1 Introduction

Economics is often defined as a field that aims to understand the process by which scarce resources are allocated to their most efficient uses, and markets are generally seen as playing a central role in this process. But, more fundamentally, the simple activity of exchange of goods and services, whether on organized exchanges or outside a market setting, is the basic first step in any production or allocation of resources. For a long time economic theory has been able to analyze formally only very basic exchange activities like the barter of two different commodities between two individuals at a given place and point in time. Most microeconomics textbooks[1] begin with an analysis of this basic situation, representing it in the classic "Edgeworth box." A slightly more involved exchange situation that can also be represented in an Edgeworth box is between two individuals trading at different points in time. Simple lending, investment, or futures contracts can be characterized in this way. However, such a simple reinterpretation already raises new issues, like the possibility of default or nondelivery by the other party in the future.

Until the 1940s or 1950s only situations of simple exchange of goods and services were amenable to formal analysis. More complex exchange activities like the allocation and sharing of risk began to be analyzed formally only with the introduction of the idea of "state-contingent" commodities by Arrow (1964) and Debreu (1959) and the formulation of a theory of "choice under uncertainty" by von Neumann and Morgenstern (1944) and others. The notion of exchange of state-contingent commodities gave a precise meaning to the exchange and allocation of risk. Preference orderings over lotteries provided a formal representation of attitudes toward risk and preferences for risk taking. These conceptual innovations are the foundations of modern theories of investment under risk and portfolio choice.

In the late 1960s and 1970s yet another conceptual breakthrough took place with the introduction of "private information" and "hidden actions" in contractual settings. The notions of "incentive compatibility" and incentives for "truth telling" provided the basic underpinnings for the theory of incentives and the economics of information. They also provided the first formal tools for a theory of the firm, corporate finance, and, more generally, a theory of economic institutions.

1. See, for example, Part 4 of the celebrated book by Mas-Colell, Whinston, and Green (1995).

Finally, much of the existing theory of long-term or dynamic contracting was developed in the 1980s and 1990s. Contract renegotiation, relational contracts, and incomplete contracts provided the first tools for an analysis of "ownership" and "control rights." These notions, in turn, complete the foundations for a full-fledged theory of the firm and organizations.

There are by now many excellent finance and economics textbooks covering the theory of investment under risk, insurance, and risk diversification. As this is already well-explored territory, we shall not provide any systematic coverage of these ideas. In contrast, to date there are only a few books covering the theory of incentives, information, and economic institutions, which is generally referred to in short as *contract theory*.[2] There has been such a large research output on these topics in the last 30 years that it is an impossible task to give a comprehensive synthesis of all the ideas and methods of contract theory in a single book. Nevertheless, our aim is to be as wide ranging as possible to give a sense of the richness of the theory—its core ideas and methodology—as well as its numerous possible applications in virtually all fields of economics.

Thus, in this book we attempt to cover all the major topics in contract theory that are taught in most graduate courses. Part I starts with basic ideas in incentive and information theory like screening, signaling, and moral hazard. Part II covers the less well trodden material of multilateral contracting with private information or hidden actions. In this part we provide an introduction to auction theory, bilateral trade under private information, and the theory of internal organization of firms. Part III deals with long-term contracts with private information or hidden actions. Finally, Part IV covers incomplete contracts, the theory of ownership and control, and contracting with externalities. Exercises are collected in a specific chapter at the end of the book.

There is obviously too much material in this book for any one-semester course in contract theory. Rather than impose our own preferences and our own pet topics, we thought that it would be better to cover all the main themes of contract theory and let instructors pick and choose which parts to cover in depth and which ones to leave to the students to read.

Consistent with our goal of providing broad coverage of the field, we have aimed for a style of exposition that favors simplicity over generality or rigor.

2. See in particular the textbooks by Salanié (1997) and Laffont and Martimort (2002).

Our primary goal is to illustrate the core ideas, the main methods in their simplest self-contained form, and the wide applicability of the central notions of contract theory. More often than not, research articles in contract theory are hard to penetrate even for a well-trained reader. We have gone to considerable lengths to make the central ideas and methods in these articles accessible. Inevitably, we have been led to sacrifice generality to achieve greater ease of understanding. Our hope is that once the main ideas have been assimilated the interested reader will find it easier to read the original articles.

In the remainder of this chapter we provide a brief overview of the main ideas and topics that are covered in the book by considering a single concrete situation involving an *employer* and an *employee.* Depending on the topic we are interested in we shall take the employer to be a *manager* hiring a *worker,* or a *farmer* hiring a *sharecropper,* or even a company *owner* hiring a *manager.* Throughout the book we discuss many other applications, and this brief overview should not be taken to be the leading application of contract theory. Before we proceed with a brief description of the multiple facets of this contracting problem, it is useful to begin by delineating the boundaries of the framework and stating the main assumptions that apply throughout this book.

The benchmark contracting situation that we shall consider in this book is one between two parties who operate in a market economy with a well-functioning legal system. Under such a system, any contract the parties decide to write will be enforced perfectly by a court, provided, of course, that it does not contravene any existing laws. We shall assume throughout most of the book that the contracting parties do not need to worry about whether the courts are able or willing to enforce the terms of the contract precisely. Judges are perfectly rational individuals, whose only concern is to stick as closely as possible to the agreed terms of the contract. The penalties for breaching the contract will be assumed to be sufficiently severe that no contracting party will ever consider the possibility of not honoring the contract. We shall step outside this framework only occasionally to consider, for example, the case of self-enforcing contracts.

Thus, throughout this book we shall assume away most of the problems legal scholars, lawyers, and judges are concerned with in practice and concentrate only on the economic aspects of the contract. We shall be primarily interested in determining what contractual clauses rational economic

individuals are willing to sign and what types of transactions they are willing to undertake.

If the transaction is a simple exchange of goods or services for money, we shall be interested in the terms of the transaction. What is the price per unit the parties shall agree on? Does the contract specify rebates? Are there penalty clauses for late delivery? If so, what form do they take? And so on. Alternatively, if the transaction is an insurance contract, we shall be interested in determining how the terms vary with the underlying risk, with the risk aversion of the parties, or with the private information the insuree or the insurer might have about the exact nature of the risk. We begin by briefly reviewing the simplest possible contractual situation an employer and employee might face: a situation involving only two parties, transacting only once, and facing no uncertainty and no private information or hidden actions.

1.1 Optimal Employment Contracts without Uncertainty, Hidden Information, or Hidden Actions

Consider the following standard bilateral contracting problem between an employer and employee: the employee has an initial endowment of time, which she can keep for herself of sell to the employer as labor services, because the employer can make productive use of the employee's time. Specifically, we can assume therefore that the parties' utility functions depend both on the allocation of employee time and on their purchasing power. Let us denote the employer's utility function as $U(l, t)$ where l is the quantity of employee time the employer has acquired and t denotes the quantity of "money"—or equivalently the "output" that this money can buy[3]—that he has at his disposal. Similarly, employee utility is $u(l, t)$, where l is the quantity of time the employee has kept for herself and t is the quantity of money that she has at her disposal.

Suppose that the initial endowment of the individuals is $(\hat{l}_1, \hat{t}_1) = (0, 1)$ for the employer (hereafter *individual 1*) and $(\hat{l}_2, \hat{t}_2) = (1, 0)$ for the employee (hereafter *individual 2*). That is, without any trade, the employer gets no employee time but is assumed to have all the money, while the employee has all of her time for herself but has no money.

3. Indeed, the utility of money here reflects the utility derived from the consumption of a composite good that can be purchased with money.

Both individuals could decide not to trade, in which case they would each achieve a utility level of $\overline{U} = U(0, 1)$ and $\bar{u} = u(1, 0)$, respectively. If, however, both utility functions are strictly increasing in both arguments and strictly concave, then both individuals may be able to increase their joint payoff by exchanging labor services l for money/output. What will be the outcome of their contractual negotiations? That is, how many hours of work will the employee be willing to offer and what (hourly) wage will she be paid?

As in most economics texts, we shall assume throughout this book that contracting parties are rational individuals who aim to achieve the highest possible payoff. The joint surplus maximization problem for both individuals can be represented as follows. If we denote by l_i the amount of employee time actually consumed and by t_i the amount of output consumed by each party $i = 1, 2$ after trade, then the parties will solve the following optimization problem:

$$\max_{l_i, t_i} U(l_1, t_1) + \mu u(l_2, t_2) \tag{1.1}$$

subject to aggregate resource constraints:

$$l_1 + l_2 = \hat{l}_1 + \hat{l}_2 = 1 \quad \text{and} \quad t_1 + t_2 = \hat{t}_1 + \hat{t}_2 = 1$$

Here μ can reflect both the individuals' respective reservation utility levels, $\overline{U}$ and $\bar{u}$, and their relative bargaining strengths.

When both utility functions are strictly increasing and concave, the maximum is completely characterized by the first-order conditions

$$U_l + \mu u_l = 0 = U_t + \mu u_t \tag{1.2}$$

which imply

$$\frac{U_l}{U_t} = \frac{u_l}{u_t}$$

See Figure 1.1, where indifference curves are drawn.

In other words, joint surplus maximization is achieved when the marginal rates of substitution between money and leisure for both individuals are equalized.

There are gains from trade initially if

$$\frac{U_l}{U_t} > \frac{u_l}{u_t}$$

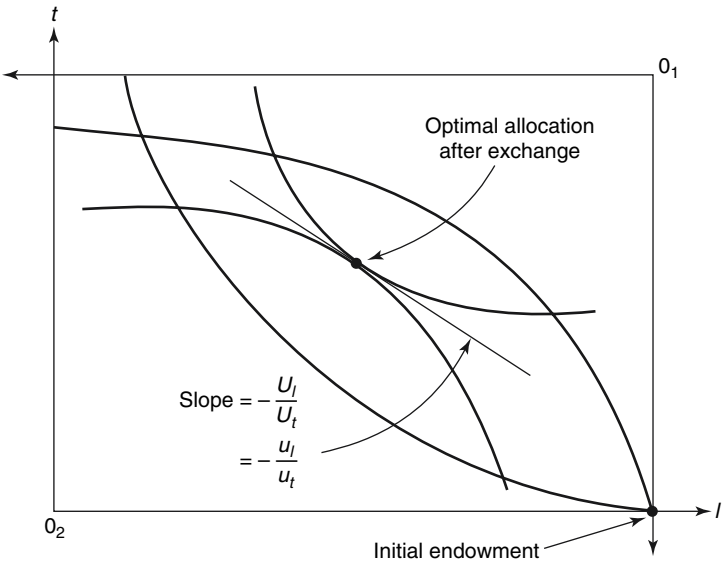

Figure 1.1
Classical Edgeworth Box

How these gains are shared between the two individuals is determined by μ. The employee gets a higher share of the surplus the higher μ is. The highest possible utility that the employee can get is given by the solution to the following optimization problem:

$$\max_{l_2, t_2} u(l_2, t_2) \quad \text{subject to} \quad U(1-l_2, 1-t_2) \geq \overline{U}$$

Similarly, the highest payoff the employer can get is given by the solution to

$$\max_{l_1, t_1} U(l_1, t_1) \quad \text{subject to} \quad u(1-l_1, 1-t_1) \geq \bar{u}$$

These extreme problems can be interpreted as simple bargaining games where one party has all the bargaining power and makes a take-it-or-leave-it offer to the other party. Note, however, that by increasing $\bar{u}$ in the employer's constrained maximization problem or $\overline{U}$ in the employee's problem one can reduce the surplus that either individual gets. Thus a given division of the surplus can be parameterized by either μ, $\overline{U}$, or $\bar{u}$, depending on how the joint surplus maximization problem is formulated.

Throughout this book we shall represent optimal contracting outcomes as solutions to constrained optimization problems like the two preceding problems. We thus take as starting point the Coase theorem (1960), that is, the efficient contracting perspective, as long as informational problems are not present.[4] Although this representation seems quite natural, it is important to highlight that behind it lie two implicit simplifying assumptions. First, the final contract the parties end up signing is independent of the bargaining process leading up to the signature of the contract. In reality it is likely that most contracts that we see partly reflect prior negotiations and each party's negotiating skills. But, if the main determinants of contracts are the parties' objectives, technological constraints, and outside options, then it is not unreasonable to abstract from the potentially complex bargaining games they might be playing. At least as a first approach, this simplifying assumption appears to be reasonable.

Second, as we have already mentioned, the other relevant dimension of the contracting problem that is generally suppressed in the preceding formal characterization is the enforcement of the contract. Without legal institutions to enforce contracts many gains from trade are left unexploited by rational individuals because one or both fear that the other will fail to carry out the agreed transaction. In the absence of courts or other modes of enforcement, a transaction between two or more parties can take place only if the exchange of goods or services is simultaneous. Otherwise, the party who is supposed to execute her trade last will simply walk away. In practice, achieving perfect simultaneity is almost impossible, so that important gains from trade may remain unexploited in the absence of an efficient enforcement mechanism.

1.2 Optimal Contracts under Uncertainty

There is more to employment contracts than the simple characterization in the previous section. One important dimension in reality is the extent to which employees are insured against economic downturns. In most developed economies employees are at least partially protected against the risk of unemployment. Most existing unemployment insurance schemes are

4. As we shall detail throughout this book, informational problems will act as constraints on the set of allocations that contracts can achieve.

nationwide insurance arrangements, funded by employer and employee contributions, and guaranteeing a minimum fraction of a laid-off employee's pay over a minimum time horizon (ranging from one year to several years with a sliding scale). A fundamental economic question concerning these insurance schemes is how much "business-cycle" and other "firm-specific" risk should be absorbed by employers and how much by employees. Should employers take on all the risk, and if so, why? One theory, dating back to Knight (1921) and formalized more recently by Kihlstrom and Laffont (1979) and Kanbur (1979), holds that employers (or "entrepreneurs") should take on all the risk and fully insure employees. The reason is that entrepreneurs are natural "risk lovers" and are best able to absorb the risk that "risk-averse" employees do not want to take.

To be able to analyze this question of optimal risk allocation formally one must enrich the framework of section 1.1 by introducing uncertainty. At one level this extension is extremely simple. All it takes is the introduction of the notions of a state of nature, a state space, and a state-contingent commodity. Arrow (1964) and Debreu (1959) were the first to explore this extension. They define a state of nature as any possible future event that might affect an individual's utility. The state space is then simply the set of all possible future events, and a state-contingent commodity is a good that is redefined to be a new commodity in every different state of nature. For example, a given number of hours of work is a different commodity in the middle of an economic boom than in a recession.

The difficulty is not in defining all these notions. The important conceptual leap is rather to suppose that rational individuals are able to form a complete description of all possible future events and, moreover, that all have the same description of the state space. Once this common description is determined, the basic contracting problem can be represented like the preceding one, although the interpretation of the contract will be different. More precisely, it is possible to represent a simple insurance contract, which specifies trades between the employer and employee in different states of nature, in an Edgeworth box. Before doing so, let us consider a pure insurance problem without production.

1.2.1 Pure Insurance

Consider the simplest possible setting with uncertainty. Assume that there are only two possible future states of nature, θ_L and θ_H. To be concrete, let θ_L represent an adverse output shock, or a "recession," and θ_H a good output

realization, or a "boom." For simplicity, we disregard time endowments. Then the state of nature influences only the value of output each individual has as endowment. Specifically, assume the following respective endowments for each individual in each state:

$(\hat{t}_{1H}, \hat{t}_{1L}) = (2, 1)$, for individual 1

$(\hat{t}_{2H}, \hat{t}_{2L}) = (2, 1)$, for individual 2

The variable $\hat{t}_{ij}$ therefore denotes the endowment of individual i in state of nature θ_j. Note that in a "recession" aggregate output—2—is lower than in a boom—4.

Before the state of nature is realized each individual has preferences over consumption bundles (t_L, t_H) represented by the utility functions $V(t_L, t_H)$ for the employer and $v(t_L, t_H)$ for the employee.

If the two individuals do not exchange any contingent commodities, their ex ante utility (before the state of nature is realized) is $\overline{V} = V(2, 1)$ and $\bar{v} = v(2, 1)$. But they can also increase their ex ante utility by coinsuring against the economic risk. Note, however, that some aggregate risk is uninsurable: the two individuals can do nothing to smooth the difference in aggregate endowments between the two states. Nevertheless, they can increase their ex ante utility by pooling their risks.

As before, the efficient amount of coinsurance is obtained when the final allocations of each contingent commodity $\{(t_{1L}, t_{2L}), (t_{1H}, t_{2H})\}$ are such that

$$V_{t_L} + \mu v_{t_L} = 0 = V_{t_H} + \mu v_{t_H} \tag{1.3}$$

which implies

$$\frac{V_{t_L}}{V_{t_H}} = \frac{v_{t_L}}{v_{t_H}}$$

See Figure 1.2, where indifference curves are drawn.

It should be clear by now that the analysis of pure exchange under certainty can be transposed entirely to the case with uncertainty once one enlarges the commodity space to include contingent commodities.

However, to obtain a full characterization of the optimal contracting problem under uncertainty one needs to put more structure on this framework. Indeed, two important elements are hidden in the preceding characterization of the optimal insurance contract: one is a description of

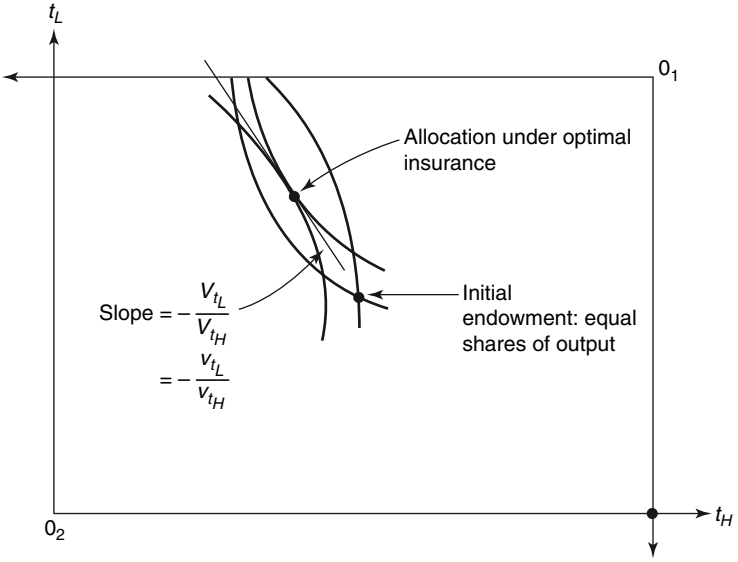

Figure 1.2
Optimal Coinsurance

ex post utility once the state of nature has been realized, and the other is the probability of each state occurring.

The first complete framework of decision making under uncertainty, which explicitly specifies the probability distribution over states and the ex post utility in each state, is due to von Neumann and Morgenstern (1944). It is this framework that is used in most contracting applications. Interestingly, even though there is by now a large literature exploring a wide range of alternative models of individual choice and behavior under uncertainty, there have been relatively few explorations of the implications for optimal contracting of alternative models of behavior under uncertainty.

In the setup considered by von Neumann and Morgenstern, individual ex post utility functions are respectively $U(t)$ and $u(t)$ for the employer and employee, where both functions are increasing in t. If we call $p_j \in (0, 1)$ the probability of occurrence of any particular state of nature θ_j, the ex ante utility function is simply defined as the expectation over ex post utility outcomes:

$$V(t_{1L}, t_{1H}) = p_L U(t_{1L}) + p_H U(t_{1H})$$

and

$$v(t_{2L}, t_{2H}) = p_L u(t_{2L}) + p_H u(t_{2H})$$

The easiest way of thinking about the probability distribution $\{p_j\}$ is simply as an objective distribution that is known by both individuals. But it is also possible to think of $\{p_j\}$ as a subjective belief that is common to both individuals. In most contracting applications it is assumed that all parties share a common prior belief and that differences in (posterior) probability beliefs among the parties only reflect differences in information. Although this basic assumption is rarely motivated, it generally reflects the somewhat vague idea that all individuals are born with the same "view of the world" and that their beliefs differ only if they have had different life experiences. Recently, however, there have been some attempts to explore the implications for optimal contracting of fundamental differences in beliefs among contracting parties.

It is instructive to consider the optimal insurance conditions (1.3) when the individuals' ex ante utility function is assumed to be the Von Neumann–Morgenstern utility function that we have specified. In that case the marginal rate of substitution between commodities 1 and 2 is given by

$$\frac{V_{t_L}}{V_{t_H}} = \frac{p_L}{p_H} \frac{U'(t_{1L})}{U'(t_{1H})}$$

As this expression makes clear, the marginal rate of substitution between the two contingent commodities varies with the probability distribution. Moreover, the marginal rate of substitution is constant along the 45° line, where $t_{1L} = t_{1H}$.

1.2.2 Optimal Employment Contracts under Uncertainty

Using the framework of von Neumann and Morgenstern, let us come back to the contracting problem of section 1.1 with two goods, leisure l and a consumption good t, which can be readily extended to include uncertainty as follows:

Let (l_{1L}, t_{1L}) and (l_{1H}, t_{1H}) represent the two different state-contingent time/output bundles of the employer, and (l_{2_L}, t_{2_L}) and (l_{2H}, t_{2H}) the two different state-contingent time/output bundles of the employee. Also let $(\hat{l}_{ij}, \hat{t}_{ij})$ denote their respective initial endowments, $(i = 1, 2; j = L, H)$. Then

the optimal insurance contract signed by the two individuals can be represented as the solution to the optimal contracting problem:

$$\max_{l_{ij}, t_{ij}} [p_L U(l_{1L}, t_{1L}) + p_H U(l_{1H}, t_{1H})]$$

subject to

$$p_L u(l_{2L}, t_{2L}) + p_H u(l_{2H}, t_{2H}) \geq \bar{u} \tag{1.4}$$

and

$$l_{1j} + l_{2j} \leq \hat{l}_{1j} + \hat{l}_{2j} \quad \text{for } j = L, H$$

$$t_{1j} + t_{2j} \leq \hat{t}_{1j} + \hat{t}_{2j} \quad \text{for } j = L, H$$

where

$$\bar{u} = p_L u(\hat{l}_{2L}, \hat{t}_{2L}) + p_H u(\hat{l}_{2H}, \hat{t}_{2H})$$

One important advantage of the von Neumann and Morgenstern formulation is that an individual's attitude toward risk can be easily characterized by the curvature of the ex post utility function. Thus, if both $U(\cdot)$ and $u(\cdot)$ are strictly concave, then both individuals are risk averse and want to share risk, whereas if both $U(\cdot)$ and $u(\cdot)$ are strictly convex, then both individuals are risk loving and want to trade gambles with each other.

For now, suppose that both individuals are risk averse, so that their ex post utility functions are strictly concave. Then the contract-maximizing joint surplus is fully characterized by the first-order conditions:

$$\frac{U_l(l_{1j}, t_{1j})}{U_t(l_{1j}, t_{1j})} = \frac{u_l(l_{2j}, t_{2j})}{u_t(l_{2j}, t_{2j})} \tag{1.5}$$

$$\frac{U_l(l_{1j}, t_{1j})}{u_l(l_{2j}, t_{2j})} \quad \text{constant across } \theta_j\text{'s} \tag{1.6}$$

$$\frac{U_t(l_{1j}, t_{1j})}{u_t(l_{2j}, t_{2j})} \quad \text{constant across } \theta_j\text{'s} \tag{1.7}$$

Condition (1.5) is the familiar condition for efficient trade ex post. This means that ex ante efficiency is achieved if and only if the contract is also ex post efficient. We shall see that when incentive considerations enter into the contracting problem there is usually a conflict between ex ante and ex post efficiency.

Conditions (1.6) and (1.7) are conditions of optimal coinsurance. Condition (1.7) is sometimes referred to as the *Borch rule* (1962): optimal coinsurance requires the equalization of the ratio of marginal utilities of money across states of nature.

A risk-neutral individual has a constant marginal utility of money. Thus, if one of the two individuals is risk neutral and the other individual is risk averse, the Borch rule says that optimal insurance requires that the risk-averse individual must also have a constant marginal utility of money across states of nature. In other words, the risk-averse individual must get perfect insurance. This is exactly the solution that intuition would suggest.

To summarize, optimal contracting under uncertainty would result in perfect insurance of the employee against economic risk only if the employer is risk neutral. In general, however, when both employer and employee are risk averse, they will optimally share business risk. Thus the simple Knightian idea that entrepreneurs perfectly insure employees is likely to hold only under special assumptions about risk preferences of entrepreneurs. An individual's attitude toward risk is driven in part by initial wealth holdings. Thus it is generally accepted that individuals' absolute risk aversion tends to decrease with wealth. If extremely wealthy individuals are approximately risk neutral and poor individuals are risk averse, then one special case where the Knightian theory would be a good approximation is when wealth inequalities are extreme and a few very wealthy entrepreneurs provide nearly perfect job security to a mass of poor employees.

It should be clear from this brief overview of optimal contracting under uncertainty that the presumption of rational behavior and perfect enforceability of contracts is less plausible in environments with uncertainty than in situations without uncertainty. In many contracting situations in practice it is possible that the contracting parties will be unable to agree on a complete description of the state space and that, as a consequence, insurance contracts will be incomplete. The rationality requirements imposed on the contracting parties and the enforcement abilities assumed of the courts should be kept in mind as caveats for the theory of contracting when faced with very complex actual contractual situations where the parties may have limited abilities to describe possible future events and the courts have limited knowledge to be able to effectively stick to the original intentions of the contracting parties.

Another important simplifying assumption to bear in mind is that it is presumed that each party knows exactly the intentions of the other

contracting parties. However, in practice the motives behind an individual's willingness to contract are not always known perfectly. As a consequence, suspicion about ulterior motives often may lead to breakdown of contracting. These considerations are the subject of much of this book and are briefly reviewed in the next sections of this chapter.

1.3 Information and Incentives

The preceding discussion highlights that even in the best possible contracting environments, where comprehensive insurance contracts can be written, it is unlikely that employees will be perfectly insured against business risks. The reason is simply that the equilibrium price of such insurance would be too high if employers were also averse to risk.

Another important reason employees are likely to get only limited insurance is that they need to have adequate incentives to work. If the output produced by employees tends to be higher when employees exert themselves more, or if the likelihood of a negative output shock is lower if employees are more dedicated or focused on their work, then economic efficiency requires that they receive a higher compensation when their (output) performance is better. Indeed, if their pay is independent of performance and if their job security is not affected by their performance, why should they put any effort into their work? This is a well-understood idea. Even in the egalitarian economic system of the former Soviet Union the provision of incentives was a generally recognized economic issue, and over the years many ingenious schemes were proposed and implemented to address the problem of worker incentives and also factory managers' incentives. What was less well understood, however, was the trade-off between incentives and insurance. How far should employee insurance be scaled back to make way for adequate work incentives? How could adequate work incentives be structured while preserving job security as much as possible? These remained open and hotly debated questions over the successive five-year plans.

Much of Part I of this book will be devoted to a formal analysis of this question. Two general types of incentive problems have been distinguished. One is the *hidden-information* problem and the other the *hidden-action* problem. The first problem refers to a situation where the employee may have private information about her inability or unwillingness to take on

certain tasks. That is, the information about some relevant characteristics of the employee (her distaste for certain tasks, her level of competence) are hidden from her employer. The second problem refers to situations where the employer cannot see what the employee does—whether she works or not, how hard she works, how careful she is, and so on. In these situations it is the employee's actions that are hidden from the employer.

Problems of hidden information are often referred to as *adverse selection,* and problems of hidden actions as *moral hazard.* In practice, of course, most incentive problems combine elements of both moral hazard and adverse selection. Also, the theoretical distinction between a hidden-action and a hidden-information problem can sometimes be artificial. Nevertheless, it is useful to distinguish between these two types of incentive problems, in part because the methodology that has been developed to analyze these problems is quite different in each case.

1.3.1 Adverse Selection

Chapters 2 and 3 provide a first introduction to optimal contracts with hidden information. These chapters examine optimal bilateral contracts when one of the contracting parties has private information. Chapter 2 explores contracting situations where the party making the contract offers is the *uninformed* party. These situations are often referred to as *screening* problems, since the uninformed party must attempt to screen the different pieces of information the informed party has. Chapter 3 considers the opposite situation where the *informed* party makes the contract offers. These situations fall under the general descriptive heading of *signaling* problems, as the party making the offer may attempt to signal to the other party what it knows through the type of contract it offers or other actions.

The introduction of hidden information is a substantial break from the contracting problems we have already considered. Now the underlying contracting situation requires specification of the private information one of the parties might have and the beliefs of the other party concerning that information *in addition to* preferences, outside options, initial endowments, a state space, and a probability distribution over states of nature.

In the context of employment contracts the type of information that may often be private to the employee at the time of contracting is her basic skill, productivity, or training. In practice, employers try to overcome this informational asymmetry by hiring only employees with some training or only high school and college graduates.

In a pathbreaking analysis Spence (1973, 1974) has shown how education can be a signal of intrinsic skill or productivity. The basic idea behind his analysis is that more-able employees have a lower disutility of education and therefore are more willing to educate themselves than less-able employees. Prospective employers understand this and therefore are willing to pay educated workers more even if education per se does not add any value. We review Spence's model and other contracting settings with signaling in Chapter 3.

Another way for employers to improve their pool of applicants is to commit to pay greater than market-clearing wages. This tends to attract better applicants, who generally have better job opportunities and are more likely to do well in interviews. Such a policy naturally gives rise to equilibrium unemployment, as Weiss (1980) has shown. Thus, as Akerlof, in his 1970 article, and Stiglitz, in many subsequent writings, had anticipated, the presence of private information about employee characteristics can potentially explain at a microeconomic level why equilibrium unemployment and other forms of market inefficiencies can arise. With the introduction of asymmetric information in contracting problems economists have at last found plausible explanations for observed market inefficiencies that had long eluded them in simpler settings of contracting under complete information.

Or at least they thought so. Understandably, given the importance of the basic economic issue, much of the subsequent research on contracting under asymmetric information has tested the *robustness* of the basic predictions of market inefficiencies and somewhat deflated early expectations about a general theory of market inefficiencies.

A first fundamental question to be tackled was, Just how efficient can contracting under asymmetric information be? The answer to this question turns out to be surprisingly elegant and powerful. It is generally referred to as the *revelation principle* and is one of the main notions in contract economics. The basic insight behind the revelation principle is that to determine optimal contracts under asymmetric information it suffices to consider only one contract for each type of information that the informed party might have, but to make sure that each type has an incentive to select only the contract that is destined to him/her.

More concretely, consider an employer who contracts with two possible types of employees—a "skilled" employee and an "unskilled" one—and who does not know which is which. The revelation principle says that it is

optimal for the employer to consider offering only two employment contracts—one destined to the skilled employee and the other to the unskilled one—but to make sure that each contract is *incentive compatible*. That is, that each type of employee wants to pick only the contract that is destined to her. Thus, according to the revelation principle, the employer's optimal contracting problem reduces to a standard contracting problem, but with additional incentive compatibility constraints.

As a way of illustrating a typical contracting problem with hidden information, let us simplify the previous problem with uncertainty with a simple form of private information added. Suppose first that employee time and output enter additively in both utility functions: define them as $U[\alpha\theta(1 - l) - t]$ for the employer and $u(\theta l + t)$ for the employee, where

- $(1 - l)$ is the employee time sold to the employer, and l is the time the employee keeps for herself;
- t is the monetary/output transfer from the employer to the employee;
- α is a positive constant; and
- θ measures the "unit value of time," or the skill level of the employee.

The variable θ is thus the state of nature, and we assume it is learned *privately* by the employee before signing any contract. Specifically, the employee knows whether she is skilled, with a value of time θ_H, or unskilled, with a value of time $\theta_L < \theta_H$. The employer, however, knows only that the probability of facing a skilled employee is p_H.

When the employer faces a skilled employee, the relevant reservation utility is $\bar{u}_H = u(\theta_H)$, and when he faces an unskilled employee, it is $\bar{u}_L = u(\theta_L)$.[5] Assume that the employee's time is more efficient when sold to the employer; that is, assume $\alpha > 1$. Then, if the employer could also learn the employee's type, he would simply offer in state θ_j a contract with a transfer $t_j = \theta_j$ in exchange for all her work time (that is, $1 - l_j = 1$). Such a contract would maximize production efficiency, and since the employee's *individual rationality constraint*, $u(t_j) \geq u(\theta_j)$, would be binding under this contract, it would maximize the employer's payoff.

When employee productivity is private information, however, the employer would not be able to achieve the same payoff, for if the employer offers a wage contract $t_j = \theta_j$ in exchange for 1 unit of work time, all

5. Indeed, in state θ_j, the employee's endowment $\hat{l}_{2j} = \theta_j$.

employee types would respond by "pretending to be skilled" to get the higher wage θ_H.

Note that for the employee type to be truly private information it must also be the case that the employee's output is not observable. If it were, the employer could easily get around the informational asymmetry by including a "money-back guarantee" into the contract should the employee's output fall short of the promised amount. Assumptions similar to the nonobservability of output are required in these contracting problems with hidden information. A slightly more realistic assumption serving the same purpose is that the employee's output may be random and that a "no-slavery" constraint prevents the employer from punishing the employee ex post for failing to reach a given output target. If that is the case, then an inefficient employee can always pretend that she was "unlucky." Even if this latter assumption is more appealing, we shall simply assume here for expositional convenience (as is often done) that output is unobservable.

Under that assumption, the only contracts that the employer can offer the employee are contracts offering a total payment of $t(l)$ in exchange for $(1 - l)$ units of work. Although this class of contracts is much simpler than most real-world employment contracts, finding the optimal contracts in the set of all (nonlinear) functions $\{t(l)\}$ could be a daunting problem. Fortunately, the revelation principle offers a key simplification. It says that all the employer needs to determine is a menu of two "point contracts": (t_L, l_L) and (t_H, l_H), where, by convention, (t_j, l_j) is the contract chosen by type j. The reason why the employer does not need to specify a full (nonlinear) contract $t(l)$ is that each type of employee would pick only one point on the full schedule $t(l)$ anyway. So the employer might as well pick that point directly. However, each point has to be incentive compatible. That is, type θ_H must prefer contract (t_H, l_H) over (t_L, l_L), and type θ_L contract (t_L, l_L) over (t_H, l_H).

Thus the optimal *menu* of employment contracts under hidden information can be represented as the solution to the optimal contracting problem under complete information:

$$\max_{(l_j, t_j)} \{p_L U[\alpha\theta_L(1-l_L)-t_L]+p_H U[\alpha\theta_H(1-l_H)-t_H]\}$$

subject to

$$u(l_L\theta_L + t_L) \geq u(\theta_L)$$

and

$$u(l_H\theta_H + t_H) \geq u(\theta_H)$$

but with two additional *incentive constraints:*

$$u(l_H\theta_H + t_H) \geq u(l_L\theta_H + t_L)$$

and

$$u(l_L\theta_L + t_L) \geq u(l_H\theta_L + t_H)$$

The solution to this constrained optimization problem will produce the most efficient contracts under hidden information. As this problem immediately reveals, the addition of incentive constraints will in general result in less efficient allocations than under complete information. In general, optimal contracts under hidden information will be second-best contracts, which do not achieve simultaneously optimal allocative and distributive efficiency. Much of Chapter 2 will be devoted to the analysis of the structure of incentive constraints and the type of distortions that result from the presence of hidden information. The general economic principle that this chapter highlights is that hidden information results in a form of *informational monopoly power* and allocative inefficiencies similar to those produced by monopolies. In the preceding example, the employer might choose to suboptimally employ skilled employees (by setting $1 - l_H < 1$) to be able to pay unskilled employees slightly less.[6] In a nutshell, the main trade-off that is emphasized in contracting problems with hidden information is one between informational rent extraction and allocative efficiency.

6. One option is to have a contract with allocative efficiency, that is, $l_H = l_L = 0$ and $t_H = t_L = \theta_H$. This leaves informational rents for the unskilled employee relative to her outside opportunity. It may therefore be attractive to lower skilled employment (that is, set $l_H > 0$)—at an allocative cost of $(\alpha - 1)l_H$—in order to lower t_L without violating the incentive constraint:

$$u(l_L\theta_L + t_L) \geq u(l_H\theta_L + t_H)$$

Intuitively, lowering skilled employment allows the employer to lower t_H by a significant amount, since the skilled employee has a high opportunity cost of time. Therefore, "pretending to be skilled" becomes less attractive for the unskilled employee (who has a lower opportunity cost of time), with the result that a lower transfer t_L becomes compatible with the incentive constraint. The trade-off between allocative efficiency and rent extraction will be detailed in the next chapter.

If the presence of hidden information may give rise to allocative inefficiencies such as unemployment, it does not follow that public intervention is warranted to improve market outcomes. Indeed, the incentive constraints faced by employers are also likely to be faced by planners or public authorities. It is worth recalling here that the centrally planned economy of the Soviet Union was notorious for its overmanning problems. It may not have had any official unemployment, but it certainly had huge problems of underemployment. Chapters 7 and 13 will discuss at length the extent to which market outcomes under hidden information may be first- or second-best efficient and when the "market mechanism" may be dominated by some better institutional arrangement.

Our discussion has focused on a situation where the employee has an informational advantage over the employer. But, in practice, it is often the employer that has more information about the value of the employee's work. Chapter 2 also explores several settings where employers have private information about demand and the value of output. As we highlight, these settings are perhaps more likely to give rise to unemployment. Indeed, layoffs can be seen as a way for employers to credibly convey to their employees that the economic environment of their firm has worsened to the extent that pay cuts may be needed for the firm to survive.

1.3.2 Moral Hazard

Chapter 4 introduces and discusses the other major class of contracting problems under asymmetric information: *hidden actions.* In contrast to most hidden information problems, contracting situations with hidden actions involve informational asymmetries arising after the signing of a contract. In these problems the agent (employee) is not asked to choose from a menu of contracts, but rather from a menu of action-reward pairs.

Contracting problems with hidden actions involve a fundamental incentive problem that has long been referred to in the insurance industry as *moral hazard:* when an insuree gets financial or other coverage against a bad event from an insurer she is likely to be less careful in trying to avoid the bad outcome against which she is insured. This behavioral response to better insurance arises in almost all insurance situations, whether in life, health, fire, flood, theft, or automobile insurance. When a person gets better protection against a bad outcome, she will rationally invest fewer resources in trying to avoid it. One of the first and most striking empirical studies of

moral hazard is that of Peltzman (1975), who has documented how the introduction of laws compelling drivers to wear seat belts has resulted in higher average driving speeds and a greater incidence of accidents (involving, in particular, pedestrians).

How do insurers deal with moral hazard? By charging proportionally more for greater coverage, thus inducing the insuree to trade off the benefits of better insurance against the incentive cost of a greater incidence of bad outcomes.

Incentive problems like moral hazard are also prevalent in employment relations. As is now widely understood, if an employee's pay and job tenure are shielded against the risk of bad earnings, then she will work less in trying to avoid these outcomes. Moral hazard on the job was one of the first important new economic issues that Soviet planners had to contend with. If they were to abolish unemployment and implement equal treatment of workers, how could they also ensure that workers would work diligently? As they reluctantly found out, there was unfortunately no miracle solution. For a time ideological fervor and emulation of model workers seemed to work, but soon major and widespread motivation problems arose in an economic system founded on the separation of pay from performance.

Employers typically respond to moral hazard on the job by rewarding good performance (through bonus payments, piece rates, efficiency wages, stock options, and the like) and/or punishing bad performance (through layoffs). As with insurance companies, employers must trade off the benefits of better insurance (in terms of lower average pay) against the costs in lower effort provision by employees. The most spectacular form of incentive pay seen nowadays is the compensation of CEOs in the United States. Arguably, the basic theory of contracting with hidden actions discussed in Chapter 4 provides the main theoretical underpinnings for the types of executive compensation packages seen today. According to the theory, even risk-averse CEOs should receive significant profit- and stock-performance-based compensation if their (hidden) actions have a major impact on the firm's performance.[7]

While it is easy to grasp at an intuitive level that there is a basic trade-off between insurance and incentives in most employment relations, it is

7. Actual CEO compensation packages have also shown some serious limitations, which the theory has addressed too; more on this topic will follow.

less easy to see how the contract should be structured to best trade off effort provision and insurance.

Formally, to introduce hidden actions into the preceding employment problem with uncertainty, suppose that the amount of time $(1 - l)$ worked by the employee is private information (a *hidden action*). Suppose, in addition, that the employee chooses the action $(1 - l)$ *before* the state of nature θ_j is realized and that this action influences the *probability* of the state of nature: when the employee chooses action $(1 - l)$, output for the employer is simply θ_H with a probability function $p_H[1 - l]$, increasing in $1 - l$ (and θ_L with a probability function $p_L[1 - l] = 1 - p_H[1 - l]$).[8] The usual interpretation here is that $(1 - l)$ stands for "effort," and more effort produces higher expected output, at cost $1 - l$ for the employee, say.[9] However, it is not *guaranteed* to bring about higher output, since the bad state of nature θ_L may still occur. Note that if output were to increase deterministically with effort then the unobservability of effort would not matter because the agent's hidden effort supply could be perfectly inferred from the observation of output.

Since effort $(1 - l)$ is not observable, the agent can be compensated only on the basis of realized output θ_j. The employer is thus restricted to offering a compensation contract $t(\theta_j)$ to the employee. Also the employer must now take into account the fact that $(1 - l)$ will be chosen by the employee to maximize her own expected payoff under the output-contingent compensation scheme $t(\theta_j)$. In other words, the employer can now make only a best guess that the effort level chosen by the employee is the outcome of the employee's own optimization problem:

$$(1-l) \in \underset{\tilde{l}}{\operatorname{argmax}}\{p_L[1-l]u[t(\theta_L)+l] + p_H[1-l]u[t(\theta_H)+l]\}$$

Therefore, when the employer chooses the optimal compensation contract $\{t(\theta_j)\}$ to maximize his expected utility, he must make sure that it is in the employee's best interest to supply the right level of effort $(1 - l)$. In other words, the employer now solves the following maximization problem:

$$\max_{t(\theta_j)}\{p_L[1-l]U[\theta_L - t(\theta_L)] + p_H[1-l]U[\theta_H - t(\theta_H)]\}$$

8. Where $p_H[\cdot]$ (respectively $p_L[\cdot]$) is an increasing (respectively, decreasing) *function* of $(1 - l)$.

9. For simplicity, we assume here that the opportunity cost of time for the employee is independent of the state of nature.

subject to

$$p_L[1-l]u[t(\theta_L)+l]+p_H[1-l]u[t(\theta_H)+l]\geq \overline{u}=u(1) \qquad \text{(IR)}$$

and

$$(1-l)\in \underset{\tilde{l}}{\operatorname{argmax}}\{p_L[1-\tilde{l}]u[t(\theta_L)+\tilde{l}]+p_H[1-\tilde{l}]u[t(\theta_H)+\tilde{l}]\} \qquad \text{(IC)}$$

As in contracting problems with hidden information, when the action supplied by the employee is not observable the employer must take into consideration not only the employee's individual rationality constraint but also her incentive constraint.

Determining the solution to the employer problem with both constraints is not a trivial matter in general. Chapter 4 provides an extensive discussion of the two main approaches toward characterizing the solution to this problem. For now we shall simply point to the main underlying idea that an efficient *trade-off* between insurance and incentives involves rewarding the employee most for output outcomes that are most likely to arise when she puts in the required level of effort and punishing her the most for outcomes that are most likely to occur when she shirks. The application of this principle can give rise to quite complex compensation contracts in general, often more complex than what we see in reality. There is one situation, however, where the solution to this problem is extremely simple: when the employee is risk neutral. In that case it is efficient to have the employee take on all the output risk so as to maximize her incentives for effort provision. That is, when the employee is risk neutral, she should fully insure the employer.

One reason why this simple theory may predict unrealistically complex incentive schemes is that in most situations with hidden actions the incentive problem may be multifaceted. CEOs, for example, can take actions that increase profits, but they can also manipulate earnings, or "run down" assets in an effort to boost current earnings at the expense of future profits. They can also undertake high-expected-return but high-risk investments. It has been suggested that when shareholders or any other employer thus face a multidimensional incentive problem, then it may be appropriate to respond with both less "high powered" and simpler incentive schemes. We explore these ideas both in Chapter 6, which discusses hidden action problems with multiple tasks, and in Chapter 10, which considers long-term incentive contracting problems where the employee takes repeated hidden actions.

Chapter 6 also considers multidimensional hidden information problems as well as problems combining both hidden information and hidden actions. All these problems raise new analytical issues of their own and produce interesting new insights. We provide an extensive treatment of some of the most important contracting problems under multidimensional asymmetric information in the research literature.

Part I of our book also discusses contracting situations with an intermediate form of asymmetric information, situations where the informed party can *credibly disclose* her information if she wishes to do so. These situations, which are considered in Chapter 5, are mostly relevant for accounting regulation and for the design of mandatory disclosure rules, which are quite pervasive in the financial industry. Besides their obvious practical relevance, these contractual situations are also of interest because they deal with a very simple incentive problem, whether to *disclose or hide* relevant information (while *forging* information is not an available option). Because of this simplicity, the contractual problems considered in Chapter 5 offer an easy introduction to the general topic of contracting under asymmetric information. One of the main ideas emerging from the analysis of contracting problems with private but verifiable information is that incentives for voluntary disclosure can be very powerful. The basic logic, which is sometimes referred to as the "unraveling result," is that any seller of a good or service (e.g., an employee) has every incentive to reveal good information about herself, such as high test scores or a strong curriculum vitae. Employers understand this fact and expect employees to be forthcoming. If an employee is not, the employer assumes the worst. It is for this reason that employees have incentives to voluntarily disclose all but the worst piece of private verifiable information. This logic is so powerful that it is difficult to see why there should be mandatory disclosure laws. Chapter 5 discusses the main limits of the unraveling result and explains when mandatory disclosure laws might be warranted.

Finally, it is worth stressing that although our leading example in this introduction is the employment relation, each chapter contains several other classic applications, whether in corporate finance, industrial organization, regulation, public finance, or the theory of the firm. Besides helping the readers to acquaint themselves with the core concepts of the theory, these applications are also meant to highlight the richness and broad relevance of the basic theory of contracting under private information.

1.4 Optimal Contracting with Multilateral Asymmetric Information

The contracting situations we have discussed so far involve only one-sided private information or one-sided hidden actions. In practice, however, there are many situations where several contracting parties may possess relevant private information or be called to take hidden actions. A first basic question of interest then is whether and how the theory of contracting with one-sided private information extends to multilateral settings. Part II of this book is devoted to this question. It comprises two chapters. Chapter 7 deals with *multilateral private information* and Chapter 8 with *multilateral hidden actions*. Besides the obvious technical and methodological interest in analyzing these more general contractual settings, fundamental economic issues relating to the constrained efficiency of contractual outcomes, the role of competition, and the theory of the firm are also dealt with in these chapters.

While the general methodology and most of the core ideas discussed in Part I extend to the general case of multilateral asymmetric information, there is one fundamental difference. In the one-sided private information case the contract design problem reduces to a problem of controlling the informed party's response, while in the multilateral situation the contracting problem becomes one of controlling the strategic behavior of several parties interacting with each other. That is, the contract design problem becomes one of *designing a game with incomplete information*.

One of the main new difficulties then is predicting how the game will be played. The best way of dealing with this issue is in fact to design the contract in such a way that each player has a unique *dominant strategy*. Then the outcome of the game is easy to predict, since in essence all strategic interactions have then been removed. Unfortunately, however, contracts where each party has a unique dominant strategy are generally not efficient. Indeed, in a major result which builds on Arrow's (1963) impossibility theorem, Gibbard (1973) and Satterthwaite (1975) have shown that it is impossible in general to attain the full-information efficient outcome when there are more than two possible allocations to choose from and when the contracting parties' domain of preferences is unrestricted (that is, when the set of possible types of each contracting party is very diverse). Rather than stick to predictable but inefficient contracts, it may then generally be desirable to agree on contracts where the outcome is less predictable but on

average more efficient (that is, contracts where each party's response depends on what the other contracting parties are expected to do). From a theorist's perspective this is a mixed blessing because the proposed efficient contracts (or "mechanisms," as they are often referred to in multilateral settings) may be somewhat fragile and may not always work in practice as the theory predicts.

1.4.1 Auctions and Trade under Multilateral Private Information

Perhaps the most important and widely studied problem of contracting with multilateral hidden information is the design of *auctions* with multiple bidders, each with his or her own private information about the value of the objects that are put up for auction.[10] Accordingly, Chapter 7 devotes considerable space to a discussion of the main ideas and derivation of key results in auction theory, such as the *revenue equivalence theorem* or *the winner's curse*. The first result establishes that a number of standard auctions yield the same expected revenue to the seller when bidders are risk neutral and their valuations for the object are independently and identically distributed. The second idea refers to the inevitable disappointment of the winner in an auction where bidders value the object in a similar way but have different prior information about its worth: when she learns that she won she also finds out that her information led her to be overoptimistic about the value of the object.

In recent years there has been an explosion of research in auction theory partly because of its relevance to auction design in a number of important practical cases. Covering this research would require a separate book, and Chapter 7 can serve only as an introduction to the subject.

Auction design with multiple informed bidders is by no means the only example of contracting with multilateral hidden information. Another leading example, which is extensively discussed in Chapter 7, is trade in situations where each party has private information about how much it values the good or the exchange.

A major economic principle emerging from the analysis of contracting with one-sided hidden information is the trade-off between allocative efficiency and extraction of informational rents. If the bargaining power lies

10. Despite its relative fragility, the theory of contracting with multilateral hidden information has proved to be of considerable practical relevance, as for example in the design of spectrum and wireless telephone license auctions (see for example Klemperer, 2002).

with the uninformed party, as we have assumed, then that party attempts to appropriate some of the informational rents of the informed party at the expense of allocative efficiency. But note that if the informed party (e.g., the employee in our example) has all the bargaining power and makes the contract offer, then the contracting outcome is always efficient. So, if the overriding objective is to achieve a Pareto efficient outcome (with, say, no unemployment), then there appears to be a simple solution when there is only one-sided hidden information: simply give all the bargaining power to the informed party.

In practice, however, besides the difficulty in identifying who the informed party is, there is also the obvious problem that generally all parties to the contract will have some relevant private information. Therefore, the natural contracting setting in which to pose the question of the efficiency of trade under asymmetric information and how it varies with the bargaining power of the different parties is one of multilateral asymmetric information. A fundamental insight highlighted in Chapter 7 is that the main constraint on efficient trade is not so much eliciting the parties' private information as ensuring their participation. Efficient trade can (almost) always be achieved if the parties' participation is obtained before they learn their information, while it cannot be achieved if participation is decided when they already know their type.

Applying this insight to our labor contracting example, the analysis in Chapter 7 indicates that labor market inefficiencies like unemployment are to be expected in an otherwise frictionless labor market when employers have market power and employees private information about their productivity, or when there is two-sided asymmetric information. It must be stressed, however, that policy intervention that is not based on any information superior to that available to the contracting parties will not be able to reduce or eliminate these inefficiencies. But labor market policies that try to intensify competitive bidding for jobs or for employees should lower inefficiencies caused by hidden information.

1.4.2 Moral Hazard in Teams, Tournaments, and Organizations

Contracting situations where several parties take hidden actions are often encountered in firms and other organizations. It is for this reason that the leading application of contracting problems involving multisided moral hazard is often seen to be the internal organization of firms and other economic institutions. Some prominent economic theorists of the firm like

Alchian and Demsetz (1972) or Jensen and Meckling (1976) go as far as arguing that the resolution of *moral-hazard-in-teams* problems (where several agents take complementary hidden actions) is the raison d'être of a firm. They contend that the role of a firm's owner or manager is to monitor employees and make sure that they take efficient actions that are hidden to others. Hence, the analysis of contracting problems with multisided moral hazard is important if only as an indirect vehicle for understanding economic organizations and firms.

Accordingly, Chapter 8 covers multiagent moral hazard situations with a particular focus on firms and their internal organization. To illustrate some of the key insights and findings covered in this chapter, consider the situation where our employer now contracts with two employees, A and B, each supplying a (hidden) "effort" $(1 - l_A)$ and $(1 - l_B)$. A key distinction in contracting problems with multisided moral hazard concerns the measure of performance: Is each employee's performance measured separately, or is there a single aggregate measure of both employees' contributions?

In the former case, when the output of each employee is observable and is given by, say, θ_{Aj} with probability p_{Aj} and θ_{Bj} with probability p_{Bj}, for $j = L, H$, the employer's problem is similar to the single-agent moral hazard problem described earlier, with the new feature that now the employer can also base compensation on each employee's *relative performance:*

$$\theta_{A_j} - \theta_{B_j}$$

An important class of incentive contracting situations in which agents are rewarded on the basis of how well they did relative to others is *rank-order tournaments*. Many sports contests are of this form, and promotions of employees up the corporate ladder can also be seen as a particular form of tournament.

Thus, in our employment problem with two employees and observable individual outputs, the employer may be able to provide better incentives with less risk exposure to the two employees by basing compensation on how well they perform relative to each other. This possibility can be seen as one reason why firms like to provide incentives to their employees through promotion schemes, appointing only the better employees to higher paying and more rewarding jobs. The reason why relative performance evaluation improves incentives is that when employees are exposed to the same exogenous shocks affecting their performance (changes in demand for their output or quality of input supplies, say), it is possible to

shield them against these risks by *filtering out the common shock* from their performance measure. To see how this works, think that the probability p_{Aj} (resp., p_{Bj}) depends not only on individual effort $(1 - l_A)$ [resp., $(1 - l_B)$] but also on a random variable *that affects both agents*. In this case, it makes sense to link an employee's compensation positively to her own performance but *negatively* to the other employee's performance. Chapter 8 discusses extensively how to make the best use of relative performance measures in general problems with multisided moral hazard.

It is worth noting here that as compelling and plausible as the case for relative performance may be, many critical commentators on CEO compensation in the United States have pointed to the absence of such relative performance evaluation for CEOs. For example, Bebchuk, Fried, and Walker (2002) have criticized CEO compensation contracts in the United States for not optimally correcting compensation by filtering out common stock market shocks through indexing. They argue that this is a major deviation from optimal incentive contracting and is evidence of a failure in corporate governance in most large U.S. companies. Others, however, have rationalized the absence of explicit indexing as an optimal way of getting managers to do their own hedging when this is cheaper, or as an optimal response to competitive pressures in the market for CEOs (see Garvey and Milbourn, 2003; Jin, 2002).

Let us now turn to the second case, where observable output is a *single aggregate measure* given by

$$\theta_{A_j} + \theta_{B_j}$$

with the probability of higher realizations that depends positively on each employee's effort. Then the employer faces a moral-hazard-in-teams problem. Indeed, the amount of time worked by either of the two employees is a public good because, by raising joint output, it benefits both employees. As is easy to understand, in such situations a major difficulty for the employer is to prevent *free riding* by one employee on the other employee's work.

Alchian and Demsetz (1972) proposed that free riding of employees can be prevented through monitoring by the employer. That is, the employer's main role in their view is one of supervising employees and making sure that they all work. They also argue that the employer should be the residual claimant on the firm's revenues and that employees should be paid fixed wages to make sure that the employer has the right incentives to monitor.

When monitoring is too costly or imperfect, however, then employees also need to be motivated through compensation based on aggregate performance. An important insight of Holmström (1982), which we discuss in Chapter 8, is that optimal provision of incentives by giving shares of aggregate output to employees requires *budget breaking* in general. That is, the sum of the shares of the team members should not always add up to one. The residual should then be sold to a third party, which can be thought of as outside shareholders.

When the number of employees to be monitored is large, it is not reasonable to think that a single employer is able to effectively monitor all employees. Multiple supervisors are then required, and someone will have to monitor the monitors. If the number of supervisors is itself large, then multiple monitors of supervisors will be needed. And so on. Thus, by specifying the *span of control* of any supervisor (the number of employees that can reasonably be monitored by a single supervisor) and the *loss of control* as more tiers of supervisors are added (intuitively, there will be an overall reduction in efficiency of supervision of bottom-layer employees as more layers are added between the top and the bottom of the *hierarchy*), one can develop simultaneously a simple theory of the optimal firm size and the optimal hierarchical internal organization of the firm. Again, Chapter 8 gives an extensive treatment of this theory of organizations.

One of the reasons why there may be a loss of control as more supervisory tiers are added is that midlevel supervisors may attempt to *collude* with their employees against top management or the firm's owners. Recent corporate scandals in the United States have painfully reminded investors of the risk of collusion between auditors and the agents they are meant to monitor. These examples vividly draw attention to the importance of considering the possibility of collusion in multiagent contracting situations. Chapter 8 provides an extensive discussion of some of the main models of optimal contracting with collusion. It emphasizes in particular the idea that beyond incentive and participation constraints, optimal multilateral contracts are also constrained by "no-collusion constraints."

1.5 The Dynamics of Incentive Contracting

In practice, many if not most contracting relations are repeated or long term. Yet the theory we develop in the first two parts of the book deals only

with static or one-shot contracting situations. In Part III we provide systematic coverage of long-term incentive contracting, mostly in a bilateral contracting framework. In Chapter 9 we discuss dynamic adverse selection and in Chapter 10 dynamic moral hazard.

Methodologically, there is no significant change in analyzing optimal multiperiod contracts as long as the contracting parties can *commit* to a single comprehensive long-term contract at the initial negotiation stage. As we have already noted in the context of intertemporal coinsurance contracting problems, when full commitment is feasible the long-term contract can essentially be reduced to a slightly more complex static contract involving trade of a slightly richer basket of state-contingent commodities, services, and transfers. What this conclusion implies in particular for contracting under hidden information is that the revelation principle still applies under full commitment.

However, if the contracting parties are allowed to *renegotiate* the initial contract as time unfolds and new information arrives, then new conceptual issues need to be addressed and the basic methodology of optimal static contracting must be adapted. Mainly, incentive constraints must then be replaced by tighter *renegotiation-proofness* constraints.

A number of new fundamental economic issues arise when the parties are involved in a long-term contractual relation. How is private information revealed over time? How is the constrained efficiency of contractual outcomes affected by repeated interactions? How does the possibility of renegotiation limit the efficiency of the overall long-term contract? To what extent can *reputation* serve as a more informal enforcement vehicle that is an alternative to courts? We discuss these and other issues extensively in this third part of the book.

1.5.1 Dynamic Adverse Selection

There are two canonical long-term contracting problems with hidden information: one where the informed party's type does not change over time and the other where a new type is drawn every period. In the first problem the main new conceptual issue to be addressed relates to *learning* and the gradual reduction of the informed party's informational advantage over time. The second class of problems is conceptually much closer to a static contracting problem, as the information asymmetry between the two contracting parties remains stationary. The main novel economic question in this class of problems concerns the trade-off

between within-period and intertemporal insurance or allocative efficiency.

To see one important implication of learning of the informed party's type over time, consider again our employment-contracting problem with private information of the employee's productivity. That is, suppose that the employee can supply labor $(1 - l)$ to produce output $\alpha\theta_H(1 - l)$ when she is skilled and output $\alpha\theta_L(1 - l)$ when she is unskilled at opportunity cost $(1 - l)\theta_j$ $(j = H,L)$, where her productivity and opportunity cost of labor are hidden information. And suppose, as before, that the employer knows only the probability of facing a skilled employee, p_H. As we have seen, in a static contracting situation the employer would pursue an optimal trade-off between rent extraction and allocative inefficiency with a menu of employment contracts. The contract for the skilled employee would specify an inefficiently low level of employment, and the contract for the unskilled employee would leave her an informational rent relative to her outside opportunity.

Now, consider a twice-repeated relation with spot contracting in each period. In this situation the menu of contracts that we have described would no longer be feasible in the first period: indeed, if in the first period the type-j employee chooses option (l_j, t_j), she will have identified herself to the employer. In the second period the employer would then know her outside opportunity, and would in particular not leave any informational rent anymore to the unskilled employee. Therefore, unless the employer commits not to respond in this way, the unskilled employee will be reluctant to reveal her type by separating. Consequently more pooling of types is to be expected in early stages of the contracting relation.

Note that this commitment issue is a very general one that arises in many different contexts. It was known for example to analysts of the Soviet system as the *ratchet effect* (see Weitzman, 1976), which denotes the behavior of central planners that dynamically increase firm performance targets when they realize that they are facing very productive firms.

Under full commitment to a comprehensive long-term contract the preceding problems disappear. But full commitment will not be feasible if the contracting parties are allowed to sign long-term contracts but cannot commit not to renegotiate them in the future if they identify Pareto-improving opportunities. Indeed, in our employment example, the contracting parties will always want to renegotiate the optimal long-term contract, as this contract specifies an inefficiently low labor supply for the

skilled employee. Once the high skill of the employee is revealed, there are gains from trade to renegotiating the contract to a higher level of labor supply. But if this renegotiation is anticipated, then the unskilled employee will again want to pretend to be skilled. In general, then, the optimal *renegotiation-proof* contract will differentiate the types less in the early stages of the relation, and the hidden information about the employee's type will only gradually be revealed to the employer. Chapter 9 provides an extensive discussion of the dynamics of contracting under adverse selection. It also illustrates the relevance of these ideas with several applications.

One general lesson emerging from our analysis in this chapter is that there are no gains from enduring relationships when the type of the informed party is fixed. Indeed, the best the contracting parties can hope to achieve is to repeat the optimal static contract. In contrast, when the informed party's type changes over time, there are substantial gains from repeating the relationship. While it is stretching our imagination to think that an employee's intrinsic productivity may change randomly over time, it is much more plausible to think of the hidden type as an unobservable income shock and to think of the contracting problem as an insurance problem with unobservable income shocks. Indeed, the first formal model of this problem by Townsend (1982) considers exactly this application.

The starting point of this analysis is that there can be no gains from contracting at all in a one-shot contracting relation because the informed party will always claim to have had a low income realization in order to receive an insurance compensation. But even in a twice-repeated contracting relation there can be substantial gains from insurance contracting. The reason is that in a relation that is repeated twice or more, greater insurance against income shocks within the first period can be traded off against better intertemporal allocation of consumption. In very concrete terms an individual who gets a low-income shock in the first period can borrow against her future income to smooth consumption. Vice versa, an individual who gets a high income in the first period can save some of this income toward future consumption. The key insight of Townsend and the subsequent literature on this problem is that borrowing and lending in a competitive debt market provides inefficiently low insurance. The optimal long-term incentive-compatible contract would provide more within-period insurance against low-income shocks. As we highlight in Chapter 9, this insight is particularly relevant for understanding the role of banks and their

greater ability in providing liquidity (that is, within-period insurance) than financial markets.

1.5.2 Dynamic Moral Hazard

Dynamic contracting problems with moral hazard have a similar structure to dynamic adverse selection problems where the type of the informed party is drawn randomly every period. As with these contracting problems, there are gains from enduring relations, and the optimal long-term contract induces a similar distortion in intertemporal consumption allocations relative to what would obtain under repeated-spot-incentive contracting and simple borrowing and lending. That is, an optimal long-term employment contract with hidden actions by the employee will induce her to consume relatively more in earlier periods. If the employee were free to save any amount of her first-period income at competitive market rates, then she would choose to save more than is optimal under a long-term incentive contract, which would directly control her savings. The broad intuition for this general result is that by inducing the employee to consume more in early periods the employer can keep her "hungry" in subsequent periods and thus does not need to raise her level of compensation to maintain the same incentives.

Chapter 10 begins with a thorough analysis of a general twice-repeated contracting problem with moral hazard and of the general result we just mentioned. It then proceeds with a detailed discussion of the different effects in play in a repeated relation with moral hazard and identifies two important sources of gains and one important source of losses from an enduring relation. A first positive effect is that repetition of the relation makes the employee less averse to risk, since she can engage in "self-insurance" and offset a bad output shock in one period by borrowing against future income. A second potential positive effect comes from better information about the employee's choice of action obtained from repeated output observations. Offsetting these two positive effects, however, is a negative effect, which comes from the greater flexibility afforded the employee to act in response to dynamic incentives. In an enduring relation she can slack off following a good performance run or make up for poor performance in one period by working extra hard the next period.

It is this ability to modulate her effort supply in response to good or bad output changes that drives a striking insight due to Holmström and Milgrom (1987) concerning the shape of the optimal long-term incen-

tive contract. One would think that when the relation between an employer and employee is enduring, the complexity of the employment contract would grow with the length of the relation, so much so that the optimal contract predicted by incentive theory in any realistic setup would become hopelessly complex. One might then fear that this extreme complexity could easily defeat the practical use of the theory. Holmström and Milgrom, however, argue that as the employee's set of possible actions grows with the length of the relation, the set of incentive constraints that constrict the shape of the optimal contract becomes so large that the incentive-compatible long-term contract ends up taking a simple linear form in final accumulated output. In short, under an enduring relation the optimal long-term contract gains in simplicity. This observation, which tends to accord well with the relative simplicity of actual employment contracts, is, however, theoretically valid only under some specific conditions on preferences and technology.

Another important simplification that is available under fairly general conditions is that the incentive effects under an optimal long-term incentive contract may be replicable with a sequence of spot contracts. This observation may be of particular relevance for evaluating long-term CEO compensation contracts. A common practice is to let CEOs exercise their stock options and sell their equity stake early but to "reload" their stock options to provide continuing incentives to CEOs. This has been viewed as an inefficient practice by some commentators (e.g., Bebchuk, Fried, and Walker, 2002), but it may also be seen as consistent with the idea of replication of the efficient long-term contract through a sequence of short-term contracts.

Explicit long-term employment contracts may also take a simple form in practice because in an ongoing employment relation efficiency may be attained by providing a combination of *explicit and implicit incentives*. The explicitly written part of the contract may then appear to be simple because it is supplemented by sophisticated implicit incentives. Loosely speaking, the term "implicit incentives" refers to notions like reputation building, career concerns, informal rewards, and *quid pro quos*. In reality many long-term employment relations do not provide a complete specification of employer obligations and employee duties. Instead they are sustained by implicit rules and incentives. Reliance on such incomplete explicit contracts may often be a way of economizing on contract-drafting costs. Accordingly, Chapter 10 provides an extensive treatment of so-called *relational contracts,*

which combine both explicit and implicit incentives, and of the implicit incentives derived from "market" perceptions about employee "talent" and their implications, for example, in terms of outside offers.

The chapter also deals with renegotiation. As with dynamic adverse selection, the possibility of renegotiation undermines efficient incentive provision. Once a risk-averse employee has taken her action, there is no point in further exposing her to unnecessary output risk, and gains from renegotiation open up by letting the employer provide better insurance to the employee. But if such renegotiation is anticipated, then the employee will have lower incentives to put in effort. This issue is particularly relevant for CEO compensation where the "action" to be taken by the CEO may be the implementation of a new investment project or a new business plan. Once the project has been undertaken, there is no point in exposing the CEO to further risk, and it may be efficient to let her sell at least part of her equity stake. This is indeed the prediction of optimal incentive contracting with renegotiation, as we explain in Chapter 10.

1.6 Incomplete Contracts

Our discussion of explicit and implicit incentives already alludes to the fact that most long-term contracts in practice are *incomplete,* in that they do not deal explicitly with all possible contingencies and leave many decisions and transactions to be determined later. It is easy to understand intuitively why this is the case. Most people find it hard to think through even relatively simple dynamic decision problems and prefer to leave many decisions to be settled at a later stage when they become more pressing. If this is true for dynamic decision problems, then this must be the case a fortiori for dynamic contracting problems, where the parties must *jointly* think through and agree on future transactions or leave them to be determined at a later stage.

Our formulation of optimal contracting problems in the first three parts of the book abstracts from all these issues. There are no contract-drafting costs, there are no limits on contract enforcement, and parties are able to instantly determine complex optimal long-term contracts. This is clearly a drastic albeit convenient simplification. In the fourth and final part of the book we depart from this simple framework and explore the implications of contractual incompleteness.

As in Part III, Part IV is concerned with long-term contracts. But the focus is different. When contracts are incomplete, some transactions and decisions must be determined by the contracting parties at some later stage. The question then arises, Who makes these decisions? The principal focus of this part of the book will be to address this question. The form of the (incomplete) long-term contract will be prespecified exogenously in at least some dimensions, and the optimizing variables will be mainly the allocation among contracting parties of ownership titles, control rights, discretion, authority, decision-making rules, and so on.

Hence, the formulation of the basic incomplete contracting problem involves a major methodological change. Indeed, to emphasize this change we consider mostly problems involving little or no asymmetric information at the contracting stage. This part of the book also involves a fundamental substantive change: In the first three parts the focus was exclusively on monetary rewards for the provision of incentives. In contrast, in Part IV the focus will be on the incentive effects of *control* and *ownership protections*. In other words, this part emphasizes other institutional factors besides monetary remuneration in the provision of incentives. In a nutshell, the incomplete contracting approach offers a vehicle to explore the analysis of economic institutions and organizations systematically.

1.6.1 Ownership and Employment

In Chapter 11 we begin our treatment of incomplete contracts by assuming that an inability to describe certain events accurately before the fact is the principal reason why contracts are incomplete. We shall, however, assume that after the fact these events are easily described and their implications fully understood. It has been a matter of debate how much limitations on language are a constraint for drafting fully comprehensive contracts both in theory and in practice. We provide an extensive discussion of this debate in Chapter 12.

In Chapter 11 we specify exogenously which events the contract cannot be based on and focus on the implications of contractual incompleteness for institution design. We shall be interested primarily in the role of two ubiquitous institutions of market economies, ownership rights and employment relations.

The first formal model of an incomplete contracting problem by Simon (1951) deals with a fundamental aspect of the employment relation we have not hitherto considered: the *authority relation* between the employer and

employee. So far we have described an employment contract like any other contract for the provision of an explicit service or "output." But in reality most employment contracts define only in very broad terms the duties of the employee and leave to the discretion of the employer the future determination of the specific tasks of the employee. In short, employment contracts are highly incomplete contracts where the employee agrees to put herself under the (limited) authority of the employer. Employment contracts thus specify a different mode of transaction from the negotiation mode prevalent in spot markets. It is for this reason that Simon, Coase, Williamson, and others have singled out the employment relation as the archetypal form of an economic institution that is different from market exchange.

Simon views the choice between the two modes of transaction as a comparison between two long-term contracts: a "sales contract," in which the service to be provided is precisely specified in a contract, and an "employment contract," in which the service is left to the discretion of the buyer (employer) within some contractually specified limits. The employment contract is preferred when the buyer is highly uncertain at the time of contracting about which service he prefers and when the seller (employee) is close to indifferent between the different tasks the employer can choose from. Chapter 11 discusses the strengths and limitations of Simon's theory and provides an extensive treatment of a "modernized" version of his theory that allows for *ex ante relation-specific investments* and ex post renegotiation. Chapter 12 further builds on Simon's theory by explicitly modeling "orders" or "commands" given by the employer, to which the employee responds by either "quitting" or "executing" the order.

The notion that the presence of relation-specific investments creates the need for modes of exchange other than trade in spot markets has been articulated and emphasized forcibly in Williamson's writings (1975, 1979, 1985). In a pathbreaking article that builds on his insights, Grossman and Hart (1986) developed a simple theory and model of *ownership rights* based on the notion of *residual rights of control*. They define a firm as a collection of assets owned by a common owner, who has residual rights of control over the use of the assets. A key new notion in their article is that ownership serves as a protection against future *holdups* by other trading partners and thus may give stronger incentives for ex ante relation-specific investments. That is, the owner of an asset has a bargaining chip in future negotiations over trades not specified in the initial incomplete contracts. He can sell

access to the productive asset to future trading partners who need the asset for production. The owner can thus protect the returns from ex ante relation-specific investments.

Building on these notions, Grossman and Hart are able to provide a simple formal theory of the costs and benefits of integration and the boundaries of the firm. Chapter 11 provides an extensive treatment of this theory. The advantage of integration is that the bargaining position of the owner of the newly integrated firm is strengthened. This stronger position may induce him to invest more. The drawback, however, is that the previous owner's bargaining position is weakened. This agent may therefore invest less. Depending on the relative size of these costs and benefits of integration, Grossman and Hart are able to determine when it is optimal to integrate or not. They are thus able to articulate for the first time a simple and rigorous theory of the boundaries of the firm, which has been further elaborated by Hart and Moore (1990) and synthesized by Hart (1995).

1.6.2 Incomplete Contracts and Implementation Theory

While this theory of the firm has improved our understanding of the special role of ownership, a major theoretical issue remains only partially resolved, at least in its initial versions. As Maskin and Tirole (1999a) have pointed out, there is a basic logical tension in the theory. On the one hand, contracting parties are assumed to be able to fully anticipate the consequences of their current actions for the future, like the potential for holdups or the protections given by ownership. And yet, on the other hand, they are also assumed to be unable to limit expected future abuse by trading partners with explicit contractual clauses. All they can do to improve their future negotiating position is to trade a very standardized contract: ownership titles.

We discuss the delicate theoretical issues relating to this basic logical tension in Chapter 12. This chapter begins by covering the theory of *Nash implementation* (Maskin, 1977) and its subsequent developments. This theory deals with issues of contract design in situations where an event is difficult to describe ex ante (or identify by a third party) but easily recognized by all contracting parties ex post. It exploits the idea that contracts can be made contingent on such events by relying on reports by the parties on which event occurred. A striking general result of implementation theory is that by designing suitable *revelation games* for the contracting parties it is often possible to achieve the same outcomes as with fully contingent contracts.

Thus the challenge for the theory of incomplete contracts, which relies on the distinction between observability and nonverifiability (or nondescribability) of an event, is to explain why the contracting parties do not attempt to contract around this constraint by designing sophisticated revelation games (or *Maskin schemes* as they are commonly called). This is not necessarily an abstruse theoretical issue, as Chapter 12 illustrates with examples of plausible optionlike contracts that achieve efficiency in contracting problems involving relation-specific investments and holdup problems. As is readily seen, these contracts can be interpreted as simple revelation games.

One way the challenge has been taken up is to argue that Maskin schemes have limited power in improving efficiency over simple incomplete contracts in *complex* contracting environments where there may be many different states of nature or potentially many different services or goods to be traded (see Segal, 1999a). Chapter 12 provides an extensive discussion of these arguments. It also explores another foundation for the theory of authority, based on actions that are *both ex ante and ex post noncontractable*. If one assumes that one can contract on who *controls* these actions, one can derive predictions about the optimal allocation of authority within organizations, either in a one-shot or in a repeated context.

1.6.3 Bilateral Contracts and Multilateral Exchange

Finally, Chapter 13 deals with another common form of contractual incompleteness: the limited participation of all concerned parties in a single comprehensive multilateral contract. Employment contracts, for example, are generally bilateral contracts between an employer and an employee even in situations where the employee works together in a team with other employees, or in situations where the employer is involved in a whole *nexus of contracts* with suppliers, clients, lenders, and other providers of capital. When (incomplete) bilateral contracts are written in such multilateral contract settings, any bilateral contract may impose an externality on the other parties. The equilibrium outcome of the contracting game may then be inefficient. Thus a central focus of this chapter is the characterization of situations where bilateral contracting results in efficient outcomes.

An important distinction that is drawn in the literature is whether the bilateral contract is *exclusive* or *nonexclusive*—that is, whether the

employee can sign only an exclusive contract with one employer or whether she can sign up with several employers for several part-time jobs. Interestingly, most employment contracts are exclusive. But this is not always the case for other contracts. For example, for health insurance it is generally possible for the insuree to acquire supplementary insurance. Similarly, with credit card debt or other loans, a borrower can build up debt on several different cards or take out loans from several different lenders. Exactly why exclusivity is required for some types of contracts but not others involving externalities has not been fully explored. Intuitively, one should expect to see exclusive contracts when the externality is potentially large and when exclusivity is easy to enforce. Whether exclusivity is enforced or not, however, one should expect inefficient equilibrium outcomes to obtain in general in bilateral contracting games, since bilateral contracts alone are insufficient to fully internalize all externalities across all affected parties. This is a central theme in the *common agency* literature, which studies multiple bilateral incentive contracting between a single agent and several principals (see, for example, Bernheim and Whinston, 1985, 1986a, 1986b). Chapter 13 provides an extensive treatment of this important contracting problem.

Because of the presence of a potential externality, an obvious concern is whether the bilateral contracting game has a well-defined equilibrium outcome. An early focus of the contracting literature has indeed been the potential nonexistence of equilibrium in such contracting games. In a landmark article Rothschild and Stiglitz (1976) have thus shown that if two insurers compete for exclusive bilateral contracts with one or several insurees who have private information about their likelihood of facing an adverse shock (or accident), then a well-defined equilibrium outcome of the contracting game may not exist. The reason is that each insurer has an incentive to respond to the contract offers of the other insurer by only *cream skimming* the good risks and leaving the high-risk insurees to contract with their rival. Chapter 13 discusses this striking result and the vast literature it has spawned.

Other important topics are touched on besides these two broad themes, such as the strategic value of contracting in duopoly or barrier-to-entry settings, or the impact of product-market competition on the size of agency problems. But it is fair to say that Chapter 13 does not attempt to provide a systematic treatment of the existing literature on bilateral contracting

with competition (whether static or dynamic, with adverse selection, moral hazard, or both) simply because the existing literature that touches on this topic is at this point both too vast and too disconnected to be able to provide a systematic and comprehensive treatment in only one chapter.[11]

1.7 Summing Up

The analysis of optimal contracting in this book highlights the fact that common contract forms and institutions that we take for granted, like employment contracts and ownership rights, are sophisticated and multifaceted "institutions." For a long time economists have been able to give only an oversimplified analysis of these contract forms, which ignored uncertainty, asymmetric information, incentives, and control issues. As this introductory chapter makes clear, however, a basic economic relation like the employment relation has to deal with these various facets.

The goal of this book is therefore to explain how existing contract theory allows one to incorporate these features, not just in employment relations but also in many other applications. In fact, we have chosen to illustrate contract-theoretic analyses in many different contexts, typically choosing the application that has been the most influential in economics for the particular general problem under consideration.

The book gives an overview of the main conceptual breakthroughs of contract theory, while also pointing out its current limitations. As contract theory has grown to become a large field, we have been forced to limit our coverage by making difficult choices on what to leave out. While we have given careful consideration to what material to cover, our choices inevitably also reflect our limited knowledge of the field and our own personal preferences. As the reader may already have noted, our biases have likely been in the direction of overemphasizing economic ideas, insights, and simplicity over generality.

11. For example, we do not cover here the growing literature on contracting in a "general equilibrium" setting, for example, the impact of credit rationing on macroeconomic fluctuations (see Bernanke and Gertler, 1989, and Kiyotaki and Moore, 1997) or income distribution (see Banerjee and Newman, 1991, 1993, and Aghion and Bolton, 1997). Nor do we cover more traditional general equilibrium analysis with moral hazard or adverse selection (see Prescott and Townsend, 1984, and Guesnerie, 1992). Indeed, providing a self-contained treatment of these various topics would require dealing with many technical issues that go beyond contract theory.

The final chapter of the book contains a set of exercises, which serve two purposes. First and foremost, these exercises help the reader to master some of the basic analytical techniques to solve optimal contracting problems and to develop a deeper understanding of the most important arguments. The second purpose of this chapter is to cover some classic articles that we have not had the space to cover in the main chapters. For this reason it may be worth leafing through this chapter even if the reader does not intend to try to solve any of the problems.