Introduction

ECONOMISTS have a very well-established theory of market trading and are on the way to possessing a similarly well-developed theory of contractual transactions. The economic analysis of institutions, however, is in a much more rudimentary state.

This book provides a framework for thinking about firms and other kinds of economic institutions. The basic idea is that firms arise in situations where people cannot write good contracts and where the allocation of power or control is therefore important. The book is divided into two parts. Part I is concerned with the boundaries of firms, Part II with the financial structure of firms. In this Introduction I want to sketch out some of the main themes. It is useful to start with a true story.

Recently my wife and I entered into negotiations to buy a piece of land from a contractor, on which the contractor would build us a house. Our first impression of the contractor was favourable; we hope that he felt the same way about us. None the less, it was clear as we talked about how to structure the deal that each of us had legitimate concerns. To mention a few of them, my wife and I were worried that we might spend a lot of money and end up with a house we did not like, or that the house might not be finished on time, or that the contractor might quit in the middle of the project and we would have to hire someone else to complete the house. On the other hand, the contractor was worried that he would build a house designed according to our wishes and we would then not pay for it, or that we would dismiss the contractor in the middle of the project and hire someone else, or that we would choose very expensive kitchen and bathroom fixtures which would raise his costs substantially.

In an ideal world, there would be a simple way to deal with many of these concerns. The three of us would write a binding contract that laid down each person's obligations in every conceivable eventuality and imposed large penalties if anybody failed to live up to them.[1] For example, the contract would provide a description of the house down to the bathroom taps and

[1] In an even more ideal world, we would not need a contract at all, since we could simply trust each other and rely on everyone behaving fairly.

light fixtures. It would specify how the price of the house would be altered if my wife and I made any changes—there would be one price for an extra bathroom, another for an extra fitted wardrobe, etc. There would be provisions for how the price of the house would change if the cost of raw materials rose during construction. A completion date would be specified, but extra days would be allowed if there were an unusually severe winter or if the foreman fell ill. And so on.

Unfortunately, it is impossible to write a contract as detailed as this; it is simply too difficult to anticipate all the many things that may happen. The contract we will write (at this point we have not yet agreed on a final contract!) will specify many aspects of the house, but many contingencies will be left out.[2] As the contractor admitted rather disarmingly to us, he is learning how to write contracts as he goes along. The contract he wrote with his last client had many missing provisions; he hopes this one will have fewer.[3]

Given that we will write an incomplete contract, it is clear that revisions and renegotiations will take place. In fact, the contract is best seen as providing a suitable backdrop or starting point for such renegotiations rather than specifying the final outcome. Thus, in thinking about the contract, my wife and I—and I believe the contractor too—have found it useful to imagine worst-case scenarios. We are all looking for a contract that will ensure that, whatever happens, each side has some protection, both against opportunistic behaviour by the other party and against bad luck. For example, here are two contracts that I do *not* think the three of us will sign:

Contract 1: My wife and I pay for the house and the land in advance. The contractor is then obliged to complete the house. He receives no further payment.

----

[2] Needless to say, it will be written in a language that only a lawyer can understand.

[3] An interesting example of an unforeseen contingency is the following. Between the times when the first and second drafts of this book were written, the location of the driveway became an issue. The land is situated between two roads, and my wife and I assumed that the driveway would lead on to the smaller of these. It now turns out that this location may violate town ordinances, and the driveway may have to lead on to a main road. I doubt that any of the three of us could have anticipated this eventuality.

Contract 2: We pay nothing in advance. However, we become the owner of the land immediately. We pay for the house and the land when the house is completed.

It is pretty clear why these contracts are unattractive. With the first one, my wife and I would be concerned that the contractor would disappear after we had paid him the money for the house and land, and we would be left holding only the land—or that, instead of literally disappearing, the contractor might work extremely slowly. With the second one, the contractor would be concerned that we would exclude him from the property that we had just become the owners of, and hire someone else to build the house. (In justification, we might claim that he was doing bad work.)

Instead of these 'extreme' contracts, it seems clear that we will end up writing something closer to the following contract (which is similar to contracts the contractor has used before):

Contract 3: We pay some amount initially which corresponds roughly to the value of the land. We then become the owners of the land. After this we make a series of specified advances to the builder, which are timed to occur when certain parts of the house are finished. (For example, the contractor receives 20 per cent of the price of the house when the foundation is completed, 10 per cent when the chimney is built, and so on.) A final payment is made some time after the house is completed.

The advantage of this kind of contract is that each side has some protection against worst-case scenarios; or, to put it another way, neither side can take too much advantage of the other. If the contractor disappears at any point, we could complete the project without incurring too much additional cost since we have paid an amount roughly commensurate with what we have received from him. On the other hand, if we sack the contractor at some point, he will not suffer too much since he has been paid for the work he has done.

The house story is of course just one example of an economic relationship, but it has two features that I believe are very general. The first is that contracts are incomplete. The second is that, because of this, the *ex post* allocation of power (or control) matters. Here power refers roughly to the position of each party if the

other party does not perform (e.g. if the other party behaves opportunistically). It was observed in the house example that contract 3 succeeded in sharing power reasonably between the contractor and my wife and me. In contrast, contract 1 gave the contractor too much power and contract 2 gave my wife and me too much power.[4]

In this book, I will argue that these two ideas—contractual incompleteness and power—can be used to understand a number of economic institutions and arrangements. Before I develop this theme a little further, it is worth pointing out that power is not a standard feature of economic theory. For example, take the frameworks that economists use to analyse the behaviour of economic agents: general equilibrium theory, game theory, mechanism design or principal–agent theory, and transaction cost theory. In general equilibrium theory, it is supposed that trade takes place through anonymous competitive markets and that every agent abides by the terms of any transaction he or she enters into. In such a setting, power is irrelevant. In game theory, agents may have market power, i.e. the ability to affect price. However, market power is not the same as the notion of power used in this book. Market power captures the idea that the contractor can charge my wife and me a lot for our house because there are not many competing contractors around; it says nothing about how we allocate power within our relationship. In mechanism design or principal–agent theory, it is supposed that it is costless to write a contract. An implication is that an optimal contract will be 'comprehensive', in the sense that, like the idealized house contract, it will lay down each person's obligations in every conceivable eventuality and impose large penalties if anybody fails to live up to them. But here again power is irrelevant, since an optimal comprehensive contract will not be breached or renegotiated.

Transaction cost theory comes closest to the framework presented here. However, although transaction cost theory puts a lot

[4] There is a sad postscript to the house transaction. Between the time when the second and third drafts of this book were written, the deal fell through. The reason is that the house site adjoined some wetlands and it turned out to be much harder to obtain building permission from the local conservation commission than my wife and I expected (or were led to believe). Both sides lost money, although the contractor appears to have lost more. The conservation issue is another example of an unanticipated contingency—in this case, a critical one.

of emphasis on the costs of writing contracts, and the consequent contractual incompleteness, less attention is paid to the idea that power is important or that institutional arrangements are designed to allocate power among agents.[5]

In the remainder of this introduction, I will try to impart a flavour of how contractual incompleteness and power can be used to understand some important economic phenomena. I will touch on the contents of each chapter except for Chapter 1, which surveys the literature, and Chapter 4, which discusses the foundations of the theory of incomplete contracts.

1. The meaning of ownership

Economists have written a great deal about why property rights are important, and in particular why it matters whether a machine, say, is privately owned or is common property. However, they have been less successful in explaining why it matters *who* owns a piece of private property. To understand the difficulty, consider a situation where I want to use a machine initially owned by you. One possibility is for me to buy the machine from you; another possibility is for me to rent the machine from you. If contracting costs are zero, we can sign a rental agreement that is as effective as a change in ownership. In particular, the rental contract can specify exactly what I can do with the machine, when I can have access to it, what happens if the machine breaks down, what rights you have to use the machine, and so on. Given this, however, it is unclear why changes in asset ownership ever need to take place.

In a world of contracting costs, however, renting and owning are no longer the same. If contracts are incomplete, not all the uses of the machine will be specified in all possible eventualities. The question then arises: who chooses the unspecified uses? A reasonable view is that the owner of the machine has this right; that

[5] Given its concern with power, the approach proposed in this book has something in common with Marxian theories of the capitalist–worker relationship, in particular, with the idea that an employer has power over a worker because the employer owns the physical capital the worker uses (and therefore can appropriate the worker's surplus); see e.g. Marx (1867: ch. 7). The connection between the two approaches has not so far been developed in the literature, however.

is, the owner has residual rights of control over the machine, or residual powers. For example, if the machine breaks down or requires modification and the contract is silent about this, the owner can decide how and when it is to be repaired or modified.

It is now possible to understand why it might make sense for me to buy the machine from you rather than to rent it. If I own the machine, I will have more power in our economic relationship since I will have all the residual rights of control. To put it another way, if the machine breaks down or needs to be modified, I can ensure that it is repaired or modified quickly, so that I can continue to use it productively. Knowing this, I will have a greater incentive to look after the machine, to learn to operate it, to acquire other machines that create a synergy with this machine, and so on.

Chapters 2 and 3 develop a formal theory of asset ownership based on these ideas.

2. The boundaries of firms

A long-standing issue in organization theory concerns the determinants of the boundaries of firms. Why does it matter if a particular transaction is carried out inside a firm or through the market or via a long-term contract? To put it another way, given any two firms A and B, what difference does it make if the firms transact through an arms-length contract or merge and become a single firm?

It has proved to be difficult to answer these questions using standard theory for the same reason that it is hard to explain why asset ownership matters. If contracting costs are zero, firms A and B can write a contract governing their relationship that specifies the obligations of all parties in all eventualities. Since the contract is all-inclusive, it is unclear what further aspects of their relationship could be controlled through a merger. This is true whether firms A and B are in a vertical relationship—firm A is buying an input from firm B—or in a horizontal or lateral relationship, e.g. where firms A and B sell complementary products and want to save on some duplicative production costs.

Once one recognizes that contracts are incomplete, however, it

is possible to explain why a merger might be desirable. Consider the well-known example of Fisher Body, which for many years has supplied car bodies to General Motors. For a long time Fisher Body and GM were separate firms linked by a long-term contract. However, in the 1920s GM's demand for car bodies increased substantially. After Fisher Body refused to revise the formula for determining price, GM bought Fisher out.[6]

Why did GM and Fisher Body not simply write a better contract? Arguably, GM recognized that, however good a contract it wrote with Fisher Body, situations similar to the one it had just experienced might arise again; that is, contingencies might occur that no contract could allow for. GM wanted to be sure that next time around it would be in a stronger bargaining position; in particular, it would be able to insist on extra supplies, without having to pay a great deal for them. It is reasonable to suppose that ownership of Fisher Body would provide GM with this extra power by giving it residual control rights over Fisher Body's assets. At an extreme, GM could dismiss the managers of Fisher Body if they refused to accede to GM's requests.[7]

Of course, although the acquisition increased GM's power and made GM more secure in its relationship with Fisher Body, it arguably had the opposite effect on Fisher Body. That is, Fisher Body may have had more to worry about since the merger. For example, if Fisher Body's costs fall, GM is now in a stronger position to force a reduction in the (transfer) price of car bodies, hence reducing the return to Fisher managers. Anticipating this, Fisher managers may have less incentive to figure out *how* to reduce costs. Thus, there are both costs and benefits from a merger.[8]

Chapters 2 and 3 develop a theory of the firm based on the idea that firm boundaries are chosen to allocate power optimally among the various parties to a transaction. I argue that power is a scarce resource that should never be wasted. One implication of the theory is that a merger between firms with highly

---

[6] For interesting and informative discussions of the GM–Fisher Body relationship, see Klein *et al*. (1978) and Klein (1988).

[7] There has been some debate about whether GM did in fact increase its power over Fisher Body by buying Fisher Body out; see Coase (1988: 45).

[8] Sometimes the costs of a merger will exceed the benefits. This may explain why GM did not merge with A. O. Smith, which has supplied a significant fraction of its automobile frames for many years. For a discussion of the A. O. Smith case, see Coase (1988: 45–6) and Klein (1988: 205).

complementary assets is value-enhancing, and a merger between firms with independent assets is value-reducing. The reason is as follows. If two highly complementary firms have different owners, then neither owner has real power since neither can do anything without the other. It is then better to give all the power to one of the owners through a merger. On the other hand, if two firms with independent assets merge, then the acquiring firm's owner gains little useful power, since the acquired firm's assets do not enhance his activities, but the acquired firm's owner loses useful power, since she no longer has authority over the assets she works with. In this case, it is better to divide the power between the owners by keeping the firms separate.

3. Financial securities

Debt

Suppose you have an interesting idea for a business venture, but do not have the capital to finance it. You go to a bank to get a loan. In deciding whether to finance the project, the bank is very likely to consider not only the return stream from the project, but also the resale value of any assets you have or will acquire using the bank's funds; in other words, the bank will be interested in the potential collateral for the loan. In addition, the durability of your assets and how quickly the returns come in are likely to determine the maturity structure of the loan. The bank will be more willing to lend long-term if the loan is supported by assets such as property or machines than if it is supported by inventory, and if the returns arrive in the distant future rather than right away.

These observations fit in well with the ideas emphasized in this book. Like the parties in the house transaction, the bank wants some protection against worst-case scenarios. If there is very little collateral underlying the loan, then the bank will worry that you will use its money unwisely or, in an extreme case, disappear with the money altogether. Similarly, if the collateral depreciates rapidly or the returns come in quickly, then the bank would be unhappy with a long-term loan since it would have little protection against your behaving opportunistically when the collateral

was no longer worth much or after the project returns had been realized (and 'consumed'). Basically, the bank wants to ensure a rough balance between the value of the debt outstanding and the value remaining in the project, including the value of the collateral, at all times. (Similar considerations made contract 3 in the house example attractive: neither party was in significant deficit or surplus with respect to the other party at any point in the transaction.) Chapter 5 builds a model of debt finance based on these ideas, and derives results about the kinds of project that will be financed.

Equity

Investors who finance business ventures sometimes take equity in the venture rather than debt. Equity, unlike debt, does not have a fixed set of repayments associated with it, with nonpayment triggering default. Rather, equity-holders receive dividends if and when the firm chooses to pay them. This puts equity-holders potentially at the mercy of those running the firm, who may choose to use the firm's profits to pay salaries or to reinvest rather than to pay out dividends. Thus, equity-holders need some protection. Typically, they get it in the form of votes. If things become bad enough, equity-holders have the power to remove those running the firm (the board of directors) and replace them with someone else.

However, giving outside equity-holders voting power brings costs as well as benefits. Equity-holders can use their power to take actions that ignore the (valid) interests of insiders. For example, they might close down an established, family-run business or force the business to terminate long-standing employees. The optimal allocation of power between insiders and outsiders is another of the topics covered in Chapter 5.

4. Dispersed power

So far I have supposed that those with power wield it. That is, I have assumed that an owner will exercise her residual control rights over assets; e.g. an equity-holder will use her votes to replace a bad manager. However, if power is held by many

people, then no one of them may have an incentive to be active in exercising this power. It is then important that there exist automatic mechanisms that will achieve what those with power are unable or unwilling to do by themselves.

A leading example of dispersed power is the case of a public company with many, small shareholders. Shareholders cannot run the company themselves on a day-to-day basis and so they delegate power to a board of directors and to managers. This creates a free-rider problem: an individual shareholder does not have an incentive to monitor management, since the gains from improved management are enjoyed by all shareholders, whereas the costs are borne only by those who are active. Because of this free-rider problem, the managers of a public company have a fairly free hand to pursue their own goals: these might include empire-building or the enjoyment of perquisites.

Chapters 6 and 8 explore two 'automatic' mechanisms that can improve the performance of management: debt (in combination with bankruptcy) and take-overs. Debt imposes a hard budget constraint on managers. If a company has a significant amount of debt, management is faced with a simple choice: reduce slack—that is, cut back on empire-building and perquisites—or go bankrupt. If there is a significant chance that managers will lose their jobs in bankruptcy, they are likely to choose the first option.

Take-overs provide a potential way to overcome collective action problems among shareholders. If a company is badly managed, then there is an incentive for someone to acquire a large stake in the company, improve performance, and make a gain on the shares or votes purchased. The threat of such action can persuade management to act in the interest of shareholders.

I derive some implications of these views of debt and take-overs. Chapter 6 shows that the view of debt as a constraining mechanism can explain the types of debt a company issues (how senior the debt is, whether it can be postponed). Chapter 8 shows that the possibility of take-overs can explain why many companies bundle votes and dividend claims together—that is, why they adopt a one share–one vote rule. One share–one vote protects shareholder property rights in the sense that it maximizes the chance that a control contest will be won by a management team that provides high value for shareholders, rather than high private benefits for itself.

Of course, if a company takes on debt, then there is always the chance that it will go bankrupt. If contracting costs were zero, there would be no need for a formal bankruptcy procedure because every contract would specify what should happen if some party could not meet its debt obligations. In a world of incomplete contracts, however, there is a role for bankruptcy procedure. In Chapter 7 I argue that a bankruptcy procedure should have two main goals. The first is that a bankrupt company's assets should be placed in their highest-value use. The second is that bankruptcy should be accompanied by a loss of power for management, so as to ensure that management has the right incentive to avoid bankruptcy. Chapter 7 suggests a procedure that meets these goals, and at the same time avoids some of the inefficiencies of existing US and UK procedures.

5. *An omitted topic: public ownership*

The book is concerned with the optimal allocation of privately owned assets. A very important topic not considered concerns the optimal balance between public and private ownership. Which assets should be publicly owned and which should be privately owned? This issue has always been a central one in the economic and political debate, but it has attracted new attention in the last few years as major industries have been privatized in the West and the socialist regimes in Eastern Europe and the former Soviet Union have dissolved.

It is natural to analyse public choice versus private choice using the ideas of incomplete contracts and power. If contracting costs are zero, there is no difference between the optimal regulation of a private firm on the one hand, and nationalization or public ownership on the other. In both cases the government will write a 'comprehensive' contract with the firm or its managers that will anticipate all future contingencies. The contract will specify the manager's compensation scheme, how the price of the firm's output should change if costs fall, how the nature of the firm's product should change if there is a technological innovation or a shift in demand, etc.

In contrast, in a world of incomplete contracts, public and private ownership are different, since in one case the government

has residual control rights over the firm's assets, while in the other case a private owner does. The public–private case is not a simple extension of the pure private property rights model, however. At least two new questions arise. First, what is the government's objective function? Much existing work views the government as a monolith, but this is unsatisfactory since, even more than in the case of a corporation, the government represents a collection of agents with conflicting goals: civil servants, politicians, and the citizens themselves. Second, what ensures that the government respects an agreed-on allocation of property rights? The government, unlike a private agent, can always change its mind: it can nationalize assets it has privatized or privatize assets it has nationalized.

There is a small, but growing, literature that analyses public versus private ownership in incomplete contracting terms.[9] However, much remains to be done. Developing a satisfactory theory, which deals among other things with the issue of the government's objective function and its commitment to property rights, is a challenging but fascinating task for future research.

---

[9] See, in particular, Schmidt (1990), Shapiro and Willig (1990), Shleifer and Vishny (1994), and Boycko *et al.* (1995).

PART I

Understanding Firms

The first part of the book, Chapters 1–4, is concerned with the nature and extent of the firm, that is, with the determinants of the boundaries of firms in a market economy. Chapter 1 contains a discussion of existing theories of the firm, including the neoclassical, principal–agent, and transaction cost theories. While these theories have proved very useful for some purposes, I shall argue that they cannot by themselves explain the boundaries of firms (or the internal organization of firms). Chapters 2 and 3 describe the more recent incomplete contracting or 'property rights' approach, which can throw some light on firm boundaries. This theory can also explain the meaning and importance of asset ownership. Finally, Chapter 4 provides a discussion of the foundations of the incomplete contracting model used in Part I and, to some extent, throughout the book.

1

Established Theories of the Firm

THIS chapter discusses some of the ways in which economists have looked at firms. I begin with the neoclassical theory of the firm, the standard approach found in all textbooks. I then move on to the principal–agent and transaction cost theories.[1]

1. Neoclassical theory

Neoclassical theory, which has been developed over the last one hundred years or so, views the firm mainly in technological terms. A single-product firm is represented by a production function which specifies the output level Q that is obtained when given levels of n inputs $x_1, \ldots, x_n$ are chosen. It is supposed that the firm is run by a selfless manager, M, who chooses input and output levels to maximize profit. This in turn implies that the manager minimizes costs.

The simplest case is where M purchases the n inputs in a competitive market at given prices $w_1, \ldots, w_n$, so that his total costs are $\sum_{i=1}^{n} w_i x_i$. Let $Q = f(x_1, \ldots, x_n)$ denote the firm's production function. Then, given a target output level Q, M will minimize costs by solving the following problem:

$$\text{Min} \sum_{i=1}^{n} w_i x_i$$

$$\text{s.t.} \quad f(x_1, \ldots, x_n) \geq Q.$$

Solving this for every value of Q generates a total cost curve $C(Q)$, from which can be deduced an average cost curve $C(Q)/Q$ and a marginal cost curve $C'(Q)$. The latter two curves are assumed to have the familiar shape indicated in Figure 1.1.

[1] Readers may find it useful to consult some other recent accounts of the theory of the firm, e.g. Holmstrom and Tirole (1989), Milgrom and Roberts (1992), and Radner (1992).
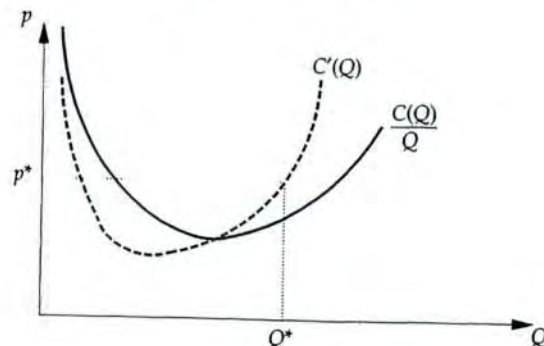
Fig. 1.1

The second stage of M's problem is to decide what level of output to produce. Under the assumption that M is a perfect competitor in the output market and faces price p^*, he maximizes $p^*Q - C(Q)$. This leads to the familiar equality between price and marginal cost, illustrated in Figure 1.1.

The U-shape of the average cost curve is justified as follows. There are some fixed costs of production (plant, machines, buildings) that must be incurred whatever the level of output. As output increases, variable costs increase, but fixed costs do not. Thus, there is a tendency for per-unit costs to fall. However, after a certain point further expansion becomes difficult because some inputs cannot be varied easily with the firm's scale. One of these is managerial talent. As output rises, the manager eventually becomes overloaded and his productivity falls. As a consequence, the firm's average cost begins to rise.[2]

How should one assess this theory of the firm? On the positive side, the theory is surely right to stress the role of technology in general, and returns to scale in particular, as important determinants of the size of firms (see e.g. Chandler 1990: 26–8). In addition, the theory has been very useful for analysing how the firm's optimal production choice varies with input and output prices, for understanding the aggregate behaviour of an industry, and for studying the consequences of strategic interaction between

[2] For a more general account of the neoclassical theory of the firm, see Mas-Colell *et al.* (1995: ch. 5).

firms once the assumption of perfect competition has been dropped (see e.g. Tirole 1988).

At the same time, the theory has several serious weaknesses. First, it completely ignores incentive problems *within* the firm. The firm is treated as a perfectly efficient 'black box', inside which everything operates perfectly smoothly and everybody does what they are told. Even a cursory glance at any actual firm suggests that this is unrealistic. Second, the theory has nothing to say about the internal organization of firms—their hierarchical structure, how decisions are delegated, who has authority. Third, and related, the theory does not satisfactorily pin down the boundaries of the firm. Among other things, it is not clear why managerial talent is a fixed factor: why can't the managerial diseconomies that lie behind the upward-sloping portion of the average cost curve be avoided through the hiring of a second manager?

It is useful to dwell on this last point since it lies at the heart of the first part of this book. Neoclassical theory is as much a theory of division or plant size as of firm size. Consider Figure 1.1 again. Imagine two 'firms' with the same production function f and cost function C, each facing the output price p^* (and both being perfect competitors). Neoclassical theory predicts that in equilibrium each firm produces Q^*. But couldn't one just as well imagine a single large firm operating with each of the smaller firms as divisions, producing $2Q^*$ altogether?

This line of reasoning suggests that it is not enough to argue that a firm will not expand because its manager has special skills and additional managers are inferior. The real issue is why it makes sense for the additional managers to be employed *outside* this firm, rather than within a division or subsidiary operated by this firm. In other words, given the original firm and a second firm employing an alternative manager, why doesn't the first firm expand—possibly laterally—by merging with the second firm?

To put it in stark terms (suggested originally by Coase 1937), neoclassical theory is consistent with there being one huge firm in the world, with every existing firm (General Electric, Exxon, Unilever, British Petroleum, . . .) being a division of this firm. It is also consistent with every plant and division of an existing firm becoming a separate and independent firm. To distinguish between these possibilities, it is necessary to introduce factors not present in the neoclassical story.

2. *The agency view*

As noted, neoclassical theory ignores all incentive problems within the firm. Over the last twenty years or so, a branch of the literature—principal–agent theory—has developed which tries to rectify this. I shall argue that principal–agent theory leads to a richer and more realistic portrayal of firms but that it leaves unresolved the basic issue of the determinants of firm boundaries.

A simple way to incorporate incentive considerations into the neoclassical model described above is to suppose that one of the inputs, input i say, has a quality that is endogenous, rather than exogenous. In particular, suppose that input i (a widget, say) is supplied by another (owner-managed) 'firm', and that the quality of this input, q, depends on the effort the supplying manager exerts, e, as well as on some randomness outside the manager's control, ϵ:

$$q = g(e, \epsilon).$$

Here ϵ is assumed to be realized after the manager's choice of e.[3]

Assume that quality q is observable and verifiable—it might represent the fraction of widgets that are not defective—but that the purchasing manager does not observe the supplying manager's effort, nor does he observe ϵ.[4] Assume also that the supplying manager dislikes high effort; represent this by a 'cost of effort' function $H(e)$. Finally, suppose (for simplicity) that only one widget is required by the purchaser ($x_i = 1$), that this yields the purchaser a revenue of $r(q)$, and that the supplying manager is risk-averse, while the purchasing manager is risk-neutral.[5]

If the purchaser could observe and verify e, he would offer the supplier a contract of the form: 'I will pay you a fixed amount P^* as long as you choose the effort level e^*.' Here e^* is chosen to be

---

[3] Another version of the principal–agent problem assumes that ϵ is realized before the manager's choice of e; see e.g. Laffont and Tirole (1993: ch. 1).

[4] The statement that q is observable and verifiable means that the parties can write an enforceable contract on the value of q.

[5] Let the purchasing manager's utility function be $U_p(r(q) - P) = r(q) - P$ and the supplying manager's utility function be $U_s(P,e) = V(P) - H(e)$, where V is concave, and P represents the payment made by the purchasing manager to the supplying manager. The principal–agent problem is interesting only if the supplying manager is risk-averse. Letting the purchasing manager be risk-neutral is a simplifying assumption.

jointly efficient for the purchaser and the supplier and P^* is determined so as to divide up the gains from trade between the two parties, according to the relative scarcity of potential purchasers and suppliers, relative bargaining power, etc.[6]

The advantage of the fixed payment P^* is that it ensures optimal risk-sharing. The purchaser bears all the risk of ϵ realizations, which is efficient since he is risk-neutral and the supplier is risk-averse.

Unfortunately, when e is not observable to the purchaser, the above contract is not feasible, since it cannot be enforced. (The purchaser would not know if the supplier deviated from $e = e^*$.) To put it another way, under the above contract, the supplier would set $e = 0$ if she dislikes working. To get the supplier to exert effort, the purchaser must pay the supplier according to *observed* performance q, i.e. he must offer the supplier an incentive scheme $P = P(q)$. In designing this incentive scheme, the two parties face the classic trade-off between optimal incentives and optimal risk-sharing. A 'high-powered' incentive scheme—that is, one where $P'(q)$ is close to $r'(q)$—is good for supplier incentives since the supplier earns a large fraction of the gains from any increase in e; but it exposes the supplier to a great deal of risk. Conversely, a low-powered scheme protects the supplier from risk but gives her little incentive to work hard.[7]

There is now a vast literature that analyses the form of the optimal incentive scheme under the above circumstances. Moreover, the basic principal–agent problem described has been extended in a number of directions. Among other things, agency theorists have allowed for: repeated relationships, several agents, several principals, several dimensions of action for the agent, career concerns and reputation effects, and so on.[8]

As a result of all this work, a rich set of results about optimal

---

[6] Given the utility functions described in n. 5, and assuming that the supplying manager receives an expected utility of U, it is easy to show that e^* maximizes $E[r(g(e,\epsilon))] - V^{-1}(U + H(e))$, and $P^* = V^{-1}(U + H(e^*))$, where E is the expectations operator. For details, see Hart and Holmstrom (1987).

[7] An optimal incentive scheme solves the following problem:

$\text{Max}_{e,P(\bullet)}\ E[r(g(e,\epsilon)) - P(g(e,\epsilon))]$

s.t. (1) $e \in \text{argmax}_{e'}\ \{E[V(P(g(e',\epsilon)))] - H(e')\}$,

(2) $E[V(P(g(e,\epsilon)))] - H(e) \geq U$.

For details, see Hart and Holmstrom (1987).

[8] For surveys, see Hart and Holmstrom (1987) and Sappington (1991).

incentive schemes has been obtained. However, although these results can throw important light on the determinants of managerial compensation packages and on certain aspects of the organization of production, the agency approach falls foul of the same criticism that besets neoclassical theory. That is, it does not pin down the *boundaries* of the firm (or say much about the internal organization of firms).

Consider again the incentive problem described above, where the optimal incentive scheme has the form $P(q)$. One interpretation of this situation is that the purchaser and supplier are independent firms linked by the arm's-length contract $P(q)$. For instance, to take an example from the Introduction, the purchaser might be General Motors and the supplier might be Fisher Body, and $P(q)$ might represent the optimal contract written by these firms when they are independent. However, another interpretation is that the purchaser and supplier are divisions of a larger firm and that $P(q)$ represents the incentive scheme that the manager of the supplying division is under. That is, according to this second interpretation, $P(q)$ is the optimal incentive scheme that the managers of the Fisher Body division face after a merger with GM. Agency theory does not distinguish between these two situations, and yet economically they seem quite different. To paraphrase Coase once again, the principal–agent view is consistent with there being one huge firm in the world, consisting of a large number of divisions linked by optimal incentive contracts; but it is also consistent with there being many small, independent firms linked by optimal arm's-length contracts. Clearly, there is something missing from the agency view of firms, just as there is something missing from the neoclassical view. The question is: what is it?

Before moving on, I should mention two possible answers to this question. First, it is sometimes claimed that asymmetries of information are reduced within the firm. For example, if the buyer of input i merges with the seller, he will be able to monitor the supplying manager's effort level better and therefore will be able to devise a better incentive scheme. The trouble with this argument is that it does not explain *why* it is easier to monitor an employee than an independent contractor. Asymmetries of information may indeed diminish within the firm, but if they do it is important to know how and why. Chapters 2 and 3 will present a

theory that can throw some light on this issue (see in particular Ch. 3, §8).

Second, it is sometimes argued that cost and/or profit-sharing becomes easier when firms merge; for example, in the above example the buyer can compensate the seller for her costs, including effort, when the buyer and seller are one firm, but not when they are two. Again, the problem with this argument is that it does not explain why cost- and profit-sharing are possible within a single firm, but not between two independent firms.[9] The theory presented in Chapters 2 and 3 will throw some light on this issue too (see in particular Ch. 3, §4).

3. Transaction cost theories

The distinction between comprehensive and incomplete contracts

One important factor missing from the principal–agent view is the recognition that writing a (good) contract is itself costly. This is a theme that lies at the heart of the large transaction cost literature which started with Ronald Coase's famous 1937 paper and has been extensively developed by Oliver Williamson and others (see in particular Williamson 1975, 1985, and Klein *et al.* 1978).

It is worth noting first that agency theory itself already incorporates some contracting costs. In the simple model considered above, it is supposed that managerial effort, e, cannot be made part of an enforceable contract since it is observed only by the agent. Another way of saying this is that the cost of putting e into the incentive contract is infinite. However, this is probably not a very useful way of looking at things. Agency theory ascribes all contracting costs to the cost of *observing* variables. If a variable is observable by both parties, then the theory assumes that it can be contracted on costlessly. But this is not the same as supposing that it is costly to *write* a contract.

This point can be made a little more sharply as follows. Although the optimal contract in a standard principal–agent

---

[9] One possible reason is that profit-sharing between independent firms may be illegal if the firms compete in the same input or output markets. This may explain some mergers.

model will not be first-best (since it cannot be conditioned directly on variables like effort that are observed by only one party), it will be 'comprehensive' in the sense that it will specify all parties' obligations in all future states of the world, to the fullest extent possible. As a result, there will never be a need for the parties to revise or renegotiate the contract as the future unfolds. The reason is that, if the parties ever changed or added a contract clause, this change or addition could have been anticipated and built into the original contract.[10]

It may be worth spelling this out. Consider the simple model of Section 2, but suppose there are two verifiable states of the world, $s = s_1$ and $s = s_2$, and that the value of s affects production quality:

$$q = g(e, \epsilon, s).$$

Assume also that s is learned by both parties at date 1 before production takes place, and that the parties contract at some prior date 0. Then in general it is optimal for the parties to write a state-contingent contract that specifies two incentive schemes, $P(q,s_1)$ and $P(q,s_2)$, one to apply when $s = s_1$ and the other to apply when $s = s_2$.

Now suppose that, after $s = s_1$ is realized, the principal and agent decide that they can do better by replacing the second-period incentive scheme $P(q,s_1)$ by $\hat{P}(q,s_1)$. Then, given that the parties have perfect foresight, they will recognize that the second-period incentive scheme $P(q,s_1)$ will not stand. But, given this, the parties may as well substitute $\hat{P}(q,s_1)$ for $P(q,s_1)$ in the original date 0 contract. In other words, given any contract that is renegotiated on the equilibrium path, there is an equivalent one that is not.[11]

---

[10] One would also not expect to see any legal disputes in a comprehensive contracting world. The reason is that, since a comprehensive contract specifies everybody's obligations in every eventuality, the courts should simply enforce the contract as it stands in the event of a dispute.

[11] In fact, not only does renegotiation not add anything in the standard principal–agent model, but the possibility of it can actually *worsen* matters! See Dewatripont (1989). The argument is subtle, but the following gives a rough idea. In some principal–agent models the presence of inefficiency in later periods can improve the agent's incentives (for truth-telling or exerting effort) in earlier periods so much that the later inefficiency is worthwhile. However, the later inefficiency is not credible: when the future arrives, the parties will renegotiate their contract to eliminate the inefficiency. In a situation like this, the parties would be better off if they could commit not to renegotiate. The problem is that it is not clear how they can achieve this.

The sources of transaction costs

In reality, contracts are not comprehensive and are revised and renegotiated all the time. According to the transaction cost literature, this is a consequence of three factors missing from the standard principal–agent story. First, in a complex and highly unpredictable world, it is hard for people to *think* very far ahead and to plan for all the various contingencies that may arise. Second, even if individual plans can be made, it is hard for the contracting parties to *negotiate* about these plans, not least because they have to find a common language to describe states of the world and actions with respect to which prior experience may not provide much of a guide. Third, even if the parties can plan and negotiate about the future, it may be very difficult for them to *write* their plans down in such a way that, in the event of a dispute, an outside authority—a court, say—can figure out what these plans mean and enforce them. In other words, the parties must be able to communicate not only with each other, but also with outsiders who may have little knowledge about the environment in which the contracting parties operate.[12]

As a result of these three contracting costs, the parties will write a contract that is incomplete. That is, the contract will contain gaps and missing provisions. In particular, it will be silent about the parties' obligations in some states of the world and will specify these obligations only coarsely or ambiguously in other states of the world. For example, the contract might not specify what is to happen if the supplier's factory burns down, because this is not anticipated; or the contract might say that the supplier must always supply one widget, rather than a number of widgets that varies with the state of the world, because it is too costly to distinguish between different states of the world. The contract might also be short-term; that is, it might specify the parties' obligations only up to some date, T.

It is useful to illustrate these points with the General Motors–Fisher Body example from the Introduction. In a constant world, GM and Fisher might find it easy to write a long-term contract specifying the quantity, quality, and price of the car bodies Fisher supplies to GM. For example, it might be optimal for them

---

[12] All of these may be regarded as different forms of 'bounded rationality'. For further discussions, see Coase (1937), Williamson (1985), and Klein *et al.* (1978).

to agree that Fisher should supply 2,000 bodies, of a particular type, each day at a particular price for the foreseeable future.

Now consider the case where the world is changing. The optimal number of car bodies, type of car bodies, and price may depend on a variety of factors, e.g. the demand for General Motors' output, Fisher Body's costs, actions of competitors, new regulations on car pollution, whether a trade agreement is reached with the Japanese, and innovations occurring in car and body production.

It may be prohibitively expensive to write a contract that conditions quantity, quality, and price on all of the external factors just described. This is not just because some of the variables are privately observed, but also because, even if publicly observable, the variables are inherently hard to specify in advance in an unambiguous manner. For example, there may be no objective way of measuring the demand for cars, or the degree of innovation, or the extent of government regulation, or the actions of competitors. Thus, a contract that tries to condition variables on these factors may not be enforceable by a court. In addition, even if it is possible for the parties to anticipate and contract on some of the factors that may be relevant for their relationship, there may be many others that the parties do not anticipate; for example, they may foresee the possibility of a new trade agreement, but not of new regulations on car pollution. Under these conditions, the parties are likely to write an incomplete contract. For example, the contract might be short-term. GM and Fisher may be able to look five years ahead, but not much further. So they might write a contract specifying that Fisher should supply 2,000 bodies per day, of a particular type, at a particular price for the next five years. Both parties realize that towards the end of the five-year period they will have further common information about demand, costs, competitors' strategies, regulations, etc., and that they can then write a new contract for another five years; and so on.

The economic implications of contractual incompleteness

As observed, an incomplete contract will be revised and/or renegotiated as the future unfolds. In fact, given that the parties can fill in the gaps as they go along, one may ask why contractual

incompleteness matters. The reason is that the renegotiation process imposes several costs. Some of these are *ex post* costs, incurred at the renegotiation stage itself, and others are *ex ante* costs, incurred in anticipation of renegotiation.[13]

First, the parties may engage in a great deal of haggling over the terms of the revised contract. Argument about division of surplus serves no overall productive purpose and, to the extent that it is time-consuming and wastes resources, it is inefficient.[14]

Second, not only may the process of *ex post* bargaining be costly, but, to the extent that the parties have asymmetric information, they may fail to reach an efficient agreement. Suppose that at the recontracting stage Fisher Body knows the current cost of producing car bodies, but General Motors does not. (It knows only the probability distribution from which costs are drawn.) GM could ensure a supply of bodies by offering Fisher an attractive price (high enough to cover Fisher's costs with probability one). However, this is expensive, since GM is overpaying in states of the world where costs are low. GM may prefer to offer a low price, knowing that in high-cost states Fisher will not supply (even though the cost of bodies is less than their value to GM). In other words, profit-maximizing behaviour by GM leads to absence of profitable trade with positive probability.[15]

It is worth noting at this point that neither of the above *ex post* costs would be significant if the parties could easily switch to new trading partners at the renegotiation stage. Any attempt by Fisher (resp. GM) to haggle for an increased share of surplus would fail if GM (Fisher) could switch to an equally efficient alternative partner. Similarly, asymmetric information does not lead to *ex post* inefficiencies if, after a bargaining breakdown, the parties can (costlessly) start the process again with new identical partners; if GM's low-price offer is turned down, GM will either eventually

13 As will become clear, it is unfair, strictly speaking, to blame these costs on renegotiation *per se*. Given contractual incompleteness, these costs would likely be even greater in the *absence* of renegotiation, i.e. if the parties had to stick to the original incomplete contract.
14 In addition, there may be costly legal disputes because an incomplete contract will be ambiguous and the parties will look to the courts to resolve the ambiguity.
15 See e.g. Fudenberg and Tirole (1991: ch. 10) or Myerson and Satterthwaite (1983). The parties may also fail to reach an agreement at the *ex ante* stage of contractual negotiation if there is asymmetric information at this point.

find a low-cost supplier, or, if all suppliers have high costs, GM will learn this and raise its price.

Thus, if the two costs described are high, it must be because there is something binding the partners together and making it difficult for them to switch at the recontracting stage. The leading candidate for that 'something' is an *ex ante relationship-specific investment*, that is, a prior investment, which creates value if the parties' economic relationship extends over time, but does not if the parties split up. In the case of GM and Fisher, examples of relationship-specific investments might be GM's decision to locate a car assembly plant near Fisher's factory, Fisher's decision to locate its factory near GM's assembly plant, GM's decision to spend money on developing cars that rely on the bodies supplied by Fisher, and Fisher's decision to spend money figuring out how to reduce the cost of producing bodies for GM (but not for car manufacturers in general).[16]

Once the existence of relationship-specific investments is recognized, it becomes apparent that there can be a third cost of contractual incompleteness that may dwarf the haggling and *ex post* inefficiency costs discussed so far. Specifically, because contracts are incomplete, the parties may be deterred from making the relationship-specific investments that would be optimal in a 'first-best' world. Suppose it is efficient for Fisher to install machinery that enables it to produce car bodies designed specifically for GM. In a comprehensive contracting world, the contract between GM and Fisher could be structured in such a way that Fisher would have an incentive to make the investment; one way to do this is to have a contract that fixes the body price for the indefinite future, so that the gains from the investment accrue to Fisher.

In an incomplete contracting world, however, such an arrangement may be impossible. Because of the difficulty of specifying quality and quantity very far in advance, both parties recognize that any long-term contract is incomplete and subject to renegotiation. Even if renegotiation proceeds smoothly—that is, if problems due to haggling and asymmetric information do not arise—the division of the gains from trade will depend on the *ex post* bargaining strengths of the parties rather than on what is

[16] Insightful discussions about relationship-specific investments may be found in Williamson (1985), Klein *et al.* (1978), and Joskow (1985).

specified in the initial contract or on economic efficiency. As a result, a party may be reluctant to invest because it fears expropriation by the other party at the recontracting stage, i.e. it fears that it will not cover its investment costs.

For example, Fisher will worry that, once it has installed the special machinery, GM will use its bargaining power to set the price of car bodies close to Fisher's variable production costs, thus causing Fisher to make a loss on its initial fixed investment. Alternatively, GM may insist that Fisher's output satisfy very demanding quality criteria. Similarly, GM will worry that, once it has sunk the costs of developing a car that uses Fisher's bodies, Fisher will use its bargaining power to set the price of car bodies close to GM's variable production benefits, thus causing GM to make a loss on *its* initial fixed investment.

Given each party's fear that the other party will 'hold it up' at the renegotiation stage, the parties are likely to make investments that are relatively non-specific.[17] For example, Fisher may decide to install general-purpose machinery that enables it to supply a range of car manufacturers; this way Fisher can play one manufacturer off against the other at the recontracting stage and achieve a higher price for its output. Similarly, GM may decide to develop a car that can use bodies produced by a number of different suppliers, rather than only Fisher bodies. Such decisions sacrifice some of the efficiency benefits of specialization, but, in a world of incomplete contracting, these efficiency losses are more than offset by the security that a general investment provides for each party.

So far I have talked about the costs that plague the relationship between two independent, i.e. non-integrated, firms. The next question to ask is: how would these costs change if the two firms merged and became a single firm? Here transaction cost theory becomes somewhat vaguer. It is often suggested that haggling problems and hold-up behaviour are reduced in a single firm. However, the precise mechanism by which this happens is not usually spelled out. In the discussion of principal–agent theory, I argued that it is unsatisfactory to assume that the informational structure changes directly as a result of a merger. In the same way, it is unsatisfactory to suppose that the agents automatically

[17] For discussions of the hold-up problem, see Williamson (1985) and Klein *et al.* (1978). For formalizations, see Grout (1984) and Tirole (1986a).

become less opportunistic. (Also, presumably opportunistic behaviour is not always reduced within the firm, since otherwise it would be optimal to carry out all economic activities within one huge firm.) If there is less haggling and hold-up behaviour in a merged firm, it is important to know *why*. Transaction cost theory, as it stands, does not provide the answer.

2

The Property Rights Approach

ALL the theories discussed in Chapter 1 suffer from the drawback that they do not explain what changes when two firms merge. I now describe a theory—the property rights approach—that tries to address this question head-on.[1] I divide the chapter into three parts. Section 1 provides a verbal description of the property rights approach. Section 2 develops a formal model under a very stylized set of assumptions. Finally, Section 3 discusses what light the theory can throw on actual organizational arrangements.

1. *A general description*

Consider two firms, A and B, and imagine that firm A acquires firm B. Ask the following question: what exactly does firm A get for its money? At least in a legal sense, the answer seems straightforward: firm A acquires, i.e. becomes the owner of, firm B's assets. Included in this category are firm B's machines, inventories, buildings, land, patents, client lists, copyrights, etc.—that is, all of firm B's physical or nonhuman assets. Excluded are the human assets of those people working for firm B; given the absence of slavery, the human capital of these workers belongs to them both before and after the acquisition.

Why does ownership of physical or nonhuman assets matter? The answer is that ownership is a source of power when contracts are incomplete. To understand this, note that an incomplete contract will have gaps, missing provisions, or ambiguities, and so situations will occur in which some aspects of the uses of nonhuman assets are not specified. For example, a contract between General Motors and Fisher Body might leave open certain aspects of maintenance policy for Fisher machines, or might not specify

---

[1] The following is based on Grossman and Hart (1986) and Hart and Moore (1990). The account in the first part of the chapter is drawn from Hart (1989).