perform, with the gains from non-performance being shared between the parties in accordance with their bargaining strength. But if the buyer has the threat of a remedy of specific performance, thereby requiring the seller to incur the costs of performance, that should allow the buyer to capture more of the gains than he could if his only legal threat were to hold the seller responsible for some smaller monetary remedy.

As a result, when *ex post* renegotiation is possible, the effects of specific performance will be felt in all of the ways discussed above. When renegotiation is costly, specific performance could, in principle, add to those negotiation costs, though it is unclear whether the negotiations required under specific performance will be any more or less costly than those required under any alternative remedy.<sup>91</sup>

---

<sup>91</sup>For analyses of this issue see, *e.g.*, Kronman (1978b), Schwartz (1979), Ulen (1984), and Bishop (1985). Most recently, Shavell (2006) has argued that negotiation costs will be relatively large where the reason for breach is a production cost increase (because in order to negotiate a release, the parties must agree on how to distribute the quasi-rents arising from the cost increase), and relatively small where the reason for breach is to sell to an third party (because the buyer can also resell to the third party, and so the quasi-rents

In any event, if specific performance results in the buyer being able to negotiate for a larger payment, this will have the same effect as an increase in the expected size of any monetary remedy. For example, the threat of having to make such a payment will strengthen the seller's incentives to take precautions against events that might expose it to such a remedy (Muris, 1982). Moreover, this threat will also increase the buyer's incentives to make reliance investments (Edlin and Reichelstein, 1996). An increase in the expected payment will also alter the risks born by each party, as discussed earlier in §5.2.9. In general, specific performance makes it more likely that the buyer will end up at least as well off as if the contract had been performed (otherwise, he would not consent to any renegotiation), whereas the award of smaller monetary remedies might not leave the buyer this well off. But whether this increase in compensation is desirable, all things considered, depends on all of the effects discussed earlier, as compared to the effects produced by whatever monetary measure of damages the court would award if it did not require specific performance. In this regard, the choice between specific performance and monetary damages has much in common with the choice between injunctive relief and monetary remedies in many other areas of law.<sup>92</sup>

Under common law, specific performance has traditionally been more difficult to obtain than monetary damages, with injunctive relief treated as matter of the court's discretion, and usually being reserved for cases in which damages are deemed insufficient to protect the non-breacher's expectation interest or in other special circumstances (for example, when the goods being traded are unique, when the breacher is insolvent, or when the non-breacher has made relational investments that would be difficult to replace). The most widely cited policy reason for these restrictions is that specific performance is thought to impose greater administrative costs on the legal system, especially in situations where the quality of a coerced performance is costly to verify. But this concern has not prevented civil law systems from making specific performance their remedy of default, even though in many circumstances their courts will still award money damages in substitution for performance (Lando and Rose, 2004).

### 5.3.4 Remedies expressly stipulated in the contract

As noted earlier (see §5.1.2), the remedies discussed above are usually supplied by the legal system as default remedies to be applied in cases where the contract is silent as to the consequences of breach. In some cases, though, the parties' contract may itself stipulate the remedy that will be required in certain states of the world. Usually this remedy will consist of a monetary payment, but in some cases, the parties may provide for particular actions to be taken, as when a sales contract provides that in the event of a defect in the goods, the seller

---

are limited to the difference between the buyer's and seller's cost of resale).

<sup>92</sup>There is a much more extensive literature analyzing the analogous issue in property and tort law. See, *e.g.*, Calabresi and Melamed (1972), Kaplow and Shavell (1996), Bebchuk (2001), and Ayres and Goldbart (2003).

will be obliged to provide repair or replacement. When such stipulations take a monetary form, they are usually referred to as “stipulated damages” or, at least when they are enforceable, “liquidated damage clauses.”<sup>93</sup>

Depending on the amount chosen by the parties to serve as the measure of damages, liquidated damage clauses can produce any and all of the effects described above.<sup>94</sup> For example, clauses that specify a large payment can give sellers a strong incentive to avoid committing a breach, while clauses that specify smaller payments will give buyers less insurance against breaches. The amount of the clause can also affect each side’s incentives to rely on the contract, or to take precautions against various risks, as the preceding sections also discussed. Indeed, while the literature summarized in the preceding sections was mostly written to provide guidance to courts or other lawmakers, that literature can just as easily be read as providing guidance to private parties who wish to select an efficient liquidated damage remedy (see Katz, 1996b).

Moreover, in some respects liquidated damage clauses (drafted by the parties) are likely to be superior to general default rules (selected by courts or legislatures). As the preceding sections make clear, most remedies involve trade-offs among various important incentives, and in many cases the contracting parties are better suited than courts to choose the particular trade-off that is best for their own transaction. In markets where parties are heterogeneous (see §5.2.8), liquidated damage clauses can be tailored to particular contracting pairs. More broadly, all of the reasons that support the enforcement of contracts generally (see §2.2) will usually argue for the enforcement of liquidated damage clauses in particular.

It is therefore striking that common-law courts refuse to enforce clauses that set damage amounts that the courts consider excessive (these are typically referred to as “penalty clauses”). This reluctance may be due partly to historical factors, and in particular to the belief (on non-economic grounds) that “penalties” or “enforcement” should be the exclusive province of the legal system, rather than being subject to private control. In addition, though, there may also be economic reasons that—in particular situations—might counsel against enforcement.

For example, in cases where the actual damages from breach turn out to differ significantly from what the parties expected, it is possible—though far from automatic—that a remedy specified by the liquidated damage clause might no longer be welfare maximizing.<sup>95</sup> In effect, denying enforcement in these cases would be similar to denying enforcement of the contract itself, under the “im-

---

<sup>93</sup>Very occasionally, parties will provide for specific performance in their contracts, but courts do not regard themselves as bound by such provisions (although they will take them into account as a factor bearing on their exercise of discretion).

<sup>94</sup>For discussions of these effects in connection with liquidated damage clauses, see Goetz and Scott (1977), Clarkson et al. (1978), Rea (1984), Schwartz (1990), and Edlin and Schwartz (2003).

<sup>95</sup>Observe, however, that if the remedy is a monetary transfer, then welfare is typically assumed to be unaffected by transfers.

plied excuse” doctrines of mistake or impracticability, in cases where unforeseen events have significantly altered the contracting parties’ situation. It is, however, a matter of debate whether courts have the ability to make such *ex post* adjustments in ways that will in fact improve the parties’ *ex ante* welfare. These doctrines are discussed above in §4.4; for discussions focusing specifically on liquidated damage clauses, see Rea (1984) and Schwartz (1990).

In addition, some liquidated damage clauses can affect the welfare of others who are not parties to the contract. In particular, if a seller with market power is concerned with defending its market against entrants, liquidated damage clauses can serve as a commitment to deter competitors from entering (Aghion and Bolton, 1987; Chung, 1991; Spier and Whinston, 1995). Even in competitive markets, if some or all buyers are unaware of seller’s liquidated damage clauses, this imperfect information could create incentives for socially undesirable clauses (just as they could in the case of any contractual clause, see generally §2.3.2).

It is worth noting that there are some situations in which common-law courts are more likely to enforce privately stipulated remedies. In particular, courts are more deferential to liquidated damage clauses that turn out to be undercompensatory *ex post*, as compared to those that turn out to be over-compensatory (perhaps because the latter raise third-party effects where the former do not). They are quite deferential to clauses that disclaim liability for consequential damages and that limit the remedy for breach of warranty to repair or replacement (because such clauses supplement the doctrine of *Hadley v. Baxendale*, of which the courts approve). They are also likelier to enforce liquidated damages that have actually been paid over as an advance deposit, although even in those cases, the breacher may be entitled to restitution of part or all of the deposit to the extent it plainly exceeds the non-breacher’s expectation loss.

## 5.4 Private enforcement of contracts

As noted, legal enforcement or its threat are not the only means by which parties are induced to honor their commitments. In this subsection, we briefly consider some of these other means. In §5.4.1, we briefly consider how the costs of using legal enforcement can either distort contracts or cause the parties to dispose of them altogether. In §5.4.2, we take up how repeated interactions or reputational concerns can deter breach. Finally, §5.4.3 discusses various legal doctrines that bear on, and in some cases support, these private alternatives to traditional legal enforcement.

### 5.4.1 Enforcement costs

One reason contracts could fail to be enforced as written is that enforcement requires expenditures by the parties that are either *ex post* incredible or can be anticipated to be so large *ex ante* that no contract is written. A partial list of examples is: (1) the agreement is illegal or exceeds the parties’ power to contract under the applicable legal system (see §2.5 above); (2) courts cannot verify critical aspects of contractual performance or whether relevant contingen-

cies have arisen (see §4.2 above); (3) litigation is costly in terms of time, risk, or material resources; (4) the defendant may be insolvent or otherwise lack the ability to comply with a judgment; and (5) the dispute arises in the course of an otherwise successful relationship that the parties do not wish to jeopardize. In addition, as a large literature makes clear, two additional reasons are also important in developing countries:<sup>96</sup> (6) the court system operates corruptly; and (7) courts are incapable of enforcing their verdicts, because police are corrupt or unavailable.

A simple model illustrates some of these issues. Suppose that party *B* employs party *A* and promises to pay *A*  $w$  upon completion of *A*'s task. Suppose it costs *A*  $k_A > 0$  to have the contract enforced or fight litigation and it costs *B*  $k_B > 0$  for the same. If *A* cannot recover its enforcement costs from *B* should she prevail at trial, then *B* knows he can underpay *A* by up to  $k_A$  without *A* seeking to enforce the contract (assuming *A* acts in a coldly rational way).<sup>97</sup> If, however, the parties anticipate this, then they could agree to a nominal wage of  $w + k_A$ , recognizing that *B* will underpay by  $k_A$ . Somewhat more problematic is non-performance by *A*. If it is feasible for *A* to overperform by an amount  $k_B$ , then the parties can simply set the performance standard in the contract  $k_B$  above what they truly intend. But such overperformance could be infeasible, in which case this trick won't work. Now it could be necessary to add a clause to the contract that *A* pay *B*  $k_B$  should *A* be found to have underperformed. To the extent the court refuses to honor that clause, citing the unenforceability of penalty clauses (see §5.3.4), or its inability to verify *A*'s performance adequately, this solution could also fail. If it is impossible to enforce a contract contingent on *A*'s performance, then the parties will either have to forgo contracting or they will have to contract around the problem (*e.g.*, use a revenue-sharing contract to give *A* better incentives).

More generally, whenever judicial enforcement is likely to occur in less than 100% of the cases, it might in principle be possible to make up for that deficiency by increasing the size of the damage award in those cases that do reach the courts. To take a simple example, if only one out of ten breaches is ever sanctioned by courts, many of the incentive effects could be restored if the damage award that would otherwise be optimal were multiplied by ten in every case.<sup>98</sup> However, courts are usually reluctant to make this adjustment in contract cases, where punitive damages are only rarely awarded. This solution will also be unattractive if either party is risk averse, as it increases the variance of both parties' returns (see §5.2.9). And in more extreme settings, where litigation is infinitely costly (*e.g.*, performance is completely unverifiable or in countries

---

<sup>96</sup>See, for instance, Anderson and Young (2002), Cungu and Swinnen (2003), or, for a more historical perspective, Greif and Kandel (1995).

<sup>97</sup>It is known from research on ultimatum games (*e.g.*, Güth et al., 1982) that players do not always act in a coldly rational way, preferring sometimes to punish others even if the act of punishing is costly to them. See Rabin (1993) for a discussion and analysis.

<sup>98</sup>This effect is often suggested as an economic rationale for punitive damages. See, *e.g.*, Polinsky and Shavell (1998) and Craswell (1999).

where a reliable court system is unavailable), reliance on court-ordered remedies is bound to fail.

Unfortunately, the issue of credible enforcement is typically ignored in most of the contract design literature.<sup>99</sup> Accordingly, we now turn briefly to enforcement methods that do not require the participation of courts.

#### 5.4.2 Self-enforcing contracts

It has long been understood from the repeated games literature that some agreements are self-enforcing in the context of an ongoing relationship.<sup>100</sup> The most prominent example of such “agreements” is tacit collusion among competing firms. That is, recognizing their repeated interaction, firms avoid undercutting each other on price. This “agreement” to keep prices high is enforced by the threat of a price war should any firm undercut.<sup>101</sup>

Within the realm of contracts, there is wide scope for such self-enforcing agreements. Moreover, self-enforcement can substitute for legal enforcement. For instance, in a one-shot game, an employer might choose to renege on a promised wage payment to an employee (recall the discussion in the previous subsection). But in a repeated context, the employee could retaliate by quitting (or possibly engaging in sabotage), which could be sufficiently costly to the employer that he chooses to pay the employee.<sup>102</sup> Of course, the employee could also take the employer to court for nonpayment, but, as we saw above, legal enforcement might not always be credible.

Self-enforcement can also complement legal enforcement. For instance, while it might not be credible for a party to enforce a contract legally in a one-shot game, the party might wish to develop a reputation for enforcement in a repeated context—just as in some models of entry deterrence whereby an incumbent firm punishes entrants to establish a reputation for toughness (see, *e.g.*, Tirole, 1988, Chapter 8), a party may want to develop a reputation as someone who can’t be cheated (*e.g.*, become known as litigious). Of course, if repetition is what is making legal enforcement of contracts credible, then one should model repetition explicitly in the analysis of the underlying contracting problem if that problem is itself repeated (*i.e.*, the game between an employer and a long-term employee). In other words, an appeal to reputation for enforcement in a static analysis of a contracting problem is most acceptable when the problem is short term (*i.e.*, a given pair meet only once to contract), while the players are long lived (*i.e.*, will play again with others).

Within the literature, self-enforcing contracts are often known as *relational contracts*. Applications have included quality assurance for experience goods

---

<sup>99</sup>Some notable exceptions are Spier (1994) and Krasa and Villamil (2000).

<sup>100</sup>For a review of repeated games see Chapter 2 of Gibbons (1992) or Chapter 5 of Fudenberg and Tirole (1991).

<sup>101</sup>See Chapter 6 of Tirole (1988) for details.

<sup>102</sup>See, also, Thomas and Worrall (1988).

(*i.e.*, goods the quality of which can only be assessed via consumption);<sup>103</sup> incentive schemes;<sup>104</sup> and social contracts within firms.<sup>105</sup> However, explicit models of reputation in the legal literature on contracts are still relatively rare.<sup>106</sup>

In addition, while reputations can, in many markets, provide powerful incentives to perform, in some markets they are less likely to be effective. For example, if performance involves a credence good (the quality of which cannot be observed even after consumption, at least not without expert diagnosis), many breaches may go undetected, with little harm to the breaching party's reputation.<sup>107</sup> In addition, the enforcement of reputations may be privately costly to those who enforce them, thus leading to free-rider problems in enforcement.<sup>108</sup> In other markets, where sellers' histories are not easily discoverable by buyers (or *vice versa*), the incentive effects of reputations may be weakened, though the involvement of kin or ethnic groups or other reputational intermediaries may help in overcoming that difficulty.<sup>109</sup> Finally, for sellers who are on the verge of bankruptcy (or are otherwise reaching their "last period"), the prospect of losing future business may be a very weak constraint at best.

### 5.4.3 Legal doctrines affecting self-enforcement

As the preceding subsection discussed, reputations are most effective in the context of a repeated game, so that a party who cheats suffers the consequence by losing the benefit of future interactions. In some contractual settings, however, the party who cheats can be made to suffer an extra-legal sanction in connection with the very contract that has been breached, as long as the other party has not yet fully performed his own end of the contract. In such a case, the other party may respond not by filing a lawsuit, but by withholding his own performance. For example, if a seller agrees to deliver goods on credit, but if the buyer discovers (before paying) that the goods are defective, the buyer might respond to this breach by refusing to accept or to pay for the goods.

To be sure, the significance of a threat to suspend performance depends partly on the value that performance has to the other party, but it also depends on how the parties have structured their transaction. To take an extreme case,

---

<sup>103</sup>Klein and Leffler (1981) and Shapiro (1983) are two examples.

<sup>104</sup>Bull (1987), Levin (2003) and MacLeod and Malcolmson (1989) are three examples.

<sup>105</sup>See Hermalin (2001) for a survey.

<sup>106</sup>For a qualitative discussion, see Charny (1990). Bernstein (1992, 1996, 2001) has presented several case studies illustrating the operation of reputational enforcement in specialized markets.

<sup>107</sup>See Darby and Karni (1973). Although see Fong (2005) for a more nuanced analysis.

<sup>108</sup>For example, Klein and Leffler (1981) model a market in which buyers, if they even once receive a defective product, follow a flat rule of never purchasing from that seller again. While such a rule does produce desirable incentive effects, it may or may not be rational for individual buyers.

<sup>109</sup>See Landa (1981) and Bernstein (1992) for discussion of kin and ethnic networks. See Mann (1999) on other reputational intermediaries.

if the contract calls for the seller to make all her deliveries before the buyer pays any of the price, that gives the buyer a good deal of leverage by threatening to withhold payment. On the other hand, if the contract instead calls for the buyer to pay the entire price in advance, before any of the goods have been shipped, this gives the buyer very little leverage, while putting the seller in the happy position of being able to threaten to suspend all of her shipments. As a consequence, parties often negotiate extensively over the exact timing of the various payment and delivery requirements.<sup>110</sup>

In addition, a party's right to suspend his or her own performance may also be regulated by various legal doctrines, as we now discuss.

**RESCISSION.** Once one party has committed a breach, the other party may sometimes be able to choose between monetary remedies (typically expectation damages, as discussed earlier in section §5.1.2) and simply walking away from the contract, without collecting any remedy at all. This latter option is usually referred to as termination or rescission. To be sure, if the contract would have been a profitable one for the non-breaching party, that party will usually prefer expectation damages over rescission, for expectation damages should give the non-breacher all of the benefits she would have received from performance (if all of those benefits can be adequately measured, see section §5.2.1 above). But if the contract in question would have been a losing one for the non-breaching party, rescission may be a more valuable remedy, as it allows that party to walk away from what might otherwise be a significant loss.

To complicate matters further, in some cases the non-breaching party may elect to rescind a contract even if she has already performed some part of her own services under the contract. By rescinding the contract, the non-breaching party would give up her right to recover the payment specified in the contract, but she could then sue for restitution to recover a judicially-determined "reasonable value" for her services.<sup>111</sup> As the court's determination of reasonable value need not be limited by the contract price for those services, this remedy could leave the non-breacher with more than she could get under any alternative remedy. Indeed, precisely because rescission is elective for the non-breacher, a rational non-breacher will not choose it unless it is more favorable to her.

However, several legal doctrines limit the use of rescission as a response to the other party's breach. At common law, breach of a service contract allows the non-breaching party to elect rescission only if the breach is "material" or "substantial," a vague test that leaves much to the courts' discretion. By contrast, breach of a contract for the sale of goods is said to allow rescission for any breach whatsoever (the so-called "perfect tender" rule). The Uniform Commercial Code has altered this latter rule, though, by limiting the buyer's

---

<sup>110</sup>See Scott and Triantis (2006). In addition to the sources cited earlier in §5.2.6, brief discussions can also be found in Goetz and Scott (1983), Kull (1991), and Kraus (1994). Note, also, the connection between this and the discussion of option contracts and holdup in §4.3.2.

<sup>111</sup>For other uses of restitution as a remedy, see section §5.1.2 *supra*.

right to rescind in cases involving the sale of goods. Under the UCC, defective goods in a single shipment of a multi-shipment (or installment) contract do not allow the buyer to rescind the entire contract unless the defect “substantially impairs” the value of the entire contract. Even in contracts calling for only a single installment, the buyer may lose the right to rescind if he fails to reject the installment within a “reasonable” time for inspection. The buyer’s right to rescind may also be limited by the seller’s right to take a “reasonable” amount of time to “cure” the defect (see the discussion of “cure” below).

While the remedy of rescission has not been analyzed as extensively as other remedies have, many of the effects are similar to those of any other remedy that is more generous to the non-breacher.<sup>112</sup> That is, as noted, the non-breacher will elect rescission only when it is more favorable to her than the other available remedies. Consequently, the availability of rescission should increase the breacher’s incentive to perform—just as would any other increase in the size of the likely remedy. Of course, the parties may be able to renegotiate *ex post* to avoid inefficient breach or inefficient performance—but, as with other remedies, the payments that parties must make in *ex post* renegotiations will still affect their *ex ante* investment incentives. Also, if either party is risk averse, the availability of rescission will also affect their risk-bearing costs, again in the same manner as any other increase in the expected size of the remedy.

**CURE.** As noted in the preceding paragraphs, the buyer’s right to rescind a contract for the sale of goods may be limited (under the UCC) by the seller’s right to cure any defects in the goods that she delivered. As long as the time specified for delivery of the goods has not yet expired, the seller has complete freedom to try to cure the defects and deliver conforming goods. However, even after the time for delivery has expired, the seller may still have some right to attempt a cure, although this right is subject to various legal limits (many of which are vague). For example, the seller may not take more than a “reasonable” time to effect such a cure; and may only do so when she had “reasonable grounds” to believe that her original, non-conforming delivery would have been acceptable “with or without a monetary allowance” for the defect. In installment contracts, where only the goods in a single shipment were defective, the seller must be able to provide “adequate assurances” that its cure will be successful. And in all cases, it is ultimately up to a court to determine whether the seller’s efforts have in fact cured the defect.

In cases where it is clear that the seller can fix the defect, the right to cure serves to limit the effect of the remedy of rescission. For example, if the market price of goods fell significantly after the contract was signed, the buyer might otherwise use a trivial defect—one that could be cured at a *de minimis* cost—to rescind the contract, thus forcing the seller to bear the loss from the market’s fall. The economic effects of this use of rescission were discussed above. Clearly, giving the seller a right to cure eliminates those effects.

---

<sup>112</sup>The earliest economic discussion is Goetz and Scott (1983). Other discussions include Kull (1991) and Kraus (1994).

In other cases, though, it may not be clear (at least initially) whether the seller will ever be successful in curing the defect. For example, many litigated cases involve the sale of cars or houses that seem to be “lemons,” whose seller makes repeated but unsuccessful attempts to find and fix the problem.<sup>113</sup> In these cases, a court’s interpretation of the right to cure has the effect of determining the point at which the contractual endeavor should be terminated in order to cut the parties’ losses. The economic implications of this decision were discussed earlier in §5.2.6.

CONDITIONS AND TERMINATION CLAUSES. Parties can also use the contract itself to specify (within limits) the conditions under which either or both parties will be released from their obligation to perform. Employment contracts, for example, may allow either the employer or the employee to terminate the relationship at any time, and for any reason. (Indeed, this is the common-law default rule for employment contracts.) Similarly, franchise contracts may specify that the franchise relationship will continue indefinitely unless one or the other party exercises its right to terminate the relationship, often with some advance notice required (*e.g.*, 30 days’ notice of termination).

Other contracts may permit one part to terminate the relationship if certain terms of the contract are violated. Technically, contractual clauses whose violation will release one party from part or all of the contract are referred to in law as conditions. By contrast, covenants or promises are clauses whose violation normally leads to some other default remedy. Violation of a covenant will release the other party from the contract only if the breach is found to be “material” (see the preceding discussion of rescission).

To be sure, just as courts sometimes refuse to enforce liquidated damage clauses (see §5.3.4), they also do not always enforce contractual termination provisions. For example, if termination would inflict on the other party a loss that seems to the court to be excessive (a “forfeiture”), courts may refuse to give effect to an express condition, thus prohibiting the other party from terminating the contract. In addition, clauses that purport to give one party the right to terminate a relationship for any reason will sometimes be interpreted more narrowly by courts, who may refuse to permit terminations that are not made “in good faith” (a phrase whose exact content is difficult to pin down). By requiring some degree of judicial approval of a termination decision, these doctrines thus limit the parties’ ability to use termination as a self-enforcement technique. Some of the economic effects of these limits are discussed in Klein (1980).

---

<sup>113</sup>In some cases, the contract may itself specify that the seller has the right to cure, and may even limit the buyer’s remedy to accepting the seller’s repair or replacement. Even in these cases, though, the UCC allows courts to disregard such a clause (and bring the seller’s right to cure to an end) if the seller’s inability to cure causes this remedy to “fail its essential purpose.”

## 5.5 Other law bearing on contract enforcement

It should also be kept in mind that the enforcement of contracts is often affected by rules and institutions from other fields of law. For example, one common way for parties to enhance the likelihood of contractual performance is to offer collateral; and this device is regulated generally by the law of property and specifically by the specialized law of secured transactions (see Schwartz, 1989; Triantis, 1992). Similarly, contracting parties often enlist third parties as guarantors on their behalf; and the value of such assurance is determined by the law of suretyship (see Katz, 1999).

Conversely, parties' ability to use reputation or repeat dealing as private enforcement devices may be restricted by other fields of law. For example, the law of antitrust generally prohibits concerted boycotts or refusals to deal; and the law of torts may treat some reputation-affecting communications as unfair competition, defamation, or invasion of privacy. To this extent, some private attempts to enforce contracts may be actionable in their own right.

## 6 Conclusions

This chapter is lengthy and many of our conclusions have already been given. Consequently, we limit ourselves here to a few remarks.

Given the vastness of the literature on the law and economics of contracts, even a survey as long as ours must omit certain topics. One topic that has been omitted is the connection between the literature on contracts and those on torts, takings, and regulation. In particular, much of the economic analysis of regulation takes the view that the regulator and the regulated entity are entering into what is, effectively, an agency contract between the regulator (the principal) and the regulated entity (the agent). See, for instance, Laffont and Tirole (1993). But even torts and takings can be related to contracts if, as some analysis has done, one views the state as seeking to approximate, in some way via law, the contract that it would have liked to have written with the tort malffeasor or the owner of the property to be taken if their identity were known in advance.<sup>114</sup>

We have also omitted the entire literature in which the state itself is a party to the contract. Government contracts raises a number of additional issues, including the need for public accountability, risks of corruption and political capture, the problem of establishing credible commitment that will survive changes of governmental regime, and the special difficulties of enforcing contract rights against a sovereign state. For instance, the risk of nations repudiating their debt contracts or abrogating licensing agreements is well documented, especially in the context of developing economies. Creditors or technology providers who contemplate entering into agreements with such governments, accordingly, must find ways to mitigate or insure against such risk. At the same time, citizens and regulators have an interest in preventing state officials from entering

---

<sup>114</sup>See, *e.g.*, Hermalin (1995) for an application of this approach on the takings issue.

into contracts that are not in the public interest *ex ante*. To an extent, these issues relate to our discussion of alternative means of enforcement in §5.4.2 and §5.4.3, but the area is broader than this.

Any chapter of this sort should close with some suggestions for future research. Some suggestions have made already: (i) more economic analysis of non-Anglo-American contract law; (ii) more positive analyses of contract law; (iii) more efforts in modeling to treat monitoring and measurement as endogenous with respect to what information is observable or verifiable; (iv) empirical studies of how courts employ certain rules, such as the parol evidence rule; (v) economic analyses of some of the doctrinal complications associated with the law of contract formation (*e.g.*, promissory estoppel); and (vi) more analysis of the interactions between private and state enforcement of contracts. To this list we would add greater use of new economic paradigms such as behavioral economics. The behavioral paradigm in particular holds out the promise of increased understanding of the phenomenon of bounded rationality, and of legal doctrines that respond to it. Consider, for instance, the literature discussed in §2.3.4 above, as well as more recent work such as that of DellaVigna and Malmendier on issues of contract design and self control and their application to questions of how do health clubs design their contracts (DellaVigna and Malmendier, 2004, in press, respectively).

## References

- Adler, Barry E.**, “The Questionable Ascent of Hadley v. Baxendale,” *Stanford Law Review*, 1999, 51 (6), 1547–1589.
- Aghion, Philippe and Benjamin E. Hermalin**, “Legal Restrictions on Private Contracts Can Enhance Efficiency,” *Journal of Law, Economics, and Organization*, Fall 1990, 6 (2), 381–409.
- **and Patrick Bolton**, “Contracts as a Barrier to Entry,” *American Economic Review*, June 1987, 77 (3), 388–401.
- **, Mathias Dewatripont, and Patrick Rey**, “Renegotiation Design with Unverifiable Information,” *Econometrica*, 1994, 62 (2), 257–282.
- Akerlof, George**, “The Market for ‘Lemons’: Qualitative Uncertainty and the Market Mechanism,” *Quarterly Journal of Economics*, 1970, 84 (3), 488–500.
- Anderlini, Luca and Leonardo Felli**, “Incomplete Written Contracts: Undescribable States of Nature,” *Quarterly Journal of Economics*, November 1994, 109 (4), 1085–1124.
- **, – , and Andrew Postlewaite**, “Should Courts Always Enforce What Contracting Parties Write?,” Working Paper, University of Pennsylvania November 2003.
- Anderson, James and Leslie Young**, “Imperfect Contract Enforcement,” Technical Report 8847, NBER, Cambridge, MA 2002.
- Arrow, Kenneth J.**, *Social Choice and Individual Values*, 2nd ed., New Haven, CT: Yale University Press, 1963.
- **and Gerard Debreu**, “Existence of an Equilibrium for a Competitive Economy,” *Econometrica*, 1954, 22 (3), 265–290.
- Atiyah, P.S.**, *An Introduction to the Law of Contract*, 6th ed., Oxford: Oxford University Press, 2003.
- Ayres, Ian**, “Valuing Modern Contract Scholarship,” *Yale Law Journal*, January 2003, 112 (4), 881–902.
- **and Eric Rasmusen**, “Mutual and Unilateral Mistake in Contract Law,” *Journal of Legal Studies*, 1993, 22 (2), 309–343.
- **and Gregory Klass**, *Insincere Promises: The Law of Misrepresented Intent*, New Haven: Yale University Press, 2005.
- **and Paul M. Goldbart**, “Correlated Values in the Theory of Property and Liability Rules,” *Journal of Legal Studies*, January 2003, 32 (1), 121–151.
- **and Robert H. Gertner**, “Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules,” *Yale Law Journal*, 1989, 99 (2), 87–130.

- **and** –, “Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules,” *Yale Law Journal*, 1992, 101 (4), 729.
- Baird, Douglas and Robert Weisberg**, “Rules, Standards, and the Battle of the Forms: A Reassessment of 2–207,” *Virginia Law Review*, 1982, 68 (6), 1217–1262.
- Bar-Gill, Oren**, “Seduction by Plastic,” *Northwestern University Law Review*, 2005, 98 (4), 1373–1434.
- **and Omri Ben-Shahar**, “The Law of Duress and the Economics of Credible Threats,” *Journal of Legal Studies*, 2004, 33 (2), 391–430.
- Barnett, Randy E.**, “A Consent Theory of Contract,” *Columbia Law Review*, 1986, 86 (2), 269–321.
- Barton, John H.**, “The Economic Basis of Damages for Breach of Contract,” *Journal of Legal Studies*, June 1972, 1 (2), 277–304.
- Bebchuk, Lucian Arye**, “Property Rights and Liability Rules: The *Ex Ante* View of the Cathedral,” *Michigan Law Review*, December 2001, 100 (3), 601–639.
- **and Omri Ben-Shahar**, “Precontractual Reliance,” *Journal of Legal Studies*, 2001, 30 (2, pt. 1), 423–457.
- **and Steven Shavell**, “Information and the Scope of Liability for Breach of Contract: The Rule of *Hadley v. Baxendale*,” *Journal of Law, Economics, and Organization*, Fall 1991, 7 (2), 284–312.
- Ben-Shahar, Omri**, “The Tentative Case Against Flexibility in Commercial Law,” *University of Chicago Law Review*, 1999, 66 (3), 781–820.
- , “Agreeing to Disagree: Filling Gaps in Deliberately Incomplete Contracts,” *Wisconsin Law Review*, 2004, (2), 389–428.
- , “An Ex-Ante View of the Battle of the Forms: Inducing Parties to Draft Reasonable Terms,” *International Review of Law and Economics*, in press.
- **and Lisa Bernstein**, “The Secrecy Interest in Contract Law,” *Yale Law Journal*, 2000, 109 (8), 1885–1925.
- Bernheim, B. Douglas and Michael D. Whinston**, “Incomplete Contracts and Strategic Ambiguity,” *American Economic Review*, 1998, 88 (4), 902–932.
- Bernstein, Lisa**, “Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry,” *Journal of Legal Studies*, January 1992, 21 (1), 115–157.
- , “Merchant Law in a Merchant Court: Rethinking the Code’s Search for Immanent Business Norms,” *University of Pennsylvania Law Review*, 1996, 144 (5), 1765–1821.

- , “The Questionable Empirical Basis of Article 2’s Incorporation Strategy: A Preliminary Study,” *University of Chicago Law Review*, 1999, 66 (3), 710–780.
- , “Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions,” *Michigan Law Review*, 2001, 99 (7), 1724–1790.
- Birmingham, Robert L.**, “Breach of Contract, Damage Measures, and Economic Efficiency,” *Rutgers Law Review*, Winter 1970, 24 (2), 273–292.
- Bishop, William**, “The Choice of Remedy for Breach of Contract,” *Journal of Legal Studies*, June 1985, 14 (2), 299–320.
- Blackorby, Charles and David Donaldson**, “Cash versus Kind, Self-selection, and Efficient Transfers,” *American Economic Review*, 1988, 78 (4), 691–700.
- Bolton, Patrick and Mathias Dewatripont**, *Contract Theory*, Cambridge, MA: MIT Press, 2005.
- Bonell, Michael Joachim**, *An International Restatement Of Contract Law: The UNIDROIT Principles Of International Commercial Contracts*, Irvington-on-Hudson, NY: Transnational Publishers, 1997.
- Boolos, George S., John P. Burgess, and Richard C. Jeffrey**, *Computability and Logic*, 4th ed., Cambridge, England: Cambridge University Press, 2002.
- Bork, Robert H.**, *The Antitrust Paradox*, New York: Basic Books, 1978.
- Buchanan, James M.**, “In Defense of Caveat Emptor,” *University of Chicago Law Review*, 1970, 38 (1), 64–73.
- Bull, Clive**, “The Existence of Self-Enforcing Implicit Contracts,” *Quarterly Journal of Economics*, 1987, 102 (1), 147–160.
- Caillaud, Bernard and Benjamin E. Hermalin**, “Hidden Action and Incentives,” Technical Report, University of California, Berkeley, <http://faculty.haas.berkeley.edu/hermalin/agencyread.pdf> 2000.
- **and** – , “Hidden-Information Agency,” Technical Report, University of California, Berkeley, <http://faculty.haas.berkeley.edu/hermalin/mechread.pdf> 2000.
- Calabresi, Guido and A. Douglas Melamed**, “Property Rules, Liability Rules, and Inalienability: One View of the Cathedral,” *Harvard Law Review*, 1972, 85 (6), 1089–1128.
- Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O’Donoghue, and Matthew Rabin**, “Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism,” *University of Pennsylvania Law Review*, 2003, 151 (3), 1211–1254.

- Charny, David**, “Nonlegal Sanctions in Commercial Relationships,” *Harvard Law Review*, December 1990, *104* (2), 373–467.
- , “Hypothetical Bargains: The Normative Structure of Contract Interpretation,” *Michigan Law Review*, 1991, *89* (7), 1815–1879.
- Che, Yeon-Koo and Donald B. Hausch**, “Cooperative Investments and the Value of Contracting,” *American Economic Review*, 1999, *89* (1), 125–147.
- Cheung, Steven N.S.**, “Transaction Costs, Risk Aversion, and the Choice of Contractual Arrangement,” *Journal of Law & Economics*, 1969, *12* (1), 23–42.
- Chirelstein, Marvin A.**, *Concepts and Case Analysis in the Law of Contracts*, New York: Foundation Press, 2006.
- Cho, In-Koo and David M. Kreps**, “Signaling Games and Stable Equilibria,” *Quarterly Journal of Economics*, May 1987, *102* (2), 179–222.
- Chung, Tai-Yeong**, “Incomplete Contracts, Specific Investments, and Risk Sharing,” *Review of Economic Studies*, 1991, *58* (5), 1031–1042.
- Clarkson, Kenneth W., Roger LeRoy Miller, and Timothy J. Muris**, “Liquidated Damages vs. Penalties: Sense or Nonsense?,” *Wisconsin Law Review*, 1978, (2), 351–390.
- Coase, Ronald H.**, “The Problem of Social Cost,” *Journal of Law and Economics*, April 1960, *3* (1), 1–44.
- Cook, Philip J. and Daniel A. Graham**, “The Demand for Insurance and Protection: The Case of Irreplaceable Commodities,” *Quarterly Journal of Economics*, February 1977, *91* (1), 143–156.
- Cooter, Robert D.**, “The Cost of Coase,” *Journal of Legal Studies*, 1982, *11* (1), 1–31.
- , “Unity in Tort, Contract, and Property: The Model of Precaution,” *California Law Review*, January 1985, *73* (1), 1–51.
- **and Melvin A. Eisenberg**, “Damages for Breach of Contract,” *California Law Review*, October 1985, *73* (5), 1432–1481.
- , **Stephen Marks, and Robert Mnookin**, “Bargaining in the Shadow of the Law: A Testable Model of Strategic Behavior,” *Journal of Legal Studies*, 1982, *11* (2), 225–251.
- Craswell, Richard**, “Precontractual Investigation as an Optimal Precaution Problem,” *Journal of Legal Studies*, June 1988, *17* (2), 401–436.
- , “Contract Law, Default Rules, and the Philosophy of Promising,” *Michigan Law Review*, 1989, *88* (3), 489–529.

- 
- , “Performance, Reliance, and One-Sided Information,” *Journal of Legal Studies*, June 1989, 18 (2), 365–401.
- , “Insecurity, Repudiation, and Cure,” *Journal of Legal Studies*, June 1990, 19 (2), 399–434.
- , “Passing on the Costs of Legal Rules: Efficiency and Distribution in Buyer-Seller Relationships,” *Stanford Law Review*, 1991, 43 (2), 361–398.
- , “Property Rules and Liability Rules in Unconscionability and Related Doctrines,” *University of Chicago Law Review*, 1993, 60 (1), 1–65.
- , “Damage Multipliers in Market Relationships,” *Journal of Legal Studies*, June 1996, 25 (2), 463–492.
- , “Offer, Acceptance, and Efficient Reliance,” *Stanford Law Review*, 1996, 48 (3), 481–553.
- , “Deterrence and Damages: The Multiplier Principle and Its Alternatives,” *Michigan Law Review*, June 1999, 97 (7), 2185–2238.
- , “Do Trade Customs Exist?,” in Jody S. Kraus and Steven D. Walt, eds., *The Jurisprudential Foundations of Corporate and Commercial Law*, Cambridge: Cambridge University Press, 2000.
- , “In that Case, What is the Question? Economics and the Demands of Contract Theory,” *Yale Law Journal*, January 2003, 112 (4), 903–924.
- , “Taking Information Seriously: Misrepresentation and Nondisclosure in Contract Law and Elsewhere,” *Virginia Law Review*, 2006, 92 (4), 565–632.
- Crocker, Keith and Scott E. Masten**, “Mitigating Contractual Hazards: Unilateral Options and Contract Length,” *RAND Journal of Economics*, 1988, 19 (3), 327–343.
- Cungu, Azeta and Johan Swinnen**, “Investment and Contract Enforcement in Transition: Evidence from Hungary,” Technical Report 127, LICOS Centre for Transition Economics, Katholieke Universiteit Leuven, Leuven, Belgium 2003.
- Dal Bó, Ernesto**, “Bribing Voters,” Working Paper, University of California, Berkeley 2003.
- Darby, Michael R. and Edi Karni**, “Free Competition and the Optimal Amount of Fraud,” *Journal of Law and Economics*, April 1973, 16 (1), 67–88.
- Debreu, Gerard**, *Theory of Value*, New Haven, CT: Yale University Press, 1959.

- Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini**, "Recent Developments in Modeling Unforeseen Contingencies," *European Economic Review*, 1998, 42 (3), 523–542.
- , – , and – , "Representing Preferences with a Unique Subjective State Space," *Econometrica*, July 2001, 69 (4), 891–934.
- DellaVigna, Stefano and Ulrike Malmendier**, "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics*, May 2004, 119 (2), 353–402.
- and – , "Paying not to go to the Gym," *American Economic Review*, in press.
- Demski, Joel S. and David E.M. Sappington**, "Resolving Double Moral Hazard Problems with Buyout Agreements," *RAND Journal of Economics*, 1991, 22 (2), 232–240.
- Diamond, Peter A.**, "Search Theory," in J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave: A Dictionary of Economics*, New York: Macmillan, 1987.
- and **Eric Maskin**, "An Equilibrium Analysis of Search and Breach of Contract I: Steady States," *Bell Journal of Economics*, Spring 1979, 10 (1), 282–316.
- and – , "An Equilibrium Analysis of Search and Breach of Contract II: A Non-Steady State Example," *Journal of Economic Theory*, October 1981, 25 (2), 165–195.
- Dworkin, Ronald M.**, "Is Wealth a Value?," *Journal of Legal Studies*, 1980, 9 (2), 191–226.
- Dye, Ronald A.**, "Costly Contract Contingencies," *International Economic Review*, February 1985, 26 (1), 233–250.
- Easterbrook, Frank H.**, "The Supreme Court, 1983 Term Foreword: The Court and the Economic System," *Harvard Law Review*, 1984, 98 (1), 4–60.
- Edlin, Aaron S. and Alan Schwartz**, "Optimal Penalties in Contracts," *Chicago-Kent Law Review*, 2003, 78 (1), 33–54.
- and **Benjamin E. Hermalin**, "Contract Renegotiation and Options in Agency Problems," *Journal of Law, Economics, and Organization*, 2000, 16 (2), 395–423.
- and – , "Implementing the First Best in an Agency Relationship with Renegotiation: A Corrigendum," *Econometrica*, September 2001, 69 (5), 1391–1395.

- **and Stefan Reichelstein**, “Holdups, Standard Breach Remedies, and Optimal Investments,” *American Economic Review*, 1996, 86 (3), 478–501.
- Eisenberg, Melvin A.**, “The Limits of Cognition and the Limits of Contracts,” *Stanford Law Review*, January 1995, 47 (2), 211–259.
- Epstein, Richard**, “Unconscionability: A Critical Appraisal,” *Journal of Law & Economics*, 1975, 18, 293–315.
- Farnsworth, E. Allan**, “Your Loss or my Gain? The Dilemma of the Disgorgement Principle in Breach of Contract,” *Yale Law Journal*, 1985, 94 (6), 1339–1393.
- , *Farnsworth on Contracts*, 3rd ed., New York: Aspen Publishers, 2004.
- Farrell, Joseph**, “Cheap Talk, Coordination, and Entry,” *RAND Journal of Economics*, 1987, 18 (1), 34–39.
- , “Information and the Coase Theorem,” *Journal of Economic Perspectives*, Fall 1987, 1 (2), 113–129.
- , “Meaning and Credibility in Cheap Talk Games,” *Games and Economic Behavior*, 1993, 5 (4), 514–531.
- Fishman, Michael J. and Kathleen M. Hagerty**, “Mandatory Versus Voluntary Disclosure in Markets with Informed and Uninformed Customers,” *Journal of Law, Economics, and Organization*, 2003, 19 (1), 45–63.
- Fong, Yuk-fai**, “When do Experts Cheat and Whom do They Target?,” *RAND Journal of Economics*, Spring 2005, 36 (1), 113–130.
- Fried, Charles**, *Contract as Promise: A Theory of Contractual Obligation*, Cambridge, MA: Harvard University Press, 1981.
- Friedmann, Daniel**, “The Efficient Breach Fallacy,” *Journal of Legal Studies*, 1989, 18 (1), 1–24.
- Fudenberg, Drew and Jean Tirole**, “Moral Hazard and Renegotiation in Agency Contracts,” *Econometrica*, November 1990, 58 (6), 1279–1319.
- **and** – , *Game Theory*, Cambridge, MA: MIT Press, 1991.
- Fuller, Lon L.**, “Consideration as Form,” *Columbia Law Review*, 1941, 41 (5), 799–824.
- **and William R. Perdue**, “The Reliance Interest in Contract Damages,” *Yale Law Journal*, November 1936, 46 (1), 52–96. January 1937, 46(3), 373–420.
- Gergen, Mark**, “The Use of Open Terms in Contract,” *Columbia Law Review*, 1992, 92 (5), 997–1081.

- Gibbons, Robert**, *Game Theory for Applied Economists*, Princeton, NJ: Princeton University Press, 1992.
- Goetz, Charles J. and Robert E. Scott**, "Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach," *Columbia Law Review*, May 1977, 77 (4), 554–594.
- **and** – , "Enforcing Promises: An Examination of the Basis of Contract," *Yale Law Journal*, 1980, 89 (7), 1261–1322.
- **and** – , "Principles of Relational Contracts," *Virginia Law Review*, 1981, 67 (6), 1089–1150.
- **and** – , "The Mitigation Principle: Toward a General Theory of Contractual Obligation," *Virginia Law Review*, September 1983, 69 (6), 967–1024.
- **and** – , "The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms," *California Law Review*, 1985, 73 (2), 261–322.
- Goldberg, Victor P.**, "Bloomer Girl Revisited or How to Frame an Unmade Picture," *Wisconsin Law Review*, 1998, 1998 (4), 1051–1084.
- , "In Search of Best Efforts: Reinterpreting Bloor v. Falstaff," *St. Louis University Law Journal*, 2000, 44 (4), 1465–1485.
- **and John R. Erickson**, "Quantity and Price Adjustment in Long-term Contracts: A Case Study of Petroleum Coke," *Journal of Law & Economics*, 1987, 30 (2), 369–398.
- Greif, Avner and Eugene Kandel**, "Contract Enforcement Institutions: Historical Perspective and Current Status in Russia," in Edward P. Lazear, ed., *Economic Transition in Eastern Europe and Russia: Realities of Reform*, Stanford, CA: Hoover Institution Press, 1995, pp. 291–321.
- Grossman, Sanford J.**, "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 1981, 21, 461–83.
- **and Oliver D. Hart**, "An Analysis of the Principal-Agent Problem," *Econometrica*, January 1983, 51 (1), 7–46.
- **and** – , "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, August 1986, 94 (4), 691–719.
- Güth, Werner, Rolf Schmittberger, and Bernd Schwarze**, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, 1982, 3 (4), 367–388.

- Hadfield, Gillian K.**, “Problematic Relations: Franchising and the Law of Incomplete Contracts,” *Stanford Law Review*, 1990, 42 (4), 927–992.
- , “Judicial Competence and the Interpretation of Incomplete Contracts,” *Journal of Legal Studies*, 1994, 23 (1), 159–184.
- Harris, Milton and Artur Raviv**, “Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement,” *American Economic Review*, March 1978, 68 (1), 20–30.
- Hart, Oliver D.**, “Incomplete Contracts,” in J. Eatwell, M. Milgate, and P. Newman, eds., *Allocation, Information, and Markets*, New York: W.W. Norton & Co., 1987.
- , “An Economist’s Perspective on the Theory of the Firm,” *Columbia Law Review*, 1989, 89 (7), 1757–1774.
- **and John H. Moore**, “Default and Renegotiation: A Dynamic Model of Debt,” *Quarterly Journal of Economics*, February 1998, 113 (1), 1–41.
- **and** – , “Foundations of Incomplete Contracts,” *Review of Economic Studies*, January 1999, 66 (1), 115–138.
- **and** – , “Agreeing Now to Agree Later: Contracts that Rule Out but do not Rule In,” 2004. Harvard Law and Economics Discussion Paper No. 465.
- Hermalin, Benjamin E.**, “An Economic Analysis of Takings,” *Journal of Law, Economics, and Organization*, 1995, 11 (1), 64–86.
- , “Economics and Corporate Culture,” in Cary L. Cooper, Sue Cartwright, and P. Christopher Earley, eds., *The International Handbook of Organizational Culture and Climate*, Chichester, England: John Wiley & Sons, Ltd., 2001.
- , “Adverse Selection, Short-Term Contracting, and the Underprovision of On-the-Job Training,” *Contributions to Economic Analysis & Policy*, 2002, 1 (1), Article 5.
- **and Michael L. Katz**, “Moral Hazard and Verifiability: The Effects of Renegotiation in Agency,” *Econometrica*, November 1991, 59 (6), 1735–1753.
- **and** – , “Judicial Modification of Contracts Between Sophisticated Parties: A More Complete View of Incomplete Contracts and Their Breach,” *Journal of Law, Economics, and Organization*, Fall 1993, 9 (2), 230–255.
- **and** – , “Privacy, Property Rights, & Efficiency: The Economics of Privacy as Secrecy,” *Quantitative Marketing and Economics*, in press.
- Hillman, Robert A.**, “Policing Contract Modifications Under the UCC: Good Faith and the Doctrine of Economic Duress,” *Iowa Law Review*, 1979, 64 (4), 849–902.

- Holmes, Oliver Wendell**, "The Path of the Law," *Harvard Law Review*, 1897, 10 (8), 457–478.
- Holmstrom, Bengt**, "Moral Hazard in Teams," *Bell Journal of Economics*, Autumn 1982, 13 (2), 324–340.
- Honnold, John O.**, *Uniform Law For International Sales Under The 1980 United Nations Convention*, Cambridge, MA: Kluwer Law International, 1999.
- Jackson, Thomas H.**, "Anticipatory Repudiation and the Temporal Element of Contract Law: An Economic Inquiry into Contract Damages in Cases of Prospective Nonperformance," *Stanford Law Review*, November 1978, 31 (1), 69–119.
- Johnston, Jason Scott**, "Strategic Bargaining and the Economic Theory of Contract Default Rules," *Yale Law Journal*, December 1990, 100 (3), 615–664.
- Jolls, Christine M.**, "Contracts as Bilateral Commitments: A New Perspective on Contract Modification," *Journal of Legal Studies*, 1997, 26 (1), 203–237.
- , **Cass R. Sunstein**, and **Richard Thaler**, "A Behavioral Approach to Law and Economics," *Stanford Law Review*, May 1998, 50 (5), 1471–1550.
- Joskow, Paul L.**, "Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets," *American Economic Review*, 1987, 77 (1), 168–85.
- Kahan, Marcel and Michael Klausner**, "Standardization and Innovation in Corporate Contracting (or The Economics of Boilerplate)," *Virginia Law Review*, 1997, 83 (4), 713–770.
- Kaplow, Louis**, "Rules Versus Standards: An Economic Analysis," *Duke Law Journal*, 1992, 42 (3), 557–629.
- , "A Model of the Optimal Complexity of Legal Rules," *Journal of Law, Economics, & Organization*, April 1995, 11 (1), 150–163.
- and **Steven Shavell**, "Property Rules Versus Liability Rules: An Economic Analysis," *Harvard Law Review*, February 1996, 109 (4), 713–790.
- Katz, Avery W.**, "The Strategic Structure of Offer and Acceptance: Game Theory and the Law of Contract Formation," *Michigan Law Review*, November 1990, 89 (2), 215–295.
- , "Your Terms or Mine? The Duty to Read the Fine Print in Contracts," *RAND Journal of Economics*, Winter 1990, 21 (4), 518–537.

- , “Transaction Costs and the Legal Mechanics of Contract Formation: When Should Silence in the Face of an Offer be Construed as Acceptance?,” *Journal of Law, Economics, and Organization*, April 1993, 9 (1), 77–97.
- , “Positivism and the Separation of Law and Economics,” *Michigan Law Review*, 1996, 94 (7), 2229–2269.
- , “Taking Private Ordering Seriously,” *University of Pennsylvania Law Review*, May 1996, 144 (5), 1745–1763.
- , “When Should an Offer Stick? The Economics of Promissory Estoppel in Preliminary Negotiations,” *Yale Law Journal*, 1996, 105 (5), 1249–1309.
- , “An Economic Analysis of the Guaranty Contract,” *University of Chicago Law Review*, 1999, 66 (1), 47–116.
- , “The Economics of Form and Substance in Contract Interpretation,” *Columbia Law Review*, March 2004, 103 (2), 496–538.
- Kelman, Steven**, *What Price Incentives? Economists and the Environment*, Boston: Auburn House Publishing Co., 1981.
- Kennedy, Duncan**, “Distributive and Paternalist Motives in Contract and Tort Law, with Special Reference to Compulsory Terms and Unequal Bargaining Power,” *Maryland Law Review*, 1982, 41 (4), 563–658.
- Kenney, Roy W. and Benjamin Klein**, “The Economics of Block Booking,” *Journal of Law & Economics*, 1983, 26 (3), 497–540.
- **and** – , “How Block Booking Facilitated Self-Enforcing Film Contracts,” *Journal of Law & Economics*, 2000, 43 (2), 427–435.
- Kessler, Friedrich**, “Contracts of Adhesion—Some Thoughts about Freedom of Contract,” *Columbia Law Review*, 1943, 43 (5), 629–642.
- Klein, Benjamin**, “Transaction Cost Determinants of Unfair Contractual Arrangements,” *American Economic Review*, May 1980, 70 (2), 356–360.
- **and Keith Leffler**, “The Role of Market Forces in Assuring Contractual Performance,” *Journal of Political Economy*, 1981, 89 (4), 615–641.
- **and Kevin Murphy**, “Vertical Restraints as Contract Enforcement Mechanisms,” *Journal of Law & Economics*, 1988, 31 (2), 265–297.
- Kornhauser, Lewis A.**, “Reliance, Reputation, and Breach of Contract,” *Journal of Law and Economics*, October 1983, 26 (2), 691–706.
- Korobkin, Russell B.**, “Inertia and Preference in Contract Negotiation: The Psychological Power of Default Rules and Form Terms,” *Vanderbilt Law Review*, 1998, 51 (6), 1583–1652.

- , “The Status Quo Bias and Contract Default Rules,” *Cornell Law Review*, 1998, 83 (3), 608–687.
- **and Thomas S. Ulen**, “Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics,” *California Law Review*, July 2000, 88 (4), 1051–1142.
- Krasa, Stefan and Anne P. Villamil**, “Optimal Contracts When Enforcement is a Decision Variable,” *Econometrica*, January 2000, 68 (1), 119–134.
- Kraus, Jody S.**, “Decoupling Sales Law from the Acceptance-Rejection Fulcrum,” *Yale Law Journal*, October 1994, 104 (1), 129–205.
- Kreps, David M.**, *A Course in Microeconomic Theory*, Princeton, NJ: Princeton University Press, 1990.
- Kronman, Anthony T.**, “Mistake, Disclosure, Information, and the Law of Contracts,” *Journal of Legal Studies*, 1978, 7 (1), 1–34.
- , “Specific Performance,” *University of Chicago Law Review*, Winter 1978, 45 (2), 351–382.
- , “Living in the Law,” *University of Chicago Law Review*, 1987, 54 (3), 835–876.
- Kull, Andrew**, “Mistake, Frustration, and the Windfall Principle of Contract Remedies,” *Hastings Law Journal*, November 1991, 43 (1), 1–55.
- Laffont, Jean-Jacques**, *Fundamentals of Public Economics*, Cambridge, MA: MIT Press, 1988.
- **and David Martimont**, *The Theory of Incentives*, Princeton, NJ: Princeton University Press, 2002.
- **and Jean Tirole**, “The Dynamics of Incentive Contracts,” *Econometrica*, 1988, 56 (5), 1153–1175.
- **and –**, *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press, 1993.
- Landa, Janet T.**, “A Theory of the Ethnically Homogeneous Middleman Group: An Institutional Alternative to Contract Law,” *Journal of Legal Studies*, June 1981, 10 (2), 349–362.
- Lando, Henrik and Caspar Rose**, “On The Enforcement Of Specific Performance In Civil Law Countries,” *International Review of Law and Economics*, 2004, 24 (4), 473–487.
- Lando, Ole and Hugh Beale, eds**, *Contract Law*, The Hague: Kluwer Law International, 2002.

- Levin, Jonathan**, "Relational Incentive Contracts," *American Economic Review*, June 2003, *93* (3), 835–857.
- Llewellyn, Karl N.**, *The Common Law Tradition: Deciding Appeals*, Boston: Little, Brown and Co., 1960.
- MacLeod, W. Bentley and James M. Malcolmson**, "Investments, Holdup, and the Form of Market Contracts," *American Economic Review*, 1993, *83* (4), 811–837.
- MacLeod, W. Bentley and James Malcolmson**, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, March 1989, *56* (2), 447–480.
- Macneil, Ian R.**, "Efficient Breach of Contract: Circles in the Sky," *Virginia Law Review*, May 1982, *68* (5), 947–969.
- Mankiw, N. Gregory and Michael D. Whinston**, "Free Entry and Social Inefficiency," *RAND Journal of Economics*, 1986, *17* (1), 48–58.
- Mann, Ronald J.**, "Explaining the Pattern of Secured Credit," *Harvard Law Review*, 1997, *110* (3), 625–683.
- , "The Role of Secured Credit in Small-Business Lending," *Georgetown Law Journal*, 1997, *86* (1), 1–44.
- , "Verification Institutions in Financing Transactions," *Georgetown Law Journal*, 1999, *87* (7), 2225–2272.
- Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green**, *Microeconomic Theory*, Oxford: Oxford University Press, 1995.
- Maskin, Eric and Jean Tirole**, "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies*, January 1999, *66* (1), 83–114.
- Masten, Scott E. and Edward A. Snyder**, "United States versus United Shoe Machinery Corporation: On the Merits," *Journal of Law & Economics*, 1993, *36* (1), 33–70.
- Matthews, Steven and Andrew Postlethwaite**, "Quality Testing and Disclosure," *RAND Journal of Economics*, 1985, *16* (3), 328–340.
- **and John H. Moore**, "Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening," *Econometrica*, March 1987, *55* (2), 441–467.
- Medema, Steven G. and Richard O. Zerbe**, "The Coase Theorem," in Boudewijn Bouckaert and Gerrit De Geest, eds., *Encyclopedia of Law and Economics*, Vol. 1, Cheltenham, England: Edward Elgar, 2000, chapter 0730, pp. 836–892.

- Menell, Peter S.**, “A Note on Private versus Social Incentives to Sue in a Costly Legal System,” *Journal of Legal Studies*, 1983, *12* (1), 41–52.
- Muris, Timothy J.**, “The Costs of Freely Granting Specific Performance,” *Duke Law Journal*, December 1982, (6), 1053–1069.
- , “Cost of Completion or Diminution in Market Value: The Relevance of Subjective Value,” *Journal of Legal Studies*, June 1983, *12* (2), 379–400.
- Myerson, Roger B.**, “Justice, Institutions, and Multiple Equilibria,” *Chicago Journal of International Law*, 2004, *5* (1), 91–108.
- Nash, John F.**, “The Bargaining Problem,” *Econometrica*, April 1950, *18* (2), 155–162.
- , “Noncooperative Games,” *Annals of Mathematics*, 1951, *54* (2), 286–295.
- Nöldeke, Georg and Klaus M. Schmidt**, “Option Contracts and Renegotiation: A Solution to the Hold-up Problem,” *RAND Journal of Economics*, 1995, *26* (2), 163–179.
- and —, “Sequential Investments and Options to Own,” *RAND Journal of Economics*, 1998, *29* (4), 633–653.
- Okun, Arthur M.**, *Equality and Efficiency: The Big Tradeoff*, Washington, DC: The Brookings Institution, 1975.
- Pildes, Richard and Elizabeth Anderson**, “Slingshot Arrows at Democracy: Social Choice Theory, Value Pluralism, and Democratic Politics,” *Columbia Law Review*, 1990, *90* (8), 2121–2214.
- Pirrong, Stephen Craig**, “Contracting Practices in Bulk Shipping Markets: A Transactions Cost Explanation,” *Journal of Law & Economics*, 1993, *36* (2), 937–976.
- Polinsky, A. Mitchell**, “Economic Analysis as a Potentially Defective Product: A Buyer’s Guide to Posner’s Economic Analysis of Law,” *Harvard Law Review*, 1974, *87* (8), 1655–1681.
- , “Risk Sharing Through Breach of Contract Remedies,” *Journal of Legal Studies*, June 1983, *12* (2), 427–444.
- and **Steven Shavell**, “Punitive Damages: An Economic Analysis,” *Harvard Law Review*, February 1998, *111* (4), 869–962.
- Posner, Eric A.**, “Contract Law in the Welfare State: A Defense of the Unconscionability Doctrine, Usury Laws, and Related Limitations on the Freedom to Contract,” *Journal of Legal Studies*, 1995, *24* (2), 283–319.
- , “The Parol Evidence Rule, the Plain Meaning Rule, and the Principles of Contractual Interpretation,” *University of Pennsylvania Law Review*, 1998, *146* (2), 533–577.

- , “Economic Analysis of Contract Law After Three Decades: Success or Failure?,” *Yale Law Journal*, January 2003, 112 (4), 829–880.
- , “There Are No Penalty Default Rules in Contract Law,” 2005. University of Chicago Law & Economics, Olin Working Paper No. 237.
- Posner, Richard A.**, “Taxation by Regulation,” *Bell Journal of Economics*, 1971, 2 (1), 22–50.
- , *Antitrust Law: An Economic Perspective*, Chicago: University of Chicago Press, 1976.
- , “Gratuitous Promises in Economics and Law,” *Journal of Legal Studies*, 1977, 6 (2), 411–426.
- , *Economic Analysis of Law*, 6th ed., Boston: Little, Brown & Co., 2003.
- **and Andrew M. Rosenfield**, “Impossibility and Related Doctrines in Contract Law: An Economic Analysis,” *Journal of Legal Studies*, 1977, 6 (1), 83–118.
- Priest, George L.**, “Breach and Remedy for the Tender of Nonconforming Goods Under the Uniform Commercial Code: An Economic Approach,” *Harvard Law Review*, 1978, 91 (5), 960–1001.
- Quillen, Gwyn D.**, “Contract Damages and Cross-Subsidization,” *Southern California Law Review*, May 1988, 61 (4), 1125–1141.
- Rabin, Matthew**, “Incorporating Fairness into Game Theory and Economics,” *American Economic Review*, December 1993, 83 (5), 1281–1302.
- , “Psychology and Economics,” *Journal of Economic Literature*, March 1998, 36 (1), 11–46.
- Rachlinski, Jeffrey J.**, “A Positive Psychological Theory of Judging in Hindsight,” *University of Chicago Law Review*, 1998, 65 (2), 571–625.
- Radin, Margaret Jane**, “Market-Inalienability,” *Harvard Law Review*, 1987, 100 (8), 1849–1937.
- Rakoff, Todd**, “Contracts of Adhesion: An Essay in Reconstruction,” *Harvard Law Review*, 1983, 96 (6), 1173–1284.
- Rasmusen, Eric B.**, “Explaining Incomplete Contracts as the Result of Contract-Reading Costs,” *Advances in Economic Analysis & Policy*, 2001, 1 (1), Article 2.
- , **J. Mark Ramseyer, and John S. Wiley**, “Naked Exclusion,” *American Economic Review*, December 1991, 81 (5), 1137–1145.
- Rea, Samuel A.**, “Nonpecuniary Loss and Breach of Contract,” *Journal of Legal Studies*, January 1982, 11 (1), 35–53.

- , “Efficiency Implications of Penalties and Liquidated Damages,” *Journal of Legal Studies*, January 1984, 13 (1), 147–167.
- Rogerson, William P.**, “Efficient Reliance and Damage Measures for Breach of Contract,” *RAND Journal of Economics*, 1984, 15 (1), 39–53.
- , “Contractual Solutions to the Hold-up Problem,” *Review of Economic Studies*, 1992, 59 (4), 777–793.
- Rubinstein, Ariel**, *Modeling Bounded Rationality* Zeuthen Lecture Book series, Cambridge, MA: MIT Press, 1998.
- Samuelson, William**, “A Comment on the Coase Theorem,” in Alvin E. Roth, ed., *Game-Theoretic Models of Bargaining*, Cambridge: Cambridge University Press, 1985, pp. 321–339.
- Schelling, Thomas C.**, *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, 1960.
- Schlechtreim, Peter**, *Commentary on the UN Convention on the International Sale of Goods*, Oxford: Clarendon Press, 1998.
- Schwartz, Alan**, “The Case for Specific Performance,” *Yale Law Journal*, December 1979, 89 (2), 271–306.
- , “A Theory of Loan Priorities,” *Journal of Legal Studies*, 1989, 18 (2), 209–261.
- , “The Myth that Promisees Prefer Supra-compensatory Remedies: An Analysis of Contracting for Damage Measures,” *Yale Law Journal*, November 1990, 100 (2), 369–407.
- , “Relational Contracts In The Courts: An Analysis Of Incomplete Agreements And Judicial Strategies,” *Journal of Legal Studies*, 1992, 21 (2), 271–318.
- **and Joel Watson**, “The Law and Economics of Costly Contracting,” *Journal of Law, Economics, and Organization*, 2003, 20 (1), 2–31.
- **and Louis L. Wilde**, “Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis,” *University of Pennsylvania Law Review*, 1979, 127 (3), 630–682.
- **and Robert E. Scott**, “Contract Theory and the Limits of Contract Law,” *Yale Law Journal*, 2004, 113 (3), 541–620.
- Schweizer, Urs**, “Externalities and the Coase Theorem: Hypothesis or Result?,” *Journal of Institutional and Theoretical Economics*, 1988, 144 (2), 245–266.

- Scott, Robert E.**, “A Relational Theory of Secured Financing,” *Columbia Law Review*, 1986, *86* (5), 901–977.
- , “The Case for Formalism in Relational Contract,” *Northwestern University Law Review*, 2000, *94* (3), 847–876.
- , “The Rise and Fall of Article 2,” *Louisiana Law Review*, 2002, *62* (4), 1009–1064.
- **and George Triantis**, “Anticipating Litigation in Contract Design,” *Yale Law Journal*, January 2006, *115* (4), 814–879.
- Segal, Ilya R.**, “Complexity and Renegotiation: A Foundation for Incomplete Contracts,” *Review of Economic Studies*, January 1999, *66* (1), 57–82.
- , “Contracting with Externalities,” *Quarterly Journal of Economics*, May 1999, *114* (2), 337–388.
- **and Michael D. Whinston**, “Naked Exclusion: Comment,” *American Economic Review*, March 2000, *90* (1), 296–309.
- **and** – , “Robust Predictions for Bilateral Contracting with Externalities,” *Econometrica*, May 2003, *71* (3), 757–791.
- Sen, Amartya**, “The Impossibility of a Paretian Liberal,” *Journal of Political Economy*, 1970, *78* (1), 152–157.
- Shapiro, Carl**, “Premiums for High Quality Products as Returns to Reputation,” *Quarterly Journal of Economics*, 1983, *98* (4), 659–680.
- Shavell, Steven**, “Damage Measures for Breach of Contract,” *Bell Journal of Economics*, 1980, *11* (2), 466–490.
- , “The Design of Contracts and Remedies for Breach,” *Quarterly Journal of Economics*, February 1984, *99* (1), 121–148.
- , “An Economic Analysis of Altruism and Deferred Gifts,” *Journal of Legal Studies*, 1991, *20* (2), 401–421.
- , “Acquisition and Disclosure of Information Prior to Sale,” *RAND Journal of Economics*, 1994, *25* (1), 20–36.
- , “Contracts, Holdup, and Legal Intervention,” March 2005. Harvard Law School Discussion Paper.
- , “Specific Performance versus Damages for Breach of Contract: An Economic Analysis,” *Texas Law Review*, March 2006, *84* (4), 831–876.
- , “On the Writing and the Interpretation of Contracts,” *Journal of Law, Economics, and Organization*, in press.

- **and Louis Kaplow**, *Fairness versus Welfare*, Cambridge, MA: Harvard University Press, 2002.
- Shepard, Andrea**, “Licensing to Enhance Demand for New Technologies,” *RAND Journal of Economics*, 1987, 18 (3), 360–368.
- Simon, Herbert A.**, “Theories of Bounded Rationality,” in C.B. McGuire and R. Radner, eds., *Decision and Organization: A volume in Honor of Jacob Marschak*, Amsterdam: North Holland, 1972.
- Spence, A. Michael**, “Job Market Signaling,” *Quarterly Journal of Economics*, 1973, 87 (3), 355–374.
- , “Monopoly, Quality, and Regulation,” *Bell Journal of Economics*, 1975, 6 (2), 417–429.
- Spier, Kathryn E.**, “Incomplete Contracts and Signalling,” *RAND Journal of Economics*, Autumn 1992, 23 (3), 432–443.
- , “Settlement Bargaining and the Design of Damage Awards,” *Journal of Law, Economics, and Organization*, April 1994, 10 (1), 84–95.
- **and Michael D. Whinston**, “On the Efficiency of Privately Stipulated Damages for Breach of Contract: Entry Barriers, Reliance, and Renegotiation,” *RAND Journal of Economics*, 1995, 26 (2), 180–202.
- Sutton, John**, “Non-Cooperative Bargaining Theory: An Introduction,” *Review of Economic Studies*, 1986, 53 (5), 709–724.
- Thomas, Jonathan and Tim Worrall**, “Self-Enforcing Wage Contracts,” *Review of Economic Studies*, 1988, 55 (4), 541–553.
- Tirole, Jean**, *The Theory of Industrial Organization*, Cambridge, MA: MIT Press, 1988.
- , “Incomplete Contracts: Where Do We Stand?,” *Econometrica*, July 1999, 67 (4), 741–781.
- Tobin, James**, “On Limiting the Domain of Inequality,” *Journal of Law & Economics*, 1970, 13 (2), 263–277.
- Townsend, Robert M.**, “Optimal Contracts and Competitive Markets with Costly State Verification,” *Journal of Economic Theory*, 1979, 20, 265–293.
- Treitel, G.H., ed.**, *The Law of Contract*, 11th ed., London: Sweet & Maxwell, 2003.
- Triantis, Alexander J. and George G. Triantis**, “Timing Problems in Contract Breach Decisions,” *Journal of Law and Economics*, April 1998, 41 (1), 163–207.

- Triantis, George G.**, "Secured Debt Under Conditions of Imperfect Information," *Journal of Legal Studies*, 1992, 21 (1), 225–258.
- Tullock, Gordon**, *The Rent-Seeking Society*, Indianapolis: Liberty Fund, 2005. edited by Charles K. Rowley.
- Ulen, Thomas S.**, "The Efficiency of Specific Performance: Toward a Unified Theory of Contract Remedies," *Michigan Law Review*, November 1984, 83 (2), 341–403.
- Varian, Hal R.**, *Microeconomic Analysis*, 3rd ed., New York: W.W. Norton, 1992.
- Walras, Leon**, *Eléments d'économie politique pure*, Lausanne: L. Corbaz, 1874.
- Webb, David C.**, "Two-Period Financial Contracts with Private Information and Costly State Verification," *Quarterly Journal of Economics*, August 1992, 107 (3), 1113–1123.
- Weinstein, Mark**, "Profit Sharing Contracts in Hollywood: Evolution and Analysis," *Journal of Legal Studies*, 1998, 27 (1), 67–112.
- White, James J. and Robert S. Summers**, *Uniform Commercial Code*, 5th ed., St. Paul: West Group, 2000.
- White, Michelle J.**, "Contract Breach and Contract Discharge due to Impossibility: A Unified Theory," *Journal of Legal Studies*, 1988, 17 (2), 353–376.
- Williamson, Oliver E.**, *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*, New York: The Free Press, 1975.
- Willig, Robert D.**, "Consumer Surplus Without Apology," *American Economic Review*, September 1976, 66 (4), 589–597.
- Wittgenstein, Ludwig**, *Tractatus Logico-Philosophicus*, London: Routledge and Kegan Paul, 1922.
- Wittman, Donald**, "Optimal Pricing of Sequential Inputs: Last Clear Chance, Mitigation of Damages, and Related Doctrines in the Law," *Journal of Legal Studies*, January 1981, 10 (1), 65–91.
- Wolcher, Louis E.**, "Price Discrimination and Inefficient Risk Allocation Under the Rule of *Hadley v. Baxendale*," *Research in Law and Economics*, 1989, 9, 9–31.