they use (in particular) the following variables: the specificity of a component to the automaker (is the component designed specifically for that automaker? Can it be bought from other suppliers on short notice?) and the complexity of the system in which the component must be inserted (engine, chassis, etc.). They show that, in particular, the specificity variable is a significant determinant of the integration decision. Masten (1984) finds analogous results for aerospace manufacturers. In a similar spirit, Anderson and Schmittlein (1984) study the integration of the sales force (manufacturer's representative versus employee salesperson) and show that the degree of human-specific capital (measured by the manager's assessment of the difficulty in learning the ins and outs of the company, the nature of the product, the nature of the consumers, and so forth) is negatively related to the probability of using independent representatives.57

Reputation as a Substitute for Contracts or Integration

The concept behind this and previous subsections is that, in order to avoid future hazards, parties should sign complete contracts, or, if contracts are too costly or impossible to write, should at least make a correct use of the authority structure (restricted contract). In practice, however, Macaulay (1963) found that relations between firms tended to be more informal than was predicted by the theory. This is often true even when firms engage in long-run relationships. Efficiency is then sustained by the firms' reputation. A firm that cheats at some date (i.e., makes decisions that are not jointly efficient) runs the risk of losing future profitable deals with its partner (see Williamson 1975, chapter 6; Kreps 1984, subsection 2.2 of this chapter, and chapters 2 and 6 below). Reputation allows a firm to save on the costs of writing complete contracts or even on the costs of distributing authority. On the other hand, informality exposes the firms to the threat of opportunism. Thus, one would expect informality to be most prevalent when specific investments are limited and when trade is sufficiently frequent that the incentive to cheat is low.

Dual Sourcing as a Substitute for Contracts

An alternative way of avoiding the ex post hold-up problem is to introduce ex post competition whenever that is feasible. Farrell and Gallini (1986) and Shepard (1986) have analyzed Williamsonian models in which the buyer invests in specific assets and the seller chooses, ex post, some ex ante noncontractible variable (call it "quality").58 Ex post, the seller has an incentive to choose low quality; therefore, ex ante, the buyer invests little in the relationship. Dual sourcing consists in having two or more suppliers, who compete ex post on quality. This raises the equilibrium level of quality and the ex ante investment. Competition can thus alleviate the ex post bilateral monopoly problem and raise efficiency. Farrell and Gallini and Shepard argue that this is a persuasive explanation of why Intel licenses its microprocessor technologies and why IBM adopts an "open architecture" policy in regard to its personal computers.

2 The Profit-Maximization Hypothesis

It is a postulate of this book, and of most economic theory, that firms maximize expected profits. There is, however, a widespread feeling that in practice their managers have other objectives (e.g., maximizing the firm's size and growth and the perquisites of the managerial position).59 This section presents arguments for and against the profit-maximization hypothesis. It also discusses the power of the current theory of industrial organization in the presence of non-profit-maximizing firms.

The shareholders of a firm are claimants for its revenue, net of various input costs. Thus, if they were able to run the firm, they would choose decisions that would

---

56. The idea behind the introduction of this variable is as follows: Both Williamson (1975) and Scherer (1980, p. 90) have argued that vertical integration permits executive fiat to obtain better coordination (because the exact timing of the production process is generally left indeterminate in the incomplete contract). A complex system may require more coordination of inputs.

57. See Williamson 1983 for other interesting examples.

58. This variable is a priori in Farrell and Gallini 1986 and a delivery lag in Shepard 1986.

59. In these models, a high prior or a long delivery lag.

60. See the models of firm behavior of the 1950s and the 1960s (e.g., Baumol 1962, Marx 1964).
minimize cost and maximize profit. Thus, non-profit-maximization is mainly associated with the separation of ownership and control. Adherents of the principal-agent theory and its eminences have, since the early 1970s, taken the approach that firms’ deviations from profit-maximizing behavior should be explained rather than postulated and that deviations should be traced to the inability of shareholders to adequately monitor the managers and discover the firm’s cost-and-demand situation. For instance, managerial discretion (e.g. the granting of perquisites) is permitted by the managers’ informational superiority relative to the shareholders. The firm’s concern for size or growth may not be attributable directly to its shareholders or its managers’ intrinsic preference for such attributes; rather, it may be due to conflicts between shareholders and managers. For instance, the shareholders’ incomplete information about the firm’s technology may allow the managers to inflate the need for personnel, lowering pressure on the job (or, equivalently, increasing on-the-job leisure). Similarly, the firm’s growth may be desired by the managers not for its own sake but because it allows them and their subordinates to enjoy greater opportunities for promotion. 

Reviewing the principal-agent literature and alternative approaches is beyond the scope of this chapter. This treatment will be restricted to the main issues. First, we will consider the basic moral-hazard problem and how direct monetary incentives, yardstick competition, takeover bids, product-market competition, and supervision can reduce managerial slack or discretion. The limits of these control mechanisms will be pointed out. It will then be argued that even if managerial slack invalidates the profit-maximization hypothesis, the implications of this hypothesis for industrial organization need not be erroneous.

The managerial rewards that will be discussed below should be taken in a broad sense. Such rewards can be monetary (as in the models below), but could also consist of promotions, tolerance of perquisites, prestige, assignment of cash flow to one’s division, and so on. Further more, it should be noted that the principal-agent relationship will be discussed mainly in the context of the separation of ownership and control between shareholders and managers. Clearly, many of the incentive devices described below apply to other tiers of the firm’s hierarchy; indeed, some of them apply more to lower tiers than to executive officers.

2.1 The Basic Incentive Problem

The agency problem in its moral-hazard form (the focus of this subsection) stems from a basic conflict between insurance and incentives. On the one hand, the theory of optimal insurance demonstrates that the optimal division of a pie of a random size (the profit) between a risk-neutral party (the shareholders) and a risk-averse one (the manager) has the risk-neutral party bear all the risk, if incentive issues are left aside. (See, e.g., Arrow 1970 and Borch 1963.) Suppose there is a pie of random size \( \Pi \) to be divided between the two parties, and that this random variable is not affected by the parties’ actions. Let \( \Pi \) take values in a discrete set \( \Pi_1 < \cdots < \Pi_\lambda \).
with probabilities \( p_i \), \( i = 1, \ldots, n \), where \( p_i > 0 \) and \( \sum_{i=1}^{n} p_i = 1 \). Let \( \Pi = w(\Pi) \) and \( u(\Pi) \) denote the allocations to the risk-neutral party and the risk-averse party when the realization is \( \Pi \). The parties' expected utilities are

\[
E\left[ u(\Pi) \right] = \sum_{i} p_i u(\Pi_i) \tag{1}
\]

and

\[
E\left[ w(\Pi) \right] = \sum_{i} p_i w(\Pi_i) \tag{2}
\]

respectively, where \( w_0 = w(\Pi_0) \). An efficient (or Pareto-optimal) contract maximizes the utility of one party given the level of utility of the other party. It satisfies

\[\max \, \sum_{i} p_i (u_i - w_i), \quad \text{s.t.} \quad \sum_{i} p_i u(\pi_i) \geq U_0, \]

where \( U_0 \) is a constant. The Lagrangian for this program is

\[L = \sum_{i} p_i (u_i - w_i) + \lambda \left( \sum_{i} p_i u(\Pi_i) - U_0 \right)\]

Taking the derivatives with respect to the \( w_i \), one gets for all \( i \)

\[u'(u_i) = 1/\lambda.\]

So \( w_i \) is independent of \( i \) if the manager is strictly risk-averse (\( u'' < 0 \)). The same result holds for a continuous distribution for \( \Pi \).

Thus, the risk-averse party should get full insurance (i.e., should have a constant income over all states of nature). This is where the issue of incentives arises. Suppose the risk-averse party takes some unobservable action that affects the size of the pie to be divided (in a stochastic sense) and that this action is costly to him. Think of this action as a level of effort (it could be a more general discretionary choice). Suppose, further, that the risk-neutral party observes only the realization of the pie (the level of profit). The risk-averse party, if given an income that does not depend on this realization, has no incentive to exert effort, because his effort does not affect his income. So full insurance conflicts with incentives. Indeed, the trade-off between insurance and incentive objectives generally leaves the parties with both suboptimal insurance and suboptimal profits.

There is one case in which this trade-off does not arise. Assume that both parties are risk-neutral (in particular, \( u' \) is constant), so that the party who takes the unobservable action (the agent) does not need to be insured. The other party (the principal) can ensure that the agent takes the jointly optimal action by "selling" the pie to him—that is, the principal receives a transfer price independent of the size of the pie, and the agent becomes the residual claimant for the remainder of the pie. Because the agent's expected income is equal (up to the fixed transfer price) to the expected size of the pie, the agent has all the incentive to pick the optimal action, i.e., the action that maximizes the expected size of the pie net of the action's cost (see section 3). The agent bears all the risk under this arrangement, but this does not matter because he is risk-neutral. Residual claimancy for the party that takes the unobservable action offers a very general solution to the incentive problem, and it will be encountered again in chapter 3 and especially in chapter 4. However, it is clear that for a risk-averse agent the residual claimancy conflicts with the insurance objective.

Finding the optimal incentive scheme when the agent is risk-averse is a complex task. (The supplementary section contains a few results.) The following simple examples illustrate the issues.

Example 1

A firm's profit may take one of two values: \( \Pi_1 \) and \( \Pi_2 \) (with \( \Pi_1 < \Pi_2 \)). The firm is run by a manager, who chooses between two levels of effort: high ('work') and low ('shirk'). The manager has utility \( U = u(w - \Phi) \) when he works and \( U = u(w) \) when he shirks, where \( w \) is the manager's wage, \( u \) is an increasing, concave function (with \( \lim_{w \to \infty} u'(w) = +\infty \)), and \( \Phi \) (the monetary disutility of high effort) is strictly positive. The manager's objective function is the expectation of \( u \). Working outside the firm, he would get \( U_0 = u(w_0) \). So, to ensure his participation, the shareholders must give him an expected utility of at least \( U_0 \). \( U_0 \) is called the reservation (net) wage. The shareholders' objective function is the expectation of the net profit \( \Pi - \Phi \).

The technology is as follows: If the manager works, the profit is \( \Pi_1 \) with probability \( x \) and equal to \( \Pi_1 \) with probability \( 1 - x \). If he does not work, the profit is \( \Pi_1 \) with probability \( y \) and \( \Pi_1 \) with probability \( 1 - y \). One has \( 0 < y < x < 1 \).

Assume that the manager's contract is chosen by the shareholders.
First, suppose that the manager’s effort is observed by the shareholders, who can then choose any level of effort they want and impose it on the manager (with the threat of a large punishment if he disobeys). Because effort is observable, there is no incentive issue; hence, the optimal contract calls for full insurance. Suppose first that they demand a low level of effort. Optimal insurance implies that \( w_1 = w_2 = w_0 \), where the second equality is due to the fact that the shareholders neither want nor need to give the manager more than his reservation wage. The shareholders’ profits are

\[
y \Pi_2 + (1 - y) \Pi_1 - w_0.
\]

Now, suppose the shareholders demand a high level of effort. Optimal insurance again calls for a constant net wage for the manager, so

\[
w_1 - \Phi = w_2 - \Phi = w_0.
\]

The shareholders’ expected profits are then

\[
x \Pi_2 + (1 - x) \Pi_1 - (w_0 + \Phi),
\]

To make things interesting, assume that it is optimal for the shareholders to demand the high effort:

\[
x \Pi_2 + (1 - x) \Pi_1 - (w_0 + \Phi) > y \Pi_2 + (1 - y) \Pi_1 - w_0,
\]

or

\[
(x - y) (\Pi_2 - \Pi_1) > \Phi. \tag{2}
\]

In words, the increase in expected profits exceeds the disutility of effort.

Now consider the more interesting case in which the manager’s effort is not observable by the shareholders. As has been noted, a high effort cannot be induced by a constant wage structure. Instead, the shareholders must reward the manager when profits are high. Suppose that the shareholders still want to induce the high effort. They must design a wage structure that satisfies the “incentive-compatibility” constraint:

\[
x u(w_2 - \Phi) + (1 - x) u(w_1 - \Phi) \geq y u(w_2) + (1 - y) u(w_1), \tag{3}
\]

where \( w_i \) is the wage paid when realized profits are \( \Pi_i \).

(Equation 3 implies that \( w_2 > w_1 \).)\textsuperscript{65}

To the incentive-compatibility constraint we must add the “individual-rationality” or “participation” constraint:

\[
x u(w_2 - \Phi) + (1 - x) u(w_1 - \Phi) \geq u(w_0).
\]

The shareholders’ expected profit is then

\[
x u(w_2) + (1 - x) (\Pi_2 - w_2).
\]

It is easily seen that, in the maximization of the shareholders’ profits with respect to equations 3 and 4, both constraints are binding. Suppose the incentive-compatibility constraint is not binding. Maximization of expected shareholders’ profit subject to the participation constraint yields full insurance \( w_1 = w_2 \), as has been shown, but this wage structure does not satisfy the incentive-compatibility constraint. Conversely, suppose that only the incentive-compatibility constraint is binding. Then the shareholders can reduce \( w_1, w_2 \), and keep this constraint satisfied; if the decrease in \( w_1 \) is not too large, the participation constraint is still satisfied. Thus, in this simple case, the optimal wage structure, given that the high effort is to be induced, is obtained from equations 3 and 4 satisfied with equality:

\[
x u(w_2 - \Phi) + (1 - x) u(w_1 - \Phi) = y u(w_2) + (1 - y) u(w_1), \tag{3’}
\]

\[
x u(w_2) + (1 - x) u(w_1 - \Phi) = u(w_0)., \tag{4’}
\]

The shareholders’ profit is lower under unobservability—equation 4’ and the concavity of \( u \) imply that the expected wage bill, \( x w_2 + (1 - x) w_1 \), strictly exceeds \( w_0 + \Phi \), as figure 3 shows.\textsuperscript{66} Hence, to induce the high effort and obtain the high profit with probability \( x \), the wage bill must be higher than under effort observability.

On the other hand, if the shareholders wished to induce the low effort without the observability, they would not suffer from effort unobservability. The full-information wage is constant \( w_1 = w_2 = w_0 \), and it also induces the low effort under unobservability. Thus, the relative desirability of inducing the high effort is lower under unobservability; that is, the shareholders may be happier

\textsuperscript{65} The left-hand side of equation 3 is strictly lower than \( x u(w_2) \) + \( y u(w_0) \), which is lower than the right-hand side if \( u_2 < u_1 \) (recall that \( x > x \)).

\textsuperscript{66} Figure 3 is an illustration of Jensen’s inequality, according to which the expectation of a concave function of a random variable is lower than the value of the function evaluated at the expectation of the random variable.
with the low effort under unobservability even if equation 2 is satisfied.

In summary: This simple model highlights the following points: Effort, if it is not observed, must be induced through incentives. The manager’s wage must grow with the realized profit. Because such incentive structures destroy insurance, the expected wage bill required to obtain effort is higher under nonobservability. This, in turn, may make the shareholders not wish to induce effort; that is, they may tolerate slacking.

Two further important points can be derived from straightforward variants of this model.

Observability, Verifiability, and Authority

The difference between “observability” and “verifiability” (which has not been explored up to this point) relates to the possibility that the principal can observe the agent’s performance but cannot verify his observations (i.e., cannot supply sufficient evidence) to a court. Because performance cannot be verified by a court, contracts that are contingent on performance (e.g., contracts that read “If the agent’s performance meets such and such standard, we will pay him so much”) cannot be made, as the courts will be unable to enforce them.67 For example, when the agent is part of a productive team, reliable accounting procedures may measure only the team’s performance, not individual contributions. However, an insider (the chief executive officer or a supervisor, say) may be able to disentangle these contributions, whereas an outsider (a judge) cannot. This applies equally well to the performance of complementary divisions (manufacturing and marketing, for instance) or to that of team workers.

Now, suppose that in an agency problem II is observable by the principal but is not verifiable, so the contract cannot depend directly on the agent’s performance. Can the principal be trusted to announce truthfully what he observes? A priori, no. Suppose that in the previous model the optimal contract induces effort when profit is verifiable. When profit is only observable, the principal has an incentive to claim that it is low (II1) even when it is high (II2), since w1 < w2. There is a simple conflict of interest for the principal.

The picture changes dramatically when the principal oversees many agents (division managers, workers, etc.). To simplify, consider a large number N of agents, each of whom produces an observable but unverifiable profit. As in the previous model, the probability that the individual performance is II1 rather than II2 is x or y, depending on whether or not the agent exerts effort. The probabilities are independent.68 Consider the following commitment by the principal: “I will pay a wage w2 to x percent of my agents (the ones I announce to be the most productive ones), and a wage w1 to the rest,” where w1 and w2 solve equations 3’ and 4’ (i.e., are the optimal wages under verifiability). Clearly, the total wage bill,

\[ N[xw_2 + (1-x)w_1], \]

67. In the model discussed above, II need not represent the firm’s profit; it could also represent an employer’s performance.

68. Correlation between the production processes is not required here, as it is in the case of the tournament mechanism given below.
is fixed, and the principal does not have an incentive to misrepresent the individual performances. Conversely, if the agents all exert effort, they know that 2 percent of them will yield profit $\Pi_2$ (by the law of large numbers) and will receive wage $w_2$. Those yielding profit $\Pi_1$ will receive wage $w_1$. Hence, the incentive-compatibility constraint and the individual-rationality constraint are satisfied. With many agents, the principal can be given the authority to choose rewards, because he can commit himself to an overall reward policy. Thus, verifiability is obtained indirectly.70

Remark The rewarding of observable but verifiable performance through authority is a more general phenomenon. Here, the fixed size of the aggregate reward prevents authority from becoming arbitrary by removing the principal’s incentive to misrepresent the agents’ performances. An alternative but similar mechanism, which works even with a single agent, exists if the principal has a reputation to defend. For instance, an employer who has a reputation for treating his employees fairly (rewarding them according to their performance)—that is, an employer who does not abuse authority—is able to offer his employees better incentives, and therefore may be reluctant to milk his reputation by not rewarding them fairly simply to increase short-run profit.

Limited Punishment and Managerial Rents

In the previous model, it was shown that the manager’s individual-rationality constraint is binding. The argument was that if it were not binding, the shareholders could reduce the wage $w_1$ a bit; this would not impair incentives and would still induce the manager to participate. In some circumstances, however, reducing the wage may not be possible. Suppose that, because of limited liability and laws against slavery, the harshest punishment that can be imposed on the manager is for him to receive $w_0$ (so $w_i \geq w_0$ for all $i$ is a new constraint in the design of the wage structure). One can interpret $w_0$ as the equivalent in terms of utility of the wage that the manager can get elsewhere net of the search costs.71 Alternatively, one might imagine that the manager becomes infinitely riskaverse under $w_0$; $w_0$ could be like a subsistence level, a fall slightly below which would yield utility $-\infty$ (“death”) to the manager.

Because $w_2$ and $w_1$ necessarily exceed $w_0$ (weakly), and because the manager can always choose not to work, the participation constraint is automatically satisfied. Suppose the shareholders still want to induce effort (which will be the case if $\Pi_2 - \Pi_1$ is large enough). To do so they need to impose a wage differential between the two levels of profit: $w_2 > w_1 \geq w_0$. Again, because the manager can always choose not to work, his expected utility is no less than

$$y w_2 + (1 - y) n(w_1) > s(w_0).$$

The participation constraint is not binding, which means that the manager enjoys a rent within the firm.72 (Here $y_0$ is equal to $w_0$, and $w_0$ is then given by equation 3.)

The possibility of rents (Calvo 1977; Calvo and Wellisz 1978, 1979) underlies much of the efficiency-wage hypothesis for involuntary unemployment, according to which workers inside a firm are given a rent relative to unemployed ones as a way of giving them work incentives in the face of limited punishments (see, e.g., Shapiro and Stiglitz 1984).73

Example 2

The purpose of this example is to confirm our earlier intuitions on a simple example with a continuous choice of effort.74 The manager chooses a level of effort $x$ on the real line. His utility is equal to $w(x - R x^2/2)$, where $R$ is a parameter of disutility of labor and $w$ satisfies the assumptions of example 1. His net reservation wage is $w_0$, so the participation constraint is

---

69. As long as he does not collide with some of the agents.
70. A similar argument is made in Shachachy 1963, Cornishell 1983, and Malcomson 1984. B is still held with a finite number of agents (i.e., not a large one, though not in such a stark way.
71. There are some subtle issues of timing of production and profit accounting here, but a dynamic model would complicate the argument. Note also that the harshest punishment can be slightly below $w_0$ without any change in the reasoning.
72. Technically, the existence of a rent is associated with the impossibility of imposing punishments beyond a given level. For a sufficient condition to rule out rents, see the earlier reasoning and (more generally) proposition 2 in Grosman and Hart 1983.
73. See Milgrom 1984 for a further discussion of rents attached to jobs.
74. This example is taken from Persson 1984, where it is attributed to Berghold (1977) and Stiglitz (1979).
where the expectation is with respect to $e$. The gross profit for the shareholders is $\Pi = e + a$, where $e$ is a random variable such that $Ee = 0$. (We will continue assuming that randomness occurs after the choice of effort, although in the current model it could occur and be observed by the agent between the signature of the contract and the choice of effort without any change in the argument.)

If the shareholders can observe effort, the optimal contract involves a fixed wage $w = \tilde{w}$. For a given effort level $e$, this wage is given by the participation constraint:

$\tilde{w} = w_0 + Re^{1/2}$.

Maximization of the shareholders’ expected profit

$$E(\tilde{w} + e - w_0 - Re^{1/2}) = e - w_0 - Re^{1/2}$$

yields $e^* = 1/R$ (assuming $w_0 \leq 1/2R$). Suppose that effort is not observable, but profit is.

We will restrict ourselves to linear incentive schemes, so let $w(\Pi) = a + b\Pi$. Now determine the optimal scheme in this class. The manager’s expected utility is

$$Ew(a + be + be - Re^{1/2}).$$

Maximization with respect to $e$ yields $e = b/R$. Effort grows with the slope of the incentive scheme, and for $b = 1$ the manager is residual claimant and $e = e^*$. The manager’s expected utility is, thus,

$$E w(a + b^2/2R + be).$$

The shareholders’ expected net profit is

$$\Pi^* = E(w + e - a - be + be) = \frac{b}{R}(1 - b) - a.$$

To find the optimal linear incentive scheme, solve

$$\max_{(a,e)} \Pi^* = \frac{b}{R}(1 - b) - a$$

subject to

$$E w(a + \frac{b^2}{2R} + be) \geq w(w_0).$$

Substituting $a$ into the participation constraint (which is binding here) yields

$$Ew \left( - \Pi^* + \frac{b^2}{2R} + be \right) = w(w_0).$$

It is clear that if $\Pi^*$ is to be maximized, the shareholders must choose $b$ so as to maximize the left-hand side of equation 6. Hence, we have

$$(Ew) \frac{1}{R} - b + E(u'(e)) = 0.$$

If the manager is risk-neutral, $u'$ is a constant independent of $e$, and equation 7 yields $b = 1$. This confirms the principle of residual claimancy for risk-neutral agents. If $u'$ is strictly concave, we claim that $b$ lies between 0 and 1. Suppose that $b \leq 0$. Then the first term on the left-hand side of equation 7 is strictly positive. The second term, which is equal to the covariance of $u'$ and $e$ (recall that $Ee = 0$), is non-negative, so equation 7 cannot hold. The reason the covariance is non-negative is that for $b \leq 0$ the manager’s income is non-increasing in $e$, so that his marginal utility, which is a decreasing function of income, is non-decreasing in $e$. The reasoning for $b \geq 1$ is similar, but here the first term is negative, and the covariance is also negative.

We thus infer that the optimal linear wage structure is a profit-sharing scheme—a compromise between a fixed wage ($b = 0$), which yields optimal insurance, and residual claimancy ($b = 1$), which yields optimal incentives.

**Remark** In the above example, the manager was assumed to be rewarded on the basis of profit. In practice, managerial compensation is contingent on the value of the firm as well as on its current profits. Lewellen (1977) documents that a firm’s stock options are often a large portion of its managerial portfolios. The general idea behind rewarding managers on the basis of stock value rather than profits is that profits are a very gamble measure of managerial performance (Lewellen 1968, chapter 4; Grossman and Hart 1980, p. 48). For instance, a profitable investment reduces current profits without reflecting managerial slack or ineptitude. But such factors, which are unverifiable because of counting manipulations, may be observable.
by the market and thus may be reflected in the firm's valuation. Stock options in particular are seen as an incentive for an otherwise transient manager to care about the firm's future profits as well as its current profits.\textsuperscript{76}

2.2 Limits to Discretion

In the preceding subsection we considered the use of performance measures to limit managerial discretion. In practice, the shareholders may want to use other pieces of information as well. We will look at other factors that restrict managerial discretion even further.

Yardstick Competition

An agent's individual performance, even if it is verifiable, is only a garbled measure of the agent's effort (see examples 1 and 2 above). For instance, a firm's low profit may be due to a decrease in demand or an increase in costs rather than to managerial slack. Such effects can be detected to some extent, by comparing the agent's performance with that of other agents placed in similar conditions.\textsuperscript{77}

To see how yardstick competition works, consider example 1 above. Suppose that the shareholders oversee two managers in charge of two similar divisions. The shareholders' profits are equal to the sum of the profits generated by each manager, net of the expected wage bill. As before, the probability of generating profit \( \Pi_x \) rather than \( \Pi_y \) is \( x \) or \( y \) depending on whether the manager works or not. Furthermore, the uncertainties facing the managers are perfectly correlated, in that the same level of effort yields the same profit. Thus, if both managers choose to work, the realized profit is either \( \Pi_x \) or \( \Pi_y \), for both (with probability \( x \)) or \( \Pi_x \) or \( \Pi_y \), and similarly when they both choose not to work.

One may have in mind, for instance, the case of two divisions serving two distinct geographical markets whose demands are perfectly correlated.

In these circumstances, the shareholders can use the following contract: "if both managers reach the same level of profit (be it \( \Pi_1 \) or \( \Pi_2 \)), both receive the full-information wage \( w_1 = w_2 = w_0 + \Phi \) if profits differ, the high-profit manager gets \( w_1 = w_0 + \Phi \) and the low-profit manager gets heavily punished." Each manager's wage thus depends on the other manager's performance as well as on his own. Clearly, both managers exerting effort is an equilibrium of the subsequent game between the managers. If a manager is expected to work, and yields the high profit, the other manager automatically reveals that he did not work by obtaining the low profit. He cannot attribute his poor performance to "adverse circumstances," and he is heavily fined.\textsuperscript{78}

Exercise 2***. In this exercise, which concerns sole sourcing versus dual sourcing, we build a simple model in which the managers' objective function is such that monetary incentives are rather ineffective. A firm has a project of a given size. The cost of the project is \( C = \beta - \epsilon \). The variable \( \beta \) is random on \( [\beta, \beta'] \) with expectation \( E \beta \). The variable \( \epsilon \) denotes the effort exerted by the manager assigned to the project. A manager has utility function \( U(\omega, \epsilon) = u(\omega) - \Phi(\epsilon) \), where \( \Phi' > 0, \Phi'' > 0, \Phi'(0) > 0 \), and

\[
\begin{align*}
\omega(\omega) &= \begin{cases} 
-\infty & \text{if } \omega < \bar{\omega} \\
\bar{\omega} + \lambda(\omega - \bar{\omega}) & \text{if } \omega \geq \bar{\omega}.
\end{cases}
\end{align*}
\]

Thus, \( \bar{\omega} \) can be interpreted as a subsistence wage. \( \lambda \) is a "small," positive parameter; mathematically, \( \lambda \leq \Phi(\epsilon) \) for all \( \epsilon \). The manager is infinitely risk-averse over the states of nature \( \beta \), so he is only interested in his utility in the worst state of nature: \( \min \{ U(\omega, \epsilon) \} \). The manager learns \( \beta \) after signing the contract and before choosing \( \epsilon \). The principal observes \( \omega \), but not \( \beta \) or \( \epsilon \). So the wage structure is a function of \( \omega(C) \), and the manager's objective can be written

\[
\min_{\epsilon} \left( \max_{\omega} \{ U(\omega(\beta - \epsilon)) - \Phi(\epsilon) \} \right).
\]

Let \( U_0 \) denote the manager's reservation utility, and let

\textsuperscript{76} Of course, to preserve incentives it is important to prevent the manager from diversifying away the risks associated with the firm's performance. Stock options that cannot be sold serve this purpose.

\textsuperscript{77} The theory of yardstick competition and tournaments was developed by Lazear and Rosen (1981), Green and Stokey (1988), Nalebuff and Stiglitz (1983), and Shleifer (1988).

\textsuperscript{78} This suggests the possibility of multiple equilibria. Indeed, with our formulation, both not working is also an equilibrium. On the issue of multiplicity and incentive design to prevent it, see Block, Stiglitz, 1984.
$e^* > 0$ be defined by $u(w) - \Phi(e^*) = U_{c'}$. The principal wants to minimize the expected cost of the project.

(i) Show that if $\beta$ and/or $e$ were observable by the principal, the optimal contract would yield $w = w^*$ and $e = e^*$ for all $\beta$, and that the expected cost of the project is $\overline{w} + E\beta - e^*$.

(ii) Under asymmetric information, show that the optimal contract is

\[
\begin{cases} 
\overline{w} & \text{if } C_i \leq \overline{\beta} - e^* \\
< \overline{w} & \text{otherwise}
\end{cases}
\]

and that the expected cost of the project is $\overline{w} + \overline{\beta} - e^*$.

(iii) Suppose that the project can be given to two managers. The cost to the principal (net of the wage bill) is $\min(C_i, C_j)$, where $C_i = \beta - e_i$ and $e_i$ is manager $i$’s effort $(i = 1, 2)$. That is, $\beta$ is the same for both managers. Show that an optimal contract is

\[
\begin{cases} 
\overline{w} & \text{if } C_i = C_j \\
\overline{w} + \Phi(e^*)(C_i - C_j)/\lambda & \text{if } C_i < C_j
\end{cases}
\]

It is concluded that the principal prefers dual sourcing to sole sourcing if and only if $\overline{w} \leq \overline{\beta} - E\beta$. Interpret.

**Remark** The preceding example and the exercise assumed perfect correlation between the agents’ technologies. But the idea of yardstick competition carries over to environments with imperfect correlation (a more reasonable assumption). Indeed, Baiman and Dienski (1980) and Holmström (1982) have used the Holmström-Shavell sufficient-statistics result (see section 3) to show that an agent’s optimal wage structure depends only on his performance if and only if performances are independent.

**Remark** Yardstick competition is somewhat analogous to the use of authority to reward agents when their performance is observable but not verifiable, in that both rest on a comparison of the agents’ performances. However, the two arguments differ in spirit. Yardstick competition relies on the correlation of the agents’ technologies, but not on the nonverifiability of performances; furthermore, comparison can be made with outside parties, such as competing firms. With authority the result stems from nonverifiability, and does not rely on the correlation of the technologies, and the comparison is made within a group of agents overseen by a principal.

The potential applications of yardstick competition are numerous. The performances of managers of divisions facing similar cost or demand conditions may be compared by chief executive officers. Similarly, the rewards of one company’s managers can be made contingent on the performance of a competitor’s managers. More generally, managerial rewards can be based on average industry profit. The Department of Defense and many private firms sometimes use dual sourcing to procure their supplies, even in spite of the possible loss in returns to scale. Medicare pays hospitals a fixed fee for treating all patients within a diagnostically related group. The size of this fee is based on the average cost of treating patients in this group at comparable hospitals (Shleifer 1985).

Yardstick competition also has its limits. The units that are to be compared may face different conditions (e.g., the correlation between the situations may be fairly imperfect). Furthermore, their performance may be garbled by accounting idiosyncrasies or measurement errors. Last, the managers’ performance depends on the assets they have inherited. (Although it may not be a problem in theory, this effect requires yardstick competition to be more sophisticated and, therefore, makes it more unlikely.) This may explain why there is little yardstick competition in the electricity-generation industry (see Jokow and Schmalensee 1986).79

**Takeovers**

Manne (1965) and Marris (1964) have argued that failure to maximize profits lowers the stock value of a firm, and that it induces outside entrepreneurs (raiders) to buy the firm, replace its management, and direct the firm toward profit maximization. The threat of such takeovers serves to discipline managers.

But why should managers be concerned with the threat of a takeover? For the argument to hold, managers must suffer greatly from the takeover. This could be true either because managers are immediately punished when their

---

79. See Astle and Smith 1986 for an empirical assessment of the use of yardstick competition as executive compensation.
firm is taken over (since a takeover is an indicator of poor management) or because, by being removed, they lose the rents they had enjoyed within the firm. The first reason does not seem convincing because of the limited liability and a prohibition on slavery, direct punishments are hard to impose. Actually, far from being punished, managers usually receive princely amounts of money ("golden parachutes") when they are fired after a raid. The threat of losing rents attached to managerial jobs is a more credible explanation. One such rent is prestige or reputation. Another may be on-the-job leisure (slack) due to the asymmetry of information between shareholders and managers. To the extent that mismanagement increases the likelihood of a takeover, the fear of losing their rents may make managers less prone to slack. Formulations of this idea have been put forth by Scharfstein (1985a) and by Denski, Sappington, and Spiller (1987).

Takeovers, however, have their limits. Costly information must be collected about the firm's inefficiencies and areas for improvement. Outsiders have an incentive to collect information and spend the takeover costs only if they can derive substantial profits from the takeover. Grossman and Hart (1986) point out a potential free-rider problem that may impede this incentive. In the event of a takeover, a shareholder may not want to tender his shares because he knows that shareholders can enjoy the increase in the stock price brought about by the raid. On the other hand, the raider can make a profit only if the tender price of shares is lower than the post-raid price. Hence, he cannot both buy the shares and make a profit on them.

There are ways out of the free-rider problem. One is a provision in the firm's constitution allowing a successful raid to sell part of the firm's assets to another company owned by the raider at advantageous terms to minority shareholders or to issue new shares—issuance of new shares—issuance to a reward to the raider, and encourages takeovers (Grossman and Hart 1986). Another possibility is that the raid may be undertaken by a large shareholder of the firm. Even if the other shareholders free-rider on him (by not tendering their shares), a large shareholder at least enjoys the increase in the value of his own shares (Shleifer and Vishny 1986).

These anti-free-rider factors are themselves limited. On the one hand, because dilution is basically a gift to the raider, incumbent shareholders may be reluctant to allow it on a large scale. Furthermore, the U.S. courts have restrained its use. On the other hand, the large shareholders internalize only the increase in the value of their shares. They do not take the positive externality on the other incumbent shareholders into account, so their incentive to monitor the firm and to undertake a takeover may be too small.

Potential resistance by the current management to a raid imposes a second limit on the efficacy of takeovers. Managers can make the firm unattractive to the raider through antitrust litigation or "poison pills." If this does not work, they can collude with the raider and buy his shares of the firm (if any) at a substantial premium over the market price in exchange for his signing a standstill agreement, which prohibits him from owning shares of the firm for a certain period of time. The other shareholders may well be losers in such a "greenmail" maneuver, because the takeover does not take place (management is not removed) and the firm purchases the raider's shares at a high price (however, see Shleifer and Vishny 1984). Finally, managerial resistance can be reduced (for instance, by offering "golden parachutes" to removed managers), but only at a substantial cost.

These effects may, to some extent, explain Scherer's (1980, p. 38) observation (made before the recent wave of takeovers) that "the available evidence provides at best only weak support for the hypothesis that takeovers generate an effective disciplinary mechanism against departures from profit maximization."

A takeover threat may have perverse incentive effects. First, it lowers the managers' incentive to make long-run investments, as they may not be around to reap the benefits. That is, managers are led to behave myopically.

80. In some instances the takeover may hurt the managers' future careers by conveying bad news about their abilities. The sanction is then imposed by the market for managerial talent and is delayed. (For models of managerial careers, see the next subsection.)

81. Denski et al. present their formulation in the related context of second sourcing (i.e., the replacement of a supplier by another). See also Antin and Yao 1987 and Callaud 1988.

82. For instance, they may purchase one of the raider's competitors, making the raider's takeover subject to antitrust laws. Or by purchasing another firm (even one not related to the raider's activities), they may increase their own firm's debt enough to scare off a cash constrained raider. "Poison pills" are preferred stock rights that are inactive unless they are triggered by a tender offer for a large fraction of the firm. They are somewhat similar to an entry fee to be paid by the raider.
(Laffont and Tirole 1987). Second, it destroys the managers’ job stability and exacerbates their career concerns, which may lead to managerial decisions that are contrary to the firm’s interest (Hermalin 1987). Third, it shortens the term of the relationship between managers and workers, and may prevent the development of trust between them (Sheifer and Summers 1987).

Managerial Incentives: The Dynamic Perspective

Another bound on the managers’ discretion is the concern for their own careers, both within the firm and outside the firm. The arguments presented here relate mainly to the interaction between adverse selection (How efficient or trustworthy is the manager?) and moral hazard (How hard does he work?).

Within a firm, a manager who has performed poorly may not be trusted to do well in the future. His prospects for internal promotion may be weak. To formalize this, some authors have focused on the repeated aspects of the relationship between a firm and its employees. Suppose that in each period the employee’s performance is of a high or a low quality. Quality is observable by the employer, but it is not verifiable. The firm offers a rent to its employees in each period (where a rent means that the employee strictly prefers to stay with the firm rather than quit—it may, for instance, represent a wage above the market wage), as long as the employee has performed with high quality before. If the employee “cheats” and turns in a low-quality performance, the firm stops offering him the rent (for instance, does not promote him or does not increase his salary). The threat of losing this rent puts some discipline on the employee. Mathematically, the formalization of this idea involves either the theory of supergames or the theory of reputation under asymmetric information (see chapters 2, 6, and 9). The idea is similar to the one in chapter 2 where consumers stop patronizing a firm that starts producing low-quality products.83

Let us now turn to the possibility that the manager will leave the firm. The threat of outside offers may discipline the firm and force it to treat its managers fairly when the quality of their work is observable but not verifiable. Suppose that a manager turns in a high-quality job, which is a signal of the manager’s ability as well as of his effort, but is not rewarded by a wage increase. If his performance is observed by other firms, the latter may be able to bid him away. Because of this threat, the firm must reward the manager fairly. Consider a professor at a university. Let the quality of his work be the quality of his research and the richness of his interactions with colleagues and students. Now, this quality may be observable by the profession but hard for a court to assess.84 His threat to move to another university tends to equalize his position within and outside his university. This idea, pioneered by Fama (1980), was studied further by Holmström (1982b), who combined the distinction between observability and verifiability and the outside pressure mechanism.

The possibility of getting good opportunities outside his present firm, like that of being rewarded within the firm, certainly gives a manager incentives to perform satisfactorily. When the manager’s internal or external supervisors observe only his performance, any amount of shirking on his part may be mistaken for a lack of trustworthiness or ability, and thus may hurt his career. Holmström shows that, in such circumstances, a manager at the beginning of his career may work even harder than is socially optimal.85

Remark Two other mechanisms that enable a firm to reward its employees’ observable-but-unverifiable performance by fiat, but fairly, have already been described. One was based on the firm’s commitment to reward a given percentage of its employees; the other was based on the firm’s reputation. The outside pressure mechanism may work better for top managers than for other employees, because the former have more outside visibility than the latter.

83. For elaborations of this idea in the context of internal organization or business relations, see Tamer 1990; Kreps 1984; Baj 1983; Hart and Holmström 1987; and MacLeod and Makornitz 1986. In the context of business relations, Kreps posed the further question of why transient managers of a firm would want to sustain the firm’s reputation. Cheating on an explicit agreement yields a current benefit, whereas some of the future losses associated with the loss of reputation will be incurred after the managers have left the firm. Kreps notes that if the managers own the firm, they internalize the full effects of their decisions, because the more of the firm reflects future as well as current profits.

84. Of course, the courts might hire experts in the profession to assess the quality (which they sometimes do). But the legal costs would generally be incommensurate with the issue at stake.

85. Gibbons (1983) shows that it may be optimal to reduce the manager’s outside visibility at the beginning of his career in order to reduce this effect. He argues that the lawyers members of a law firm often have such a reduced visibility. Wolfinson (1985) demonstrates that a reputation effect alleviates the moral hazard problem somewhat in oil drilling ventures.

The Theory of the Firm
Supervision

Except in the discussion of takeovers, we have mainly taken the principal's information as given and studied the optimal use of this information to control the agent. We now turn to the internal incentive to monitor.

Alchian and Demsetz (1972) argue that nonseparabilities or increasing returns to scale are essential to the understanding of organization design. The product of a team exceeds the sum of the products of its members in isolation; however, Alchian and Demsetz note, team production may hinder the metering of productivity and rewards. For instance, unless each worker is accountable for a particular part of production (as is the case with piecework), accounting data measure only the output of the shop. In one of their examples, Alchian and Demsetz show that it may be hard to disentangle the relative performances of two workers who jointly lift heavy cargo into trucks. At a more aggregate level, a high level of sales of a product may be due to a good design, to the high quality of the product, or to an appropriate marketing campaign, and there may be no clean measure of each functional division's contribution.

Nonseparabilities create a free-rider problem among the members of a team. Suppose, for instance, that two employees in a team share equally every extra dollar the team creates. Then each employee, when generating $1 for the team, receives only 50c. This means that each employee has too small an incentive to contribute to the team's production. One solution is to increase the monetary incentives further (see the remark below). The other solution is the introduction of a third party (a monitor or a supervisor) to measure the individual performance of each employee. How is the monitor monitored? What are his incentives to supervise the employees? Alchian and Demsetz suggest that the monitor be given a title to the net earnings of the team (net of payments to other inputs). In the jargon of incentives theory, the supervisor is made the residual claimant or sink. At the margin, he captures any extra profit of the team. He thus has a strong incentive to exert effort to measure the employees' individual performances.

In summary: Alchian and Demsetz's organization (or "firm") is a particular policing device utilized in the presence of team production. Indeed, among the bundle of rights associated with ownership of a firm, Alchian and Demsetz list the right to be residual claimant and to observe the behavior of employees.86

Remark Holmström (1982b) offers an alternative theory of the firm based on team production and the impossibility of measuring individual performance. He argues that one of a corporation's roles is to break the rule that the benefits of an organization must be split among employees (the way it is done in a partnership). The idea is as follows: To face the right incentives in his production decisions, an employee must be the "residual claimant" for these decisions; if a decision yields one extra dollar to the firm or unit, the employee must receive that extra dollar. Consider a two-employee unit, and suppose that only the unit's performance (equal to the sum of the two employees' performances) is observable. Then, each time the unit's profit grows by $1, each employee must be given $1. The marginal distribution of profits must be $2 for each extra dollar earned by the unit. This is possible only if there exists a "source" (shareholders? other units?) that "breaks the budget constraint" at the margin.87 The Holmström and Alchian-Demsetz theories differ except for the starting point (the impossibility of disentangling individual contributions at an accounting level) and for the conclusion that the budget constraint of a team must be broken. Alchian and Demsetz rely on supervision by a third party who is made a sink in order to give him the incentive to monitor (at the margin, the employees receive $0 for every $1 they create, i.e., they have a fixed wage). In contrast, Holmström disciplines team members through monetary incentives provided by a marginal source.

Williamson (1975, p. 49) argues that the importance of nonseparabilities in production should not be overstressed. Because a monitor can supervise only a limited number of employees, the theory can at best explain the organiza-

86. Other listed rights are those to be the central party common to all contracts with inputs, to alter the membership of the team, and to sell these rights.

87. For an argument that follows Holmström's see exercise 4.4. Note that the use of a source requires that the employers be unable to collude against the source. Because every time an employee increases output by $1 the other employee also receives $1, the employees have an incentive to get together to expand production beyond its noncooperative level.

45
tion of small groups. In large firms, supervision must be delegated.

This brings us to an interesting class of hierarchical models which were pioneered by Williamson (1967) and by Calvo and Wellisz (1978, 1979) and developed further by Rosen (1982) and by Keren and Levhari (1983). Suppose that the firm is organized according to a pyramidal structure, as shown in figure 4. Level 3 is the productive tier of the firm (the workers). Level-1 employees are supervised by level-2 employees, presumably because of nonseparabilities in production. A level-2 employee does not supervise all level-1 employees, because the quality of supervision decreases with the span of supervision. In turn, level-2 employees' incentives to supervise are provided by level-3 supervision. Level-3 is assumed to be composed by a single agent (or unit), who is the residual claimant for the firm's profit net of wage and input payments. For instance, level 3 can be the shareholders (respectively, the executive officers) and level 2 the executive officers (respectively, the division officers). An attractive property of these models is that the horizontal and vertical sizes of the firm are not fixed. The span of control of each employee and the number of tiers are supposed to be chosen by the top tier to maximize profits. What may then put a lid on the firm's size is the deterioration of supervision as the firm grows. More workers require more level-2 supervisors, and the supervision of level-2 employees by the top tier worsens. (Alternatively, an extra layer can be created between the top tier and level 2, but this addition is costly.) Calvo and Wellisz also derive some interesting implications for the wage structure within the firm from this type of model.

In Tirole 1986b it is argued that the exercise of supervision and authority in an organization is limited by the possibility of collusion between groups of its members. An employee obtains power over another when given the right to evaluate the other's performance (supervision) or to make decisions in unforeseen contingencies that affect him (authority). This power gives rise to the possibility of collusion between the two parties, which is enforced by side-transfers (promise of a similar counter favor, money, evolution of their personal relationship, etc.). Organizational design is then partly geared toward preventing collusion by recognizing a supervisor's incentive to act as an advocate for rather than a prosecutor of the supervisee, and by limiting authority through the imposition of bureaucratic rules.

Product-Market Competition

As Scherer (1980, p. 38) notes, when forced into the trenches on the question of whether firms maximize profits, economists resort to the ultimate weapon in their arsenal: a variant of Darwin's natural selection theory. Over the long pull, there is one simple criterion for the survival of a business enterprise: Profits must be nonnegative. No matter how strongly managers prefer to pursue other objectives, and no matter how difficult it is to find profit-maximizing strategies in a world of uncertainty and high information costs, failure to satisfy this criterion means ultimately that a firm will disappear from the economic scene.

For instance, Winter (1971) suggests that a competitive firm that makes inefficient decisions (about techniques, for instance) incurs losses because it cannot simply transmit its extra costs to the consumers (the market price is taken as given). The firm is thus led to search for new and better decisions in order to survive. Hence, firms in a competitive environment are more hard-pressed to reduce costs, and end up being more efficient. This is clearly a rather realistic story; however, as Hart (1983) notes, "a question which Winter does not analyze, at least formally, is why firms choose inefficient techniques in the first place." The

---

88. Indeed, most of the supervisors in organizations have only limited stakes in them. In particular, they are usually far from being the residual claimants for the teams they supervise.
answer to this question may lie in the problems of delegation and monitoring mentioned above.\textsuperscript{99}

The effects of the threat of bankruptcy and of competitive pressure on managerial incentives have not yet been formulated in a satisfactory way. However, some of the effects of competitive pressure on incentives (not those related to the survival issue) have been studied.

One obvious effect stems from the possibility of yardstick competition. The shareholders of a competitive firm can base managerial rewards on the competitors’ profits or on the market price, which would not be possible were the firm a monopoly.

The effect of product-market competition on internal incentives becomes more subtle when outside data, such as competitors’ profits and the market price, are not available to the firm. Hart (1983) has shown that the form of competition in the product market still has an influence on internal control when the owners of the firm observe only that firm’s performance.\textsuperscript{90}

In Hart’s model, competition operates through the variability of the market price, which, together with a cost shock, determines the variability of the firm’s profit and therefore determines the extent to which the firm’s managers can manipulate the owners’ uncertainty in order to engage in slack. Roughly, a profit that is exogenously more variable leaves more leeway for misrepresentation by the managers and less opportunity for control by the shareholders. Hart considers a competitive industry with two types of firms: managerial firms (in which the shareholders delegate the decision power to managers and are therefore exposed to the above-mentioned control hazards) and entrepreneurial firms (which are run by the entrepreneurs themselves).\textsuperscript{91}

Intuitively, when the marginal cost of production (which is perfectly correlated across the firms) is low, entrepreneurial firms expand their output. The managers of the managerial firm, if they do not respond much to monetary incentives, take advantage of the good times to slack. If the proportion of entrepreneurial firms increases, the output is thus more sensitive to a cost decrease. This reduces the market price in good times, thus mitigating the influence of the low cost on profit. In particular, the managerial firms’ profits become less sensitive to outside uncertainty when the proportion of entrepreneurs grows. This makes their control by shareholders easier, which results in less slack. To the extent that entrepreneurs are a symbol of competition,\textsuperscript{92} more competition in the output market leads to less slack in managerial firms. Unfortunately, this result is sensitive to the description of the managers’ utility function, as Scharfstein (1985b) shows. If managers do react to monetary incentives sufficiently, a larger proportion of entrepreneurs increases slack in managerial firms.\textsuperscript{93}

An Application: The M-Form Firm

The analysis of managerial incentives may shed some light on the emergence of the so-called multidivisional-form (M-form) firm.\textsuperscript{94}

Recall from section 1 that the technologically rational organization of a firm is the unitary form, with its specialization by function. However, this type of organization collapsed with the horizontal expansion of firms. According to Chandler (1966), this collapse is mainly due to the loss of control by the top management. This might be explained as follows. To control the functional divisions,
the top management can basically use one of two methods: rewarding each functional division for good performance (i.e., basing incentive schemes on output) and supervising the divisions directly in order to assess individual contributions (i.e., by measuring inputs). The first method clearly faces the accounting problem of separating the contributions of the various divisions. The sales of a product or the profits of the firm depend on the quality of each division's performance, which may be hard to measure. This gives rise to an Alchian-Demsetz-type team problem. The second method can be employed only if the firm is small. A loss in the supervisory possibilities of top management may have accounted for the difficulties of the unitary-form firm in an expansionary phase.

The multidivisional form emerged in the 1920s and became prevalent after World War II. It consists of organizing the firm into quasi-firms resembling "scaled down, specialized U-form structures" (Williamson 1975, p. 136). These divisions are defined by product, brand, or geographic lines, and are fairly autonomous. See, for example, figure 5.

In the M-form firm, a relatively precise measure of divisional performance is available. And, indeed, the role of the general office (top management) is to audit and allocate resources among the competing divisions. Within a division, by contrast, the supervisory mode is more prevalent, and that allows some assessment of the relative contributions of functional subdivisions.

This rough analysis leaves an important question unanswered: If divisions are quasi-firms, why should they be organized within a single structure? Why aren't they separate legal entities? Indeed, Williamson views the M-form firm as a "miniature capital market"; however, he argues further that the general office has better auditing capabilities, as well as better takeover capabilities, than the capital market (1975, pp. 146–148). His argument is related to incomplete contracts.

2.3 Doubts about the Neoclassical Methodology

Progress made since the early 1970s has shown that the neoclassical firm is not the unrealistic, profit-maximizing entity that organization theorists ridiculed in the 1950s and the 1960s. Still, the neoclassical theory leaves many questions unanswered, and this raises some doubts about its ability to cope with certain complex organizational phenomena. Simon 1970 and chapters 3 and 5 of Nelson and Winter 1982 are particularly instructive in this respect. Among the neglected topics are the following.

Optimizing Behavior

The neoclassical theory presumes that members of an organization act so as to maximize an objective function of a few standard variables (such as income and effort). This raises two questions: Does their objective include other variables (prestige, ego, power, number of subordinates, size of the budget, conviviality, friendship, etc.)? Do members even have a well-defined objective function, or do they use rules of thumb or "satisfice"?

Concerning the first issue (arguments of the objective function), we must recognize that a number of goals other than income and effort seem to motivate the members' energy. Of course, it is interesting to wonder whether these enter directly into the member's utility function or whether they are intermediate objectives that help achieve the primitive goals. As was mentioned earlier, it might be the case that a manager enjoys a large number
of subordinates not per se but rather because this acts as a signal to the labor market (a manager in charge of an important division is likely to be an able manager) or because it allows for smoother operation and thus reduces on-the-job pressure (effort). The economist faces a familiar dilemma here: An increase in the number of explanatory variables (arguments of the objective function) makes it easier to explain real-world phenomena. At the same time, the theory loses predictive power: By adding enough arguments, one can always "explain" any kind of behavior. By contrast, the restriction to a small number of predetermined primitive variables restores some discipline. How "inspired" the neoclassical economists have been in their choices of primitive variables is not yet known.

The second issue—the possibility that economic agents do not optimize—is, of course, a matter of concern. Indeed, members of an organization often use rules of thumb instead of performing complex computations. However, many behaviors that look nonoptimizing may actually be the outcomes of optimization under constraints, and therefore may not be at all irrational. For instance, a member generally does not have time to sample all the information that is relevant to decision making; he may thus make decisions that, viewed with better information, look irrational. However, the decisions may be rational, given the shadow price of the time necessary to gather the information. Similarly, the time and effort required to compute optimal decisions in complex problems are traded off against inefficiencies in decision making. It remains to be seen how successful such "rational explanations" of bounded rationality will be.96

Communication and Knowledge

Neoclassical theory pays only lip service to the issue of communication. Information flows between members of an organization are limited only because of incentive compatibility. A member keeps his information private when he would suffer from its becoming public. Private incentives certainly limit information flows considerably. However, even well-intentioned members of an organization (i.e., members who do not manipulate information to their advantage) may have trouble communicating all the information they possess to their relevant co-members, because it is too time consuming or because the information is hard to "codify" (see Arrow 1974) to make it understandable to its receivers. Thus, decisions that would be profit maximizing under full communication will not be made under imperfect communication.97

This brings us to the notion of knowledge. Several contributions (among the most recent Nelson and Winter 1982, Kreps 1984, and Cremer 1986) have endowed a firm with a stock of knowledge that guides the organization's members in their decisions and coordination in a world of imperfect communication:

The context of the information possessed by an individual member is established by the information possessed by all other members. To view organization memory as reducible to individual member memories is to overlook, or understate, the linking of those individual memories by shared experiences in the past, experiences that have established the extremely detailed and specific communication system that underlies routine performance. (Nelson and Winter 1982, p. 105)

Organizational culture is the pattern of basic assumptions that a given group has learned, discovered or developed in learning how to cope with its problems of external adaptation and internal integration, and that have worked well enough to be considered valid, and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to these problems. (Schein 1984, p. 3)

Organizational memory helps the members to find relatively satisfactory decisions in the presence of complex decision making ("satisficing" or individual bounded rationality) and to coordinate their respective actions in the absence of perfect communication (collective bounded rationality).

Dynamics of Organizations

Neoclassical theory has focused on the optimal design of an organization at a given point of time. Little has been said about reorganizations. Because most reorganizations

---

96. An important area where this information has been fairly successfully in search theory, which explicitlyformulates the trade-off between costly acquisition of information and insufficient decision making.

97. Neoclassical theory has recently made some progress in the formalization of these issues (Chomsky and Milgrom 1984) and their consequences (Sah and Stiglitz 1983), but much work remains to be done.
are, in general, not specified in the initial organizational
design, they are much influenced by the authority and
bargaining relationships at the time of the reorganization.
As the formalization of these relationships is still in its
infancy, even in the static neoclassical theory of the firm,
it is not surprising that neoclassical theory has had little
to say about reorganizations.

Group Behavior

Sociologists (e.g., Dalton [1959] and Crozier [1967]) and
organization theorists (e.g., Cyert and March [1963]) have
emphasized that organizational behavior is often best
predicted by the analysis of group incentives as well as
individual incentives. In contrast, neoclassical theory has
tended to focus only on individual incentives.98

The reader will have perceived in this discussion of the
current lacunae of the neoclassical approach much
intellectual imperialism. Most of it conveys the following
message: "Sociologists and organization theorists are right
to claim that neoclassical theory has not tackled some
important aspects of organizations. But give the theory
enough time to develop and resolve adequate models." Of
this, of course, is a pure act of faith, and much of the
controversy may revolve around whether the neoclassical
approach is indeed able to cope with these aspects.

2.4 The Profit-Maximization Hypothesis and
Industrial Organization

As has been noted, there are many ways in which discre-
tion can be curbed. However, none of them is perfect, and
we should expect some possibly important deviations
from profit-maximizing behavior.99 Does this mean that
the profit-maximizing models of this book are funda-
mentally flawed? Not necessarily. Consider the familiar
problem of monopoly pricing (see chapter 1). Let a mo-

opoly have the profit function

\[ \Pi = P(q)q - c(q)q - w, \]

where \( q \) is the firm's output, \( P(q) \) is the inverse demand
function, \( w \) is the manager's wage, and \( c(q) \) is the unit cost
(a function of the manager's effort \( e \) and some random
variable \( z \)). Suppose that the shareholders observe every
variable except \( e \) and \( z \). In particular, they observe the
realization of the unit cost \( c \). This unit cost plays the same
role that the profit variable plays in examples 1 and 2 of
subsection 2.1. It can easily be shown that the optimal
wage structure, \( w(e) \), is based solely on the realization of
because \( c \) conveys all the relevant information about
effort (mathematically, it is a sufficient statistic for effort).
Now, we know from subsection 2.1 that if the manager is
risk-averse, the effort \( e \) induced by the optimal wage
structure differs from \( e^* \), the optimal level that would be
obtained if effort were observable by the shareholders.
The firm is profit-maximizing only in a constrained sense,
and the manager engages in X-inefficiency \( (\varepsilon \neq e^*) \) in the
sense of Leibenstein 1966.

However, given the effort \( e \) induced by the optimal
wage structure, \( w(e) \), let \( \varepsilon \equiv \varepsilon(e) \) denote the resulting
random cost. The shareholders' expected profit is

\[ E\Pi = P(q)q - (E\varepsilon(q) - Ew(e)). \]

Because the choice of \( q \) does not change the managerial-
control problem (recall that \( \varepsilon \) is a sufficient statistic), the
shareholders (or the manager) might as well pick \( q \) so as
to maximize \( P(q)q - (E\varepsilon(q) - Ew(e)) \). Hence, for an outside
observer, the firm's behavior is observationally equivalent to
that of a firm that does not engage in X-inefficiency, but
has a cost distribution \( \varepsilon \) for the efficient (full-information)
level of effort. So never mind that the distribution of the
unit cost is biased toward high levels because of informa-
tional asymmetries: the monopoly-pricing model devel-
oped in chapter 1 is still valid.

This is only an example; it is not meant to convey the
impression that this "separability" between internal
organization and product-market or input-market deci-
sions is the rule.100 Indeed, one of the most exciting
research agendas in industrial organization for years to

press frequently reports the internal policy struggles of large firms in a manner
that clearly involves informal use of a coalition model, there is little scholarly
literature in economics that takes this perspective. The proposals of March
[1962] and Cyert and March [1963] have been largely ignored."

99. Scharf (1980, p. 43) concludes: "...assuming profit maximization provides
e a good first approximation in describing business behavior. Deviations, both
intentional and inadvertent, undoubtedly exist in abundance, but they are kept
within narrow or less narrow bounds by competitive forces, the self-interest
of stock-owning management, and the threat of managerial displacement by
important outside stockholders and takeover raiders."

100. For a counterexample, see the long-nurse story in chapter 9.
come is the determination of the scope and importance of such interactions. However, this author feels and hopes that many of the conclusions of the theory of industrial organization will remain valid (at least at a descriptive level) when the profit-maximisation postulate is abandoned for a full-bladed model of internal organization.

3 Supplementary Section: The Principal-Agent Relationship

The purpose of this section is to introduce the moral-hazard issue in a more general framework than the one in the text.

Suppose that shareholders are risk-neutral. Their objective function is equal to the expected gross profit of the firm minus the expected wage payment. For ease of exposition, assume that there is a unique manager. This manager makes an unobservable decision \( e \) in an interval \([e,7]\). This decision will be interpreted as an effort level, but more generally it could be any discretionary or moral-hazard variable (perquisites, care, etc.). It is not observable by the shareholders. Given \( e \), the realization of the profit depends on the realization of a random variable \( \Pi(e, \varepsilon) \).

Presumably, \( \Pi \) increases with \( e \). The shareholders observe only the profit level and reward the manager according to a wage function of the only observable variable: \( w(\Pi) \).

Thus, the shareholders’ objective function is

\[
E_e [\Pi(e, \varepsilon) - w(\Pi(e, \varepsilon))].
\]

The manager’s objective function is the expectation of his utility. The latter depends on the monetary reward and on the level of effort exerted: \( U(e, \varepsilon) \). Presumably, \( U \) increases with \( w \) and decreases with \( e \). Assume also that \( U \) is concave in \( w \) (income risk aversion). Thus, the manager’s objective function is

\[
E [U(e, \varepsilon) - U(e, \varepsilon)].
\]

(Hereafter, all expectations are taken with respect to \( \varepsilon \).

In the traditional principal-agent framework, the shareholders design a wage contract \( w(e) \). There exists ex ante a competitive supply of identical managers, with some reservation utility \( U_0 \), the expected utility that they would obtain by working somewhere else. The shareholders are able to fill the managerial position only if the highest possible expected utility for a manager over all potential levels of effort exceeds \( U_0 \). In the jargon of incentive theory, the manager’s “individual rationality” or “participation” constraint.

101 This supplementary section is partly inspired by part 1 of Hart and Holmstrom 1987.
max \ E L(w(\Pi(e, \epsilon), \epsilon)) \geq U_0, \tag{8}

must be satisfied. Next, if the shareholders want to induce a given level of effort \( \epsilon^* \) from the manager, they must design a wage structure that is "incentive compatible":

\[ \epsilon^* \] maximizes \( \mathbb{E} L(w(\Pi(e, \epsilon), \epsilon)) \) over all \( e \). \tag{9}

Now consider the shareholders' problem. Choose a wage structure \( w^*(\epsilon) \) and induce a level of effort \( \epsilon^* \) for the manager that maximizes

\[ \mathbb{E}[\Pi(e, \epsilon) - w(\Pi(e, \epsilon))] \]

subject to constraints 8 and 9.

Solving this problem is, in general, a complex task.\(^{102}\)

There are two polar cases, however, in which the solution is straightforward.

Before exposing those cases, it is useful to derive the solution in the case in which \( e \) and \( \epsilon \) are observable (full information). Because the shareholders observe \( e \), they can impose any level they want consistent with the manager's participation (by threatening a large penalty if the manager does not conform). Thus, the only relevant constraint is 8. Now, from the theory of optimal insurance, for a given \( e \), the risk-neutral shareholders should give a constant wage \( w \) to the manager, i.e., should give him full insurance.\(^{103}\) The choice of the level of effort is in general, slightly more complex. Let us simply assume that the optimal effort under full information is not \( g \) (the lowest possible effort for the manager).

Let us now return to our asymmetric-information framework, in which the shareholders observe only the profit. Obviously, because the contract is contingent on fewer variables, the shareholders have less control and therefore can expect, at most, their expected full-information profit.

(The allocation they induce by using a small number of observables can always be duplicated with a larger one by dropping the extra observables.)

The first polar case is that of an income-risk-neutral manager. In this case,

\[ L(w, \epsilon) = w - \Phi(\epsilon). \]

The term \( \Phi(\epsilon) \) represents the disutility of effort expressed in terms of money. The shareholders' objective function can then be written as

\[ \mathbb{E}[\Pi(e, \epsilon) - w(\Pi(e, \epsilon)) - \Phi(\epsilon)] = [\mathbb{E} \Pi(e, \epsilon) - \Phi(\epsilon)] - U_0, \]

where use is made of the manager's risk neutrality and of the individual rationality constraint (which can easily be shown to be binding in this case). Let \( \epsilon^* \) optimize \( \mathbb{E} \Pi(e, \epsilon) - \Phi(\epsilon) \) (the expected profit net of the disutility of effort). By definition, \( \epsilon^* \) is the optimal effort under full information (because under full information only the individual rationality constraint matters). Thus, under full information the shareholders' net profit is

\[ \mathbb{E} \Pi(e^*, \epsilon) - \Phi(\epsilon^*) - U_0. \]

Under asymmetric information, suppose the shareholders offer to sell the firm to the manager at price \( p = \mathbb{E} \Pi(e^*, \epsilon) - \Phi(\epsilon^*) - U_0 \). If the manager accepts, he becomes the residual claimant for the firm's profit. The sale of the firm to the manager is formally equivalent to an incentive scheme in which the shareholders would remain the claimants for the firm's profit but would pay a wage \( w(\epsilon) = \Pi - p \). To see whether the manager accepts, look at his objective function:

\[ \max_{\epsilon} \{\mathbb{E} \Pi(e, \epsilon) - \Phi(\epsilon) - p\} = U_0. \]

So the manager accepts, and the shareholders make exactly the same expected profit as under full information.

As is explained above, the intuition for this result is that the manager faces an incentive scheme that exactly reflects the vertical structure's objective. Thus, he chooses the full-information level of effort. The potential drawback is that the manager may bear all the risk. That would not matter, however, because the manager is risk-neutral.

The other polar case is that of an infinitely income-risk-
averse manager. Such a manager prefers a random wage \( \hat{w}_t \) and effort \( e_t \) to a random wage \( \tilde{w}_t \) and effort \( e_t \) if and only if \( \min \hat{w}_t > \min \tilde{w}_t \) or \( \min \hat{w}_t = \min \tilde{w}_t \) and \( e_1 < e_2 \). That is, the manager is concerned primarily with his minimum wage, but in case of a tie he prefers the allocation with the lowest level of effort. Suppose further that the distribution of \( \Pi \) given \( e \) has support equal to an interval \([\Pi, \Pi + 1]\) whatever \( e \), so that the distribution of \( e \) but not its support moves with \( e \). Then, whatever his effort, the manager’s lowest possible wage is effort-independent, so he chooses \( e = \varepsilon \). In this case, no incentive can be given to the manager. He is given a constant wage, the value of which is chosen to make the individual rationality constraint binding given effort \( e \).

The case of a general managerial utility function \( U(w, e) \) is more complex. Let us simplify the model by assuming that the utility function is separable in income and effort: \( U(w, e) = u(w) - \Phi(e) \). Assume that \( u' > 0 \), \( u'' < 0 \), \( \Phi' > 0 \), \( \Phi'(0) = 0 \), and \( \Phi'(\infty) = \infty \). We will also use the “parametrized distribution formulation of uncertainty” (pioneered by Mirrlees [1974] and Holmström [1979]), according to which the cumulative distribution of \( \Pi \) given \( e \) is described by a cumulative distribution function \( F(\Pi, e) \) on \([\Pi, \Pi + 1]\), with density \( f(\Pi, e) > 0 \). These functions are assumed to be differentiable in effort. That effort increases profit (stochastically) is formalized by the first-order stochastic dominance relation on \([\Pi, \Pi + 1] \):

\[
e_1 > e_2 \Rightarrow F(\Pi, e_1) < F(\Pi, e_2)
\]

that is, the distribution for \( e_1 \) puts more weight on the upper tail. For a differentiable cumulative distribution function, this means that \( f(\Pi, e) < 0 \).

The manager’s optimization problem for a given incentive scheme \( u(\cdot) \):

\[
\max_e \left( \int_\Pi u(\mu(\Pi)) f(\Pi, e) d\Pi - \Phi(e) \right)
\]

yields the first-order condition

\[
\int_\Pi u(\mu(\Pi)) f(\Pi, e) d\Pi - \Phi'(e) = 0.
\]

Of course, this condition is not sufficient for optimality of effort; the second-order condition must also be satisfied if a maximum is to be obtained. For the moment, ignore the second-order condition. If the optimal solution for the shareholders found by ignoring the second-order condition for the manager can later be shown to satisfy the second-order condition, then it is truly optimal.

The manager must also be willing to participate:

\[
\int_\Pi u(\mu(\Pi)) f(\Pi, e) d\Pi - \Phi(e) \geq U_0.
\]

The “first-order approach” for the shareholders consists of finding a wage structure \( u(\cdot) \) and an effort level \( e \) that maximizes

\[
L = \int_\Pi \left( [U(\Pi) - u(\Pi)] f(\Pi, e) + \frac{\lambda}{\Pi} u(\Pi)(\Pi - U_0) f(\Pi, e) + \frac{\eta}{\Pi} (\Pi - U_0) u(\Pi) f(\Pi, e) \right) d\Pi
\]

where \( \lambda \) and \( \eta \) are positive (they are actually strictly positive, as can easily be shown). The derivative of \( L \) is taken with respect to \( e \) and that of the integrand with respect to \( u(\cdot) \) for all \( \Pi \). We will be concerned only with the second differentiation, which yields

\[
-\frac{1}{\mu'(\Pi)} + \lambda f(\Pi, e) u'(\Pi) + \frac{\eta}{\Pi} f(\Pi, e) u'(\Pi) = 0,
\]

or

\[
\frac{1}{\mu'(\Pi)} = \lambda + \frac{\eta}{\Pi} f(\Pi, e).
\]

(Equation 12 need only be true for almost every \( \Pi \), i.e., everywhere except possibly on a set of \( \Pi \) with measure zero.) It is left to the reader to compute the derivative of \( L \) with respect to \( e \) and to check that the second-order conditions for the choice of \( u(\Pi) \) and \( e \) are satisfied (conditional on the manager’s second-order condition being satisfied).

The interpretation of the result is particularly simple when there are only two possible levels of effort—low and high (\( e_1 \) and \( e_2 \))—rather than a continuum. On the lower bound \( \mu(\Pi_0) \). This deter the manager from choosing any level of effort \( e < e_0 \). Such extreme punishments are not feasible if \( u(\cdot) \) is constrained, say, by limited liability. Thus, a moving support is basically equivalent to full information. The same phenomenon can also occur with nonmoving, infinite support, as was shown by Mirrlees [1974].

The Theory of the Firm
assumption that the shareholders want to induce the high level of effort, the incentive-compatibility constraint becomes
\[
\mathbb{E}_{\Pi} \left[ \frac{1}{u'(w(\Pi))} \right] = \lambda + \eta \left( 1 - \frac{f_0(\Pi)}{f_0(\overline{\Pi})} \right), \quad \lambda, \eta > 0. \tag{12'}
\]
where \(f_0(\cdot)\) and \(f_0(\cdot)\) denote the densities for effort levels \(e_0\) and \(e_0\). Equation 12 then becomes
\[
\frac{1}{u'(w(\Pi))} = \lambda + \eta \left( 1 - \frac{f_0(\Pi)}{f_0(\overline{\Pi})} \right) \quad \lambda, \eta > 0. \tag{12''}
\]
Because \(u'\) is decreasing, \(1/u'\) is increasing. Thus, the higher the relative probability that effort was high when profit \(\Pi\) was observed, the higher the manager's wage will be. The term \(f_0(\Pi)/f_0(\overline{\Pi})\) is called the likelihood ratio. The optimal wage function is increasing with the realized profit if the likelihood ratio is decreasing. (See Milgrom 1981 for a useful discussion of likelihood ratios.)

This property is a natural one to assume (although it is trivial to construct conditional distributions that do not satisfy it). If higher profits are indeed a correct signal of higher effort, managerial compensation increases with observed profits.

When is the first-order approach valid? In other words, when does the optimal solution satisfy the manager's second-order condition? As was shown with increasing generality by Mirrlees (1975), Grossman and Hart (1983), and Rogerson (1985), the following conditions are sufficient for the first-order approach to be valid (and, in particular, for equation 12 or 12' to describe the optimal compensation scheme):

1. Monotonic-likelihood-ratio property: \(f_0/f_0(\cdot)\) increases with \(\Pi\) (or in the two-level-of-effort case, \(f_0/f_0(\cdot)\) increases with \(\Pi\).
2. Convexity of the distribution function: \(F_{e_0} \geq 0\), or (more generally) for all \(e_1, e_2, \Pi, \) and any \(x \in [0, 1],\)
\[
F(\Pi; x_1 + (1 - x)x_2) \leq x F(\Pi; x_1) + (1 - x) F(\Pi; x_2).
\]

That is, the deterministic effort \(x_1 + (1 - x)x_2\) is stochastically superior to effort \(x_1\) with probability \(x\) and to effort \(x_2\) with probability \(1 - x\). This convexity assumption has the flavor of decreasing returns to scale.\(^{109}\)

Equation 12 also teaches us something interesting about the value of a signal. Suppose that the shareholders observe not only profit \(\Pi\) but also some other signal, \(s\). This signal could be the price of inputs, the performances of other firms, or news about the economy. If \(G(\Pi, s, e)\) denotes the joint distribution of \(\Pi\) and \(s\) for an effort \(e\), with density \(g(\Pi, s, e)\), then equation 12 becomes
\[
\frac{1}{u'(w(\Pi, s))} = \lambda + \eta \frac{g_{e}(\Pi, s, e)}{g(\Pi, s, e)}. \tag{12''}
\]

Equation 12' gives the same wage structure as equation 12 (i.e., \(w\) depends only on \(\Pi\)) if
\[
g_{e}(\Pi, s, e) = \frac{f_{e}(\Pi, s)}{f(\Pi, s)} \tag{13}
\]

However, equation 13, if integrated with respect to \(e\), is equivalent to the existence of two functions \(m\) and \(n\) such that
\[
g(\Pi, s; e) = m(\Pi, s)n(\Pi, s) \tag{14}
\]

Equation 14 says that \(\Pi\) is a sufficient statistic for \((\Pi, s)\) with respect to \(e\). Thus, the optimal incentive scheme uses the extra information \(s\) if and only if \(s\) is informative about \(e\) given that \(\Pi\) is already available. This theorem is proved more formally in Holmström 1979 and in Shavell 1979.

General results for the moral-hazard problem are rare. For instance, a reasonable prediction would be that a \$1 increase in profit results in an increase in the manager's wage of between \$0 (full insurance) and \$1 (residual claimancy). However, this need not hold. Even for a separable utility function for the manager (\(U = \phi(e)u(w) - \Phi(e)\) — a special case of which, \(u \equiv 1\), was used above), one can prove only that the reward function must be increasing at some level of profit and has slope < 1 at some
(possibly different) level of profit (Grossman and Hart 1983). Only when there are two possible profit levels, \( \Pi_1 \) and \( \Pi_2 \), does one necessarily have
\[
0 < \frac{w_2 - w_1}{\Pi_2 - \Pi_1} < 1
\]
(where \( w \) corresponds to \( \Pi_1 \)). Alternatively, one can suppose that the likelihood ratio is monotonic and that the distribution function is convex in the effort to apply the first-order approach and its conclusions (such as the monotonicity of the reward function). Getting more specific results requires stronger assumptions on the distribution and the utility functions.

### Answers and Hints

#### Exercise 1

(i) The efficient amount of investment is given by
\[
\max_{\ell} [v(\ell) - I].
\]
This yields \( I^* = 2 \).
(ii) Ex post bargaining leads to \( v(\ell) - p = p - c, \) so
\[
p = \frac{v(\ell) + c}{2}. \text{ Ex ante, the supplier maximizes}
\]
\[
\frac{v(\ell) + c}{2} - c - I = \frac{1}{2} \left( \frac{c}{2} - I^2 \right).
\]
This yields \( I = 1 \) (suboptimal investment). Not investing at all is not profitable, because \( c < \frac{1}{2} \).
(iii) If the buyer has the right to buy at price \( p \), the supplier either will not invest or will invest the minimum amount so that the buyer buys: \( v(\ell) = p \). The second strategy yields profit \( p - c - v^{-1}(p) \). If \( p = v(2) = 4 \), the investment is 2 and the profit is \( 2 - c > 0 \). If \( p < 4 \), investment is suboptimal or nonexistent. If the supplier has the right to sell at a given price, investment is nul.
(iv) Efficient investment.

#### Exercise 2

(i) \( U(w, e) \) should be constant for insurance purposes. Thus, \( u(w) = \Phi(e) = U_a \) for all \( \beta \). But the minimization of \( w + \beta - e \) subject to the previous constraint yields \( w = \bar{w} = e \) for all \( \beta \) if \( \lambda \) is small (i.e., if \( \lambda < \Phi'(v^*) \)).
(ii) Let \( w(\beta) \) and \( e(\beta) \) denote the wage and the effort in state \( \beta \). Incentive compatibility requires that
\[
\bar{w} + \lambda (w(\beta) - \bar{w}) - \Phi(e(\beta)) \\
\geq \bar{w} + \lambda (w(\bar{\beta}) - \bar{w}) - \Phi(e(\bar{\beta}) - (\bar{\beta} - \beta))
\]
or
\[
\lambda (w(\beta) - w(\bar{\beta})) \geq \Phi(e(\beta)) - \Phi(e(\bar{\beta}) - (\bar{\beta} - \beta)).
\]
Using the convexity of \( \Phi \) and the condition \( \lambda \leq \Phi'(v) \) for all \( e \), we get
\[
w(\beta) + \beta - e(\beta) \geq w(\bar{\beta}) + \bar{\beta} - e(\bar{\beta}).
\]
Thus, the cost for the principal cannot be lower in state \( \beta \) than in state \( \bar{\beta} \). For any wage structure, the manager's utility is
max\{\alpha(w(\beta - \epsilon)) - \Phi(\epsilon)\}.

(Because of the previous incentive-compatibility constraint, the worst state of nature for the manager is \(\beta\).) These two properties imply that the contract in the text is optimal. Because monetary incentives are too costly, the principal contents himself with a given cost target.

(iii) The cost of having two managers under full information is \(2\beta + E\beta - \epsilon^*\). But this cost is reached under the yardstick-competition scheme in the text: In state \(\beta\), manager 1 maximizes

\[
\max_{\epsilon^*} [\beta + \Phi(\epsilon^*)(\beta - \epsilon^*) - \Phi(\epsilon)].
\]

Hence, \(\epsilon = \epsilon^*\) for all \(\beta\).

References


56 The Theory of the Firm


