We measure welfare here by means of Marshallian aggregate surplus (see Section 10.E). In this case, social welfare when there are J active firms is given by

$$W(J) = \int_0^{Jq_J} p(s)\,ds - Jc(q_J) - JK. \tag{12.E.5}$$

The socially optimal number of active firms in this oligopolistic industry, which we denote by J^o, is any integer number that solves $\text{Max}_J\, W(J)$. Example 12.E.3 illustrates that in contrast with the conclusion arising in the case of a competitive market, the equilibrium number of firms here need not be socially optimal.

Example 12.E.3: Consider the Cournot model of Example 12.E.1. For the moment, ignore the requirement that the number of firms is an integer, and solve for the number of firms $\bar{J}$ at which $W'(\bar{J}) = 0$. This gives

$$(\bar{J} + 1)^3 = \frac{(a-c)^2}{bK}. \tag{12.E.6}$$

If $\bar{J}$ turns out to be an integer, then the socially optimal number of firms is $J^o = \bar{J}$. Otherwise, J^o is one of the two integers on either side of $\bar{J}$ [recall that $W(\cdot)$ is concave]. Now, recall from (12.E.4) that $\pi_J = (1/b)[(a-c)/(J+1)]^2$. As noted in Example 12.E.1, if we let $\tilde{J}$ be the real number such that

$$(\tilde{J} + 1)^2 = \frac{(a-c)^2}{bK}, \tag{12.E.7}$$

the equilibrium number of firms is the largest integer less than or equal to $\tilde{J}$. From (12.E.6) and (12.E.7), we see that

$$(\tilde{J} + 1) = (\bar{J} + 1)^{3/2}.$$

Thus, when the demand and cost parameters are such that the optimal number of firms is exactly two ($J^o = \bar{J} = 2$), four firms actually enter this market ($J^* = 4$, since $\tilde{J} \cong 4.2$); when the social optimum is for exactly three firms to enter ($J^o = \bar{J} = 3$), seven firms actually do ($J^* = 7$, since $\tilde{J} = 7$); when the social optimum is for exactly eight firms to enter ($J^o = \bar{J} = 8$), 26 actually enter ($J^* = 26$, since $\tilde{J} = 26$). ■

Can we say anything general about the nature of the entry bias? It turns out that we can as long as stage 2 competition satisfies three weak conditions [we follow Mankiw and Whinston (1986) here]:

(A1) $Jq_J \geq J'q_{J'}$ whenever $J > J'$;
(A2) $q_J \leq q_{J'}$ whenever $J > J'$;
(A3) $p(Jq_J) - c'(q_J) \geq 0$ for all J.

Conditions (A1) and (A3) are straightforward: (A1) requires that aggregate output increases (price falls) when more firms enter the industry, and (A3) says that price is not below marginal cost regardless of the number of firms entering the industry. Condition (A2) is more interesting. It is the assumption of *business stealing*. It says that when an additional firm enters the market, the sales of existing firms fall (weakly). Hence, part of the new firm's sales come at the expense of existing firms. These conditions are satisfied by most, although not all, oligopoly models. [In the Bertrand model, for example, condition (A3) does not hold.]

For markets satisfying these three conditions we have the result shown in Proposition 12.E.1.

Proposition 12.E.1: Suppose that conditions (A1) to (A3) are satisfied by the post-entry oligopoly game, that $p'(\cdot) < 0$, and that $c''(\cdot) \geq 0$. Then the equilibrium number of entrants, J^*, is at least $J^\circ - 1$, where J° is the socially optimal number of entrants.[23]

Proof: The result is trivial for $J^\circ = 1$, so suppose that $J^\circ > 1$. Under the assumptions of the proposition, π_J is decreasing in J (Exercise 12.E.2 asks you to show this). To establish the result, we therefore need only show that $\pi_{J^\circ - 1} \geq K$.

To prove this, note first that by the definition of J° we must have $W(J^\circ) - W(J^\circ - 1) \geq 0$, or

$$\int_{Q_{J^\circ - 1}}^{Q_{J^\circ}} p(s)\,ds - J^\circ c(q_{J^\circ}) + (J^\circ - 1)c(q_{J^\circ - 1}) \geq K,$$

where we let $Q_J = Jq_J$. We can rearrange this expression to yield

$$\pi_{J^\circ - 1} - K \geq p(Q_{J^\circ - 1})q_{J^\circ - 1} - \int_{Q_{J^\circ - 1}}^{Q_{J^\circ}} p(s)\,ds + J^\circ[c(q_{J^\circ}) - c(q_{J^\circ - 1})].$$

Given $p'(\cdot) < 0$ and condition (A1), this implies that

$$\pi_{J^\circ - 1} - K \geq p(Q_{J^\circ - 1})[q_{J^\circ - 1} + Q_{J^\circ - 1} - Q_{J^\circ}] + J^\circ[c(q_{J^\circ}) - c(q_{J^\circ - 1})]. \quad (12.E.8)$$

But since $c''(\cdot) \geq 0$, we know that $c'(q_{J^\circ - 1})[q_{J^\circ} - q_{J^\circ - 1}] \leq c(q_{J^\circ}) - c(q_{J^\circ - 1})$. Using this inequality with (12.E.8) and the fact that $q_{J^\circ - 1} + Q_{J^\circ - 1} - Q_{J^\circ} = J^\circ(q_{J^\circ - 1} - q_{J^\circ})$ yields

$$\pi_{J^\circ - 1} - K \geq [p(Q_{J^\circ - 1}) - c'(q_{J^\circ - 1})]J^\circ(q_{J^\circ - 1} - q_{J^\circ}).$$

Conditions (A2) and (A3) then imply that $\pi_{J^\circ - 1} \geq K$.[24] ■

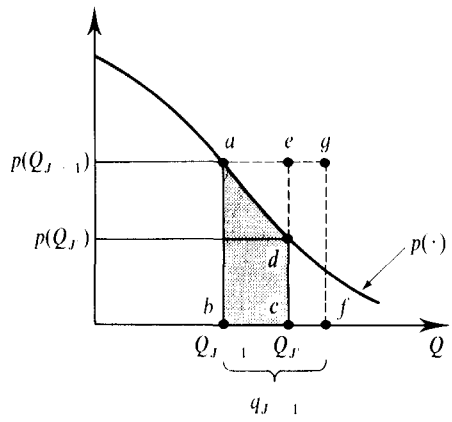
The idea behind the proof of Proposition 12.E.1 is illustrated in Figure 12.E.1 for the case where $c(q) = 0$ for all q. In the figure, the incremental welfare benefit of the J°th firm, before taking its entry cost into account, is represented by the shaded area $(abcd)$. Since entry of this firm is socially efficient, this area must be at least K. But area $(abcd)$ is less than area $(abce)$, which equals $p(Q_{J^\circ - 1})(Q_{J^\circ} - Q_{J^\circ - 1})$. Moreover, business stealing implies that $(Q_{J^\circ} - Q_{J^\circ - 1}) = J^\circ q_{J^\circ} - (J^\circ - 1)q_{J^\circ - 1} \leq q_{J^\circ - 1}$, and so we see that area $(abce) \leq p(Q_{J^\circ - 1})q_{J^\circ - 1} = \pi_{J^\circ - 1}$ [the value of $\pi_{J^\circ - 1}$ is represented in Figure 12.E.1 by area $(abfg)$]. Hence $\pi_{J^\circ - 1} \geq K$.

The tendency for excess entry in the presence of market power is fundamentally driven by the business-stealing effect. When business stealing accompanies new entry and price exceeds marginal cost, part of a new entrant's profit comes at the expense of existing firms, creating an excess incentive for the new firm to enter.

Of course, as Proposition 12.E.1 indicates, we may also see too few firms in an industry. The classic example concerns a situation in which the socially optimal number of firms is one. A single firm deciding whether to enter a market as a

23. If there is more than one maximizer of $W(J)$, say $\{J_1^\circ, \ldots, J_N^\circ\}$, then $J^* \geq \text{Max}\{J_1^\circ, \ldots, J_N^\circ\} - 1$.

24. Note that if (A1) holds with strict inequality, then this conclusion can be strengthened to $\pi_{J^\circ - 1} > K$ [a strict inequality appears in (12.E.8)]. In this case, $J^* \geq J^\circ - 1$ even if firms do not enter when indifferent.



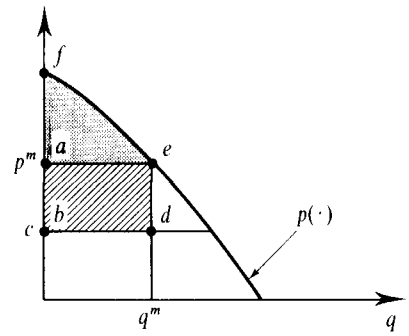


Figure 12.E.1 (left)
Diagrammatic explanation of Proposition 12.E.1.

Figure 12.E.2 (right)
An insufficient entry incentive.

monopolist compares its monopoly profit—the hatched area (*abde*) in Figure 12.E.2—with the entry cost K. However, the firm fails to capture, and therefore ignores, the increase in consumer surplus that its entry generates—the shaded area (*fae*). As a result, the firm may find entry unprofitable even though it is socially desirable. Proposition 12.E.1 tells us, however, that if we have too little entry in a homogeneous-good market, this can be at most by a single firm.

What happens when product differentiation is present? It turns out that we can then say very little of a general nature. The reason is that the sort of problem illustrated in Figure 12.E.2 can now happen for many products, leading to many "too few by one" conclusions. An additional issue is that, with product differentiation, the number of firms is not all that matters. We may also fail to have the right selection of products.[25]

An alternative approach to the two-stage entry game models the actions of entry and quantity/price choice as simultaneous. In this *one-stage entry game*, a firm incurs its setup cost only if it sells a positive amount. For example, the one-stage versions of Examples 12.E.1 and 12.E.2 are Cournot and Bertrand games, respectively, with cost functions

$$C(q) = \begin{cases} K + c(q) & \text{if } q > 0 \\ 0 & \text{if } q = 0 \end{cases}$$

and an infinite (or very large) number of firms. For models of price competition, this change can have dramatic consequences. Consider the effect on the result of Example 12.E.2 that is illustrated in Example 12.E.4.

Example 12.E.4: *The One-Stage Entry Model with Bertrand Competition.* Suppose that $p > [K + cx(p)]/x(p)$ for some p (the parameter $c > 0$ is the cost per unit); that is, suppose there is some price level at which a monopolist can earn strictly positive profits after paying its set up cost K. Assume that many firms simultaneously name prices and that a firm incurs the setup cost K only if it actually makes sales. Any equilibrium of this game has all sales occurring at price $p^* = \text{Min}\{p: p \geq [K + cx(p)]/x(p)\}$ (if price is above p^*, some firm could gain by setting a price $p^* - \varepsilon$; if price is below p^*, some firm must be making strictly negative profits), and one firm satisfying all demand at this price (if the demand were split among

25. See Spence (1976), Dixit and Stiglitz (1977), Salop (1979), and Mankiw and Whinston (1986) for more on the case of product differentiation.

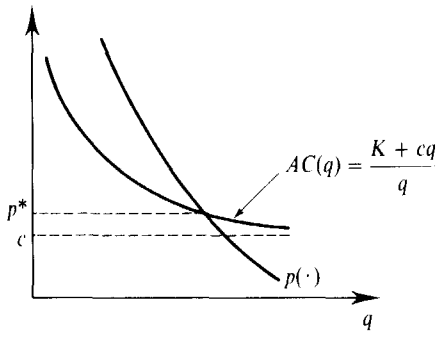


Figure 12.E.3
Equilibrium in the one-stage entry game discussed in Example 12.E.4.

several firms at price p^*, none of them could cover their cost).[26] In this equilibrium, all firms make zero profits. The equilibrium outcome is depicted in Figure 12.E.3. Observe that it is strictly superior in welfare terms to the outcome that arises in the two-stage entry process considered in Example 12.E.2, where there is also a single firm active but it quotes a monopoly price.[27] ■

What is the critical difference between the one-stage and two-stage entry processes? In the two-stage model an entrant must sink its fixed costs prior to competing, whereas in the one-stage model it can compete for sales while retaining the option not to sink these costs if it does not make any sales. We can think of the two-stage case as a model of a firm incurring a once-and-for-all sunk entry cost that allows for many later periods of competitive interaction, whereas the one-stage case captures a setting in which "hit-and-run" entry is possible (i.e., entry for just one period while paying only the one-period rental price of capital). When a firm must incur a sunk cost in entering it must consider the reaction of other firms to its entry. In the Bertrand model with constant costs this reaction is severe: price falls to cost and the firm loses money by entering. In contrast, in the one-stage game the firm can enter and undercut active firms' prices without fearing their reactions. This makes entry more aggressive and leads to a lower equilibrium price. This one-stage entry model with price competition provides one formalization of what Baumol, Panzar, and Willig (1982) call a *contestable market*.

12.F The Competitive Limit

In Chapter 10, we introduced the idea that a competitive market might usefully be thought of as a limiting case of an oligopolistic market in which firms' market power grows increasingly small (see Section 10.B). We also noted that this view could provide a framework for reconciling cases in which competitive equilibria fail to exist in the presence of free entry and average costs that exhibit a strictly positive efficient

26. Note that we now allow consumer demand to be given entirely to one firm when several firms name the same price (before, we had taken the division of demand in this case to be exogenously given). This is the only division of demand that is compatible with equilibrium in this example. It can be formally justified as the limit of the equilibria that arise when prices must be quoted in discrete units as the size of these units grows small.

27. In fact, this equilibrium outcome is the solution to the problem faced by a welfare-maximizing planner who can control the outputs q_j of the firms but must guarantee a nonnegative profit to all active firms, that is, who faces the constraint that $p(\sum_k q_k)q_j \geq cq_j + K$ for every j with $q_j > 0$.

scale (see Section 10.F). In this situation, we argued, as long as many firms could fit into the market, the market outcome ought to be close to the competitive outcome that would arise if industry average costs were actually constant at the level of minimum average cost. In this section, we elaborate on these points and develop, in a setting of free entry, the theme that if the size of individual firms is small relative to the size of the market, then the equilibrium will be nearly competitive.

We have already seen one example of this phenomenon in Example 12.E.1. Here we establish the point in a more general way. We now let market demand be $x_\alpha(p) = \alpha x(p)$, where $x(p)$ is differentiable and $x'(\cdot) < 0$. Increases in α correspond to proportional increases in demand at all prices. Letting $p(q)$ be the inverse demand function associated with $x(p)$, the inverse demand function associated with $x_\alpha(p)$ is then $p_\alpha(q) = p(q/\alpha)$. All potential firms have a strictly convex cost function $c(q)$ and entry cost $K > 0$. We denote the level of minimum average cost for a firm by $\bar{c} = \text{Min}_{q>0}\,[K + c(q)]/q$, and we let $\bar{q} > 0$ denote a firm's (unique) efficient scale.

As in Example 12.E.1, we focus on the case of a two-stage entry model with Cournot competition in the second stage, in which the cost K is incurred only if the firm decides to enter in stage 1. We let $b(Q_{-j})$ denote active firm j's optimal output level for any given level of aggregate output by its rivals, Q_{-j}, and we assume that this best response is unique for all Q_{-j}.

Finally, we let p_α and Q_α denote the price and aggregate output in a subgame perfect Nash equilibrium (SPNE) of the two-stage Cournot entry model when the market size is α. We denote by P_α the set of all SPNE prices for market size α.

Proposition 12.F.1: As the market size grows, the price in any subgame perfect Nash equilibrium of the two-stage Cournot entry model converges to the level of minimum average cost (the "competitive" price). Formally,

$$\underset{p_\alpha \in P_\alpha}{\text{Max}}\, |p_\alpha - \bar{c}| \to 0 \quad \text{as} \quad \alpha \to \infty.$$

Proof: The argument consists of three steps:

(i) First, you are asked in Exercise 12.F.1 to show that for large enough α, an active firm's best-response function $b(Q_{-j})$ is (weakly) decreasing in Q_{-j}.

(ii) Second, we argue that if $b(Q_{-j})$ is decreasing, then we must have $Q_\alpha \geq \alpha x(\bar{c}) - \bar{q}$ in any SPNE of the two-stage entry game with market size α. To see why this is so, suppose that with market size α we had an SPNE with J_α firms entering and an aggregate output level $Q_\alpha < \alpha x(\bar{c}) - \bar{q}$. Consider any firm j whose equilibrium entry choice is "out" in this equilibrium, and suppose that firm j instead decided to enter and produce quantity $\bar{q}$. Because $b(\cdot)$ is decreasing, it is intuitively plausible that the aggregate output level of the original J_α active firms cannot increase when firm j enters in this way (see the small-type paragraph that follows for the formal argument behind this claim). As a result, aggregate output in the market following firm j's entry is no more than $(Q_\alpha + \bar{q})$; and since $(Q_\alpha + \bar{q}) < \alpha x(\bar{c})$, the resulting (post-entry) price is above $\bar{c}$. Hence, firm j would earn strictly positive profits by entering in this fashion, contradicting the hypothesis that we were at an SPNE to start with.

The argument that the output of the existing J_α firms cannot increase following entry of firm j is as follows: Let Q_{-j} be the initial equilibrium level of these firms' aggregate output,

and let $\tilde{Q}_{-j}$ be their post-entry aggregate output. Suppose that $\tilde{Q}_{-j} > Q_{-j}$. Then at least one of these firms, say firm k, must have increased its output level in response to firm j's entry, say from q_k to $\tilde{q}_k > q_k$. Because $b(\cdot)$ is decreasing, it must be that $\tilde{Q}_{-k} < Q_{-k}$; that is, the post-entry output $\tilde{Q}_{-k}$ of active firms other than k (which includes firm j) must be less than their pre-entry output, Q_{-k}. By part (c) of Exercise 12.C.8, this implies that $q_k + Q_{-k} \geq \tilde{q}_k + \tilde{Q}_{-k}$. But $Q_{-j} = q_k + Q_{-k}$ (since firm j initially produces nothing), and $\tilde{q}_k + \tilde{Q}_{-k} \geq \tilde{Q}_{-j}$ (because firm j's post-entry output is nonnegative). Hence, $Q_{-j} \geq \tilde{Q}_{-j}$, which is a contradiction.

(iii) Finally, we argue that the conclusion of (ii) implies the result. To see this, consider how much above $\bar{c}$ the price can be if aggregate output is no more than $\bar{q}$ below $\alpha x(\bar{c})$. This is given by

$$\begin{aligned}\Delta p_\alpha &= p_\alpha(\alpha x(\bar{c}) - \bar{q}) - p_\alpha(\alpha x(\bar{c})) \\ &= p\left(\frac{\alpha x(\bar{c}) - \bar{q}}{\alpha}\right) - p(x(\bar{c})).\end{aligned}$$

But as $\alpha \to \infty$, $[\alpha x(\bar{c}) - \bar{q}]/\alpha \to x(\bar{c})$, so that $\Delta p_\alpha \to 0$. ■

There are two forces driving Proposition 12.F.1. First, the entry process ensures that firms will enter if there is too much "room" left in the market. Second, in a market that is very large relative to the minimum efficient scale, a reduction of output equal to the level of minimum efficient scale has very little effect on price. The consequence of these two facts is that as the market size grows large, firms' market power is dissipated and price approaches the level of minimum average cost (the competitive level). In this limiting outcome, welfare approaches its optimal level.[28]

In Example 12.E.2, we saw that in a two-stage Bertrand market, no such limiting result holds.[29] Because price drops to marginal cost if even two firms enter, the market is always monopolized, no matter what its size. However, the two-stage Bertrand model's limiting properties are quite special. As long as, for any market size, price is above marginal cost for any finite number of firms that enter the market, and approaches marginal cost as the number of firms grows large, a limiting result like that in Proposition 12.F.1 holds.

Finally, Proposition 12.F.1 applies only for the case of homogeneous-good markets. With product differentiation, we must be careful. Firms may be small relative to the size of the entire set of interrelated markets, but they may still be large relative to their own particular niche. In this case, each firm may maintain substantial market power even in the limit, and the limiting equilibrium can be far from efficient (see Exercise 12.F.4).

28. The sense of approximation is relative to the size parameter of the market α. Assuming that α is a proxy for the number of consumers, this means that the welfare loss per consumer relative to the social optimum goes to zero.

29. Strictly speaking, firms' cost functions in Example 12.E.2 differ from the cost functions assumed in Proposition 12.F.1 (average costs including K are declining everywhere in Example 12.E.2). Nevertheless, for the two-stage Cournot model, Proposition 12.F.1 can be shown to be valid for the cost function of Example 12.E.2 (letting $\bar{c}$ in the statement of the proposition now be the limiting value of average cost as a firm's output grows large).

12.G Strategic Precommitments to Affect Future Competition

An important feature of many oligopolistic settings is that firms attempt to make strategic precommitments in order to alter the conditions of future competition in a manner that is favorable to them. Examples of strategic precommitments abound. For example, investments in cost reduction, capacity, and new-product development all lead to long-lasting changes that can affect the nature of future competition. In practice, these types of decisions can be among the most important competitive decisions that firms make.

Some general features of these types of strategic precommitments can be usefully illuminated through examination of the following simple two-stage duopoly model:

Stage 1: Firm 1 has the option to make a strategic investment, whose level we denote by $k \in \mathbb{R}$. This choice is observable.

Stage 2: Firms 1 and 2 play some oligopoly game, choosing strategies $s_1 \in S_1 \subset \mathbb{R}$ and $s_2 \in S_2 \subset \mathbb{R}$, respectively. Given investment level k and strategy choices (s_1, s_2), profits for firms 1 and 2 are given by $\pi_1(s_1, s_2, k)$ and $\pi_2(s_1, s_2)$, respectively.

For example, k might be an investment that reduces firm 1's marginal cost of production with the stage 2 game being Cournot competition (so $s_j = q_j$, firm j's quantity choice). Alternatively, stage 2 competition could be differentiated products price competition.

We suppose that there is a unique Nash equilibrium in stage 2 given any choice of k, $(s_1^*(k), s_2^*(k))$, and we assume for convenience that it is differentiable in k. We also assume for purposes of our discussion that $\partial\pi_1(s_1, s_2, k)/\partial s_2 < 0$ and $\partial\pi_2(s_1, s_2)/\partial s_1 < 0$, that is, that stage 2 actions are "aggressive" in the sense that a higher level of s_{-j} by firm j's rival lowers firm j's profit. Hence, firm 1 would be better off, all else being equal, if it could induce firm 2 to lower its choice of s_2.

When can investment by firm 1 cause firm 2 to lower s_2? Letting $b_1(s_2, k)$ and $b_2(s_1)$ denote firm 1's and firm 2's stage 2 best-response functions (note that firm 1's best response depends on k), we can differentiate the equilibrium condition $s_2^* = b_2(b_1(s_2^*, k))$ to get

$$\frac{ds_2^*(k)}{dk} = \frac{db_2(s_1^*(k))}{ds_1}\left(\frac{\partial b_1(s_2^*(k), k)/\partial k}{1 - [\partial b_1(s_2^*(k), k)/\partial s_2][db_2(s_1^*(k))/ds_1]}\right). \quad (12.G.1)$$

The denominator of the second term on the right-hand side of (12.G.1) being nonegative is often called the *stability condition.* It implies that the simple dynamic adjustment process in which the firms take turns myopically playing a best response to each others' current strategies converges to the Nash equilibrium from any strategy pair in a neighborhood of the equilibrium. We shall maintain this assumption for the remainder of our discussion. Thus, the effect of k on s_2 can be seen to depend on two factors: (i) Does k make firm 1 more or less "aggressive" in stage 2 competition [i.e., what is the sign of $\partial b_1(s_2^*(k), k)/\partial k$]? and (ii) Does firm 2 respond to the anticipation of more aggressive play by firm 1 with more aggression itself or with less [i.e., what is the sign of $db_2(s_1^*(k))/ds_1$]?

	Strategic Substitutes: $\frac{db_2(\cdot)}{ds_1} < 0$	Strategic Complements $\frac{db_2(\cdot)}{ds_1} > 0$
$\frac{\partial b_1(\cdot)}{\partial k} > 0$	$\frac{ds_2^*(k)}{dk} < 0$	$\frac{ds_2^*(k)}{dk} > 0$
$\frac{\partial b_1(\cdot)}{\partial k} < 0$	$\frac{ds_2^*(k)}{dk} > 0$	$\frac{ds_2^*(k)}{dk} < 0$

Figure 12.G.1
Determinants of the sign of $ds_2^*(k)/dk$.

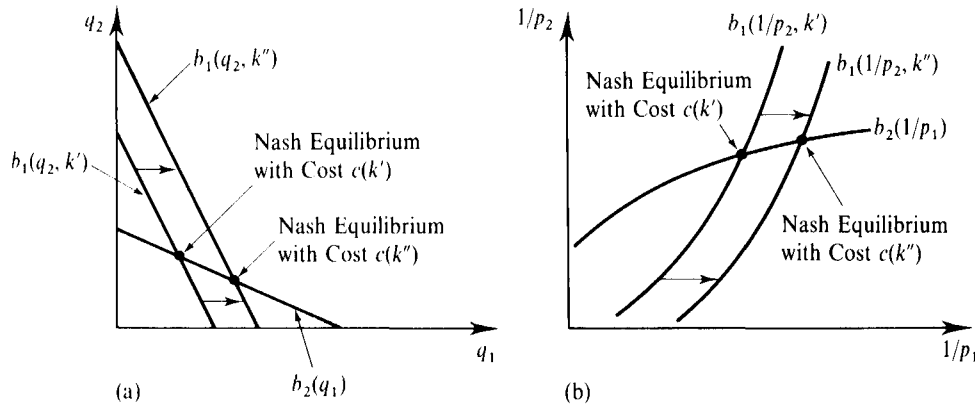


Figure 12.G.2
Strategic effects of a reduction in marginal cost from $c(k')$ to $c(k'') < c(k')$.
(a) Quantity model.
(b) Price model.

When firm 2 responds in kind to more aggressive choices of s_1 by firm 1 [i.e., when $db_2(s_1^*(k))/ds_1 > 0$], we say that s_2 is a *strategic complement* of s_1; and if firm 2 becomes less aggressive in the face of more aggressive play by firm 1 [i.e., if $db_2(s_1^*(k))/ds_1 < 0$], s_2 is a *strategic substitute* of s_1. [This terminology is derived from Bulow, Geanakoplos, and Klemperer (1985); see also Fudenberg and Tirole (1984) for a related taxonomy.]

Figure 12.G.1 summarizes these two determinants of firm 2's response, $ds_2^*(k)/dk$.

Example 12.G.1: *The Strategic Effects from Investment in Marginal Cost Reduction.* The importance for strategic behavior of the distinction between cases of strategic complements and strategic substitutes is nicely illustrated by examining the strategic effects of investments in marginal cost reduction for models of quantity versus price competition.

Suppose that if firm 1 invests k then its (constant) per-unit production costs are $c(k)$, where $c'(k) < 0$. Consider, first, the case in which stage 2 competition takes the form of the Cournot model of Example 12.C.1, so that the stage 2 strategic variable is $s_j = q_j$, firm j's quantity choice. In this model, we have a situation of strategic substitutes because firm 2's best-response function in stage 2 is downward sloping [$db_2(q_1)/dq_1 < 0$ at all q_1 such that $b_2(q_1) > 0$]. As shown in Figure 12.G.2(a), the lowering of firm 1's marginal cost because of an increase in k from, say, k' to $k'' > k'$, shifts firm 1's best-response function outward from $b_1(q_2, k')$ to $b_1(q_2, k'')$; with lower marginal costs, firm 1 will wish to produce more for any quantity choice of its rival

[and so, in terms of our earlier analysis, $\partial b_1(q_2^*(k), k)/\partial k > 0$]. Thus, in this model, investment in cost reduction leads to a reduction in firm 2's output level, an effect that is beneficial for firm 1 [see Figure 12.G.2(a)].

In contrast, suppose that stage 2 competition takes the form of the differentiated price competition model of Example 12.C.2. Here we take $s_j = (1/p_j)$ to conform with the interpretation of s_j as an "aggressive" variable [i.e., $\partial \pi_1(s_1, s_2, k)/\partial s_2 < 0$]. In this model, we have a situation of strategic complements: an anticipated reduction in firm 1's price causes firm 2 to reduce its price also [i.e., $db_2(1/p_1)/d(1/p_1) > 0$]. As depicted in Figure 12.G.2(b), a reduction in firm 1's marginal cost because of an increase in k from k' to $k'' > k'$ once again makes firm 1 more aggressive, leading it to choose a lower price given any price choice of its rival; its best-response function shifts to the right from $b_1(1/p_2, k')$ to $b_1(1/p_2, k'')$ [hence, in terms of our earlier analysis, $\partial b_1(1/p_2^*(k), k)/\partial k > 0$]. With strategic complements, the result of the reduction in firm 1's marginal cost is therefore to lower firm 2's equilibrium price, an effect that is undesirable for firm 1.

Thus, the strategic effects of a reduction in firm 1's marginal cost differ between the two models, being beneficial to firm 1 in the quantity model and detrimental in the price model.[30] Which model more accurately captures the nature of competitive interaction depends on the particulars of an industry's situation. For example, if firms in a mature industry have excess capacity, the price model is likely to be more descriptive, and the strategic effect will be detrimental. On the other hand, in a new market where firms are investing in capacity, the strategic effect is likely to be better captured by the quantity model (recall our interpretation of the Cournot model in terms of capacity choices in Section 12.C). ■

In deciding on its level of investment, firm 1 must therefore consider not only the direct effects of its investment (say, the direct benefit of lower costs), but also the strategic effects that arise through induced changes in its its rival's behavior. Formally, the derivative of firm 1's profits with respect to a change in k can be written as

$$\frac{d\pi_1(s_1^*(k), s_2^*(k), k)}{dk} = \frac{\partial \pi_1(s_1^*(k), s_2^*(k), k)}{\partial k} + \frac{\partial \pi_1(s_1^*(k), s_2^*(k), k)}{\partial s_1}\frac{ds_1^*(k)}{dk} + \frac{\partial \pi_1(s_1^*(k), s_2^*(k), k)}{\partial s_2}\frac{ds_2^*(k)}{dk}.$$

Since at a Nash equilibrium in stage 2 given investment level k we have $\partial \pi_1(s_1^*(k), s_2^*(k), k)/\partial s_1 = 0$, this simplifies to

$$\frac{d\pi_1(s_1^*(k), s_2^*(k), k)}{dk} = \frac{\partial \pi_1(s_1^*(k), s_2^*(k), k)}{\partial k} + \frac{\partial \pi_1(s_1^*(k), s_2^*(k), k)}{\partial s_2}\frac{ds_2^*(k)}{dk}. \quad (12.G.2)$$

The first term on the right-hand side of (12.G.2) is the *direct effect* on firm 1's profits from changing k; the second term is the *strategic effect* that arises because of firm 2's equilibrium response to the change in k. Since $\partial \pi_1(s_1^*(k), s_2^*(k), k)/\partial s_2 < 0$, the strategic effect on firm 1's profits is positive if $ds_2^*(k)/dk < 0$, that is, if firm 2's response to increases in firm 1's investment is to lower its choice of s_2.

30. Best-response functions need not always slope this way in the price and quantity models, but the particular examples considered here represent the "normal" cases; see Exercise 12.C.12.

In the above discussion, we have considered situations in which a firm makes a strategic precommitment to affect future competition with another firm who is (or will be) in the market. A particularly striking example of strategic precommitment to affect future market conditions, however, arises when one firm is the first into an industry and seeks to use its first-mover advantage to deter further entry into its market. We can analyze this case formally by introducing a stage between stages 1 and 2, say stage 1.5, at which firm 2 decides whether to be in the market and by supposing that if firm 2 chooses "in" then it must pay a set-up cost $F > 0$. Firm 2 will therefore choose "out" given firm 1's stage 1 choice of k if its anticipated profit in stage 3, $\pi_2(s_1^*(k), s_2^*(k))$, is less than F. Given this fact, the incumbent would, of course, like simply to announce that in response to any entry it will engage in predatory pricing (i.e., it will choose a very high level of s_1 in stage 3). The problem, however, is that this threat must be *credible* (recall the discussion in Chapter 9). Thus, what the incumbent needs to do to deter entry is choose a level of k that precommits it to sufficiently aggressive behavior that firm 2 chooses not to enter. In any particular problem, this may or may not be possible, and it may or may not be profitable. As a general matter, there are many potential mechanisms (i.e., many types of variables k) by which such precommitments can be made. In Appendix B, we examine in some detail the classic mechanism of entry deterrence through capacity expansion first studied by Spence (1977) and Dixit (1980).

APPENDIX A: INFINITELY REPEATED GAMES AND THE FOLK THEOREM

In this appendix, we extend the discussion in Section 12.D of infinitely repeated games to a more general setting. Our primary aim is to develop a formal statement of a version of the *folk theorem* of infinitely repeated games. Infinitely repeated games have a very rich theoretical structure and we shall only touch on a limited number of their properties. Fudenberg and Tirole (1992) and Osborne and Rubinstein (1994) provide more extended discussions.

The Model

An infinitely repeated game consists of an infinite sequence of repetitions of a one-period simultaneous-move game, known as the *stage game*. For expositional simplicity, we focus here on the case in which there are two players.

In the one-period stage game, each player i has a compact strategy set S_i; $q_i \in S_i$ is a particular feasible action for player i. Denote $q = (q_1, q_2)$ and $S = S_1 \times S_2$. Player i's payoff function is $\pi_i(q_i, q_j)$. We restrict our attention throughout to pure strategies. It will be convenient to define player i's one-period best-response payoff given that his rival plays q_j by $\hat{\pi}_i(q_j) = \text{Max}_{q \in S_i} \pi_i(q, q_j)$.[31] We assume that the stage game has a unique pure strategy Nash equilibrium $q^* = (q_1^*, q_2^*)$ (the assumption of uniqueness is for expositional simplicity only).

In the infinitely repeated game, actions are taken and payoffs are earned at the beginning of each period. The players discount payoffs with discount factor $\delta < 1$.

31. We assume that conditions on the sets S_i and functions $\pi_i(q_i, q_j)$ hold such that this function exists (i.e., such that each player's best response is always well defined).

Players observe each other's action choices in each period and have perfect recall. A pure strategy in this game for player i, s_i, is a sequence of functions $\{s_{it}(\cdot)\}_{t=1}^{\infty}$ mapping from the history of previous action choices (denoted H_{t-1}) to his action choice in period t, $s_{it}(H_{t-1}) \in S_i$. The set of all such pure strategies for player i is denoted by Σ_i, and $s = (s_1, s_2) \in \Sigma_1 \times \Sigma_2$ is a profile of pure strategies for the two players.

Any pure strategy profile $s = (s_1, s_2)$ induces an *outcome path* $Q(s)$, an infinite sequence of actions $\{q_t = (q_{1t}, q_{2t})\}_{t=1}^{\infty}$ that will actually be played when the players follow strategies s_1 and s_2. Player i's discounted payoff from outcome path Q is given by $v_i(Q) = \sum_{\tau=0}^{\infty} \delta^{\tau}\pi_i(q_{1+\tau})$. We also define player i's *average payoff* from outcome path Q to be $(1-\delta)v_i(Q)$; this is the per-period payoff that, if infinitely repeated, would give player i a discounted payoff of $v_i(Q)$. Finally, it is also useful to define the discounted continuation payoff from outcome path Q from some period t onward (discounted to period t) by $v_i(Q, t) = \sum_{\tau=0}^{\infty} \delta^{\tau}\pi_i(q_{t+\tau})$.

We can note immediately the following fact: The strategies that call for each player i to play his stage game Nash equilibrium action q_i^* in every period, regardless of the prior history of play, constitute an SPNE for *any* value of $\delta < 1$. In the discussion that follows, we are interested in determining to what extent repetition allows other outcomes to emerge as SPNEs.

Nash Reversion and the Nash Reversion Folk Theorem

We begin by considering strategies with the Nash reversion form that we considered for the Bertrand pricing game in Section 12.D.

Definition 12.AA.1: A strategy profile $s = (s_1, s_2)$ in an infinitely repeated game is one of *Nash reversion* if each player's strategy calls for playing some outcome path Q until someone defects and playing the stage game Nash equilibrium $q^* = (q_1^*, q_2^*)$ thereafter.

What outcome paths Q can be supported as outcome paths of an SPNE using Nash reversion strategies? Following logic similar to that discussed in Section 12.D, we can derive the test in Lemma 12.AA.1.

Lemma 12.AA.1: A Nash reversion strategy profile that calls for playing outcome path $Q = \{q_{1t}, q_{2t}\}_{t=1}^{\infty}$ prior to any deviation is an SPNE if and only if

$$\hat{\pi}_j(q_{jt}) + \frac{\delta}{1-\delta}\pi_j(q_1^*, q_2^*) \leq v_j(Q, t) \qquad (12.AA.1)$$

(where $j \neq i$) for all t and $i = 1, 2$.

Proof: As discussed in Section 12.D, the prescribed play after any deviation is a Nash equilibrium in the continuation subgame; so we need only check whether these strategies induce a Nash equilibrium in the subgame starting in any period t when there has been no previous deviation. Note first that if for some i and t condition (12.AA.1) did not hold, then we could not have an SPNE. That is, if no deviation had occurred prior to period t, then in the continuation subgame, player i would not find following path Q to be his best response to player j's doing so (in particular, a deviation by player i in period t that maximizes his payoff in that period, followed by his playing q_i^* thereafter, would be superior for him).

In the other direction, suppose that condition (12.AA.1) is satisfied for all i and t but that we do not have an SPNE. Then there must be some period t in which some player i finds it worthwhile to deviate from outcome path Q if no previous deviation has occurred. Now, when his opponent follows a Nash revision strategy, player i's optimal deviation will involve deviating in a manner that maximizes his payoff in period t and then playing q_i^* thereafter. But his payoff from this deviation is exactly that on the left side of condition (12.AA.1), and so this deviation cannot raise his payoff. ■

Condition (12.AA.1) can be written to emphasize the trade-off between one-period gains and future losses as follows:

$$\hat{\pi}_i(q_{jt}) - \pi_i(q_{1t}, q_{2t}) \leq \delta\left(v_i(Q, t+1) - \frac{\pi_i(q_1^*, q_2^*)}{1-\delta}\right) \tag{12.AA.2}$$

for all t and $i = 1, 2$. The left-hand side of condition (12.AA.2) gives player i's one-period gain from deviating in period t, and the right-hand side gives player i's discounted future losses from reversion to the Nash equilibrium starting in period $t + 1$.

For stationary outcome paths of the sort considered in Section 12.D [where each player i takes the same action q_i in every period, so that $Q = (q_1, q_2), (q_1, q_2), \ldots$], the infinite set of inequalities that must be checked in condition (12.AA.2) reduce to just two: infinite repetition of (q_1, q_2) is an outcome path of an SPNE that uses Nash reversion if and only if, for $i = 1$ and 2,

$$\hat{\pi}_i(q_j) - \pi_i(q_1, q_2) \leq \frac{\delta}{1-\delta}[\pi_i(q_1, q_2) - \pi_i(q_1^*, q_2^*)]. \tag{12.AA.3}$$

How much better than the static Nash equilibrium outcome $q^* = (q_1^*, q_2^*)$ can the players do using Nash reversion? First, under relatively mild conditions (which the Bertrand game considered in Section 12.D does not satisfy), the players can sustain a stationary outcome path that has strictly higher discounted payoffs than does infinite repetition of $q^* = (q_1^*, q_2^*)$ as long as $\delta > 0$. This fact is developed formally in Proposition 12.AA.1.

Proposition 12.AA.1: Consider an infinitely repeated game with $\delta > 0$ and $S_i \subset \mathbb{R}$ for $i = 1, 2$. Suppose also that $\pi_i(q)$ is differentiable at $q^* = (q_1^*, q_2^*)$, with $\partial\pi_i(q_i^*, q_j^*)/\partial q_j \neq 0$ for $j \neq i$ and $i = 1, 2$. Then there is some $q' = (q_1', q_2')$, with $[\pi_1(q'), \pi_2(q')] \gg [\pi_1(q^*), \pi_2(q^*)]$ whose infinite repetition is the outcome path of an SPNE that uses Nash reversion.

Proof: At $q = (q_1^*, q_2^*)$, condition (12.AA.3) holds with equality. Consider a differential change in q, (dq_1, dq_2), such that $[\partial\pi_i(q_i^*, q_j^*)/\partial q_j]\, dq_j > 0$ for $i = 1, 2$. The differential change in firm i's profits from this change is

$$\begin{aligned} d\pi_i(q_i^*, q_j^*) &= \frac{\partial\pi_i(q_i^*, q_j^*)}{\partial q_i} dq_i + \frac{\partial\pi_i(q_i^*, q_j^*)}{\partial q_j} dq_j \\ &= \frac{\partial\pi_i(q_i^*, q_j^*)}{\partial q_j} dq_j, \end{aligned} \tag{12.AA.4}$$

since q_i^* is a best response to q_j^*. Thus,

$$d\pi_i(q_i^*, q_j^*) > 0. \tag{12.AA.5}$$

On the other hand, the envelope theorem (see Section M.L of the Mathematical Appendix) tells us that at any q_j

$$d\hat{\pi}_i(q_j) = \frac{\partial \pi_i(b_i(q_j), q_j)}{\partial q_j} dq_j,$$

where $b_i(\cdot)$ is player i's best response to q_j in the stage game. Hence,

$$d\hat{\pi}_i(q_j^*) = \frac{\partial \pi_i(q_i^*, q_j^*)}{\partial q_j} dq_j. \tag{12.AA.6}$$

Together, (12.AA.4) and (12.AA.6) imply that, to first order, the value of the left-hand side of condition (12.AA.3) is unaffected by this change. However, (12.AA.5) implies that the right-hand side of (12.AA.3), to first order, increases. Hence, for a small enough change $(\Delta q_1, \Delta q_2)$ in direction (dq_1, dq_2), infinite repetition of $(q_1 + \Delta q_1, q_2 + \Delta q_2)$ is sustainable as the outcome path of an SPNE using Nash reversion strategies and, by (12.AA.5), yields strictly higher discounted payoffs to the two players than does infinite repetition of $q^* = (q_1^*, q_2^*)$. ■

Proposition 12.AA.1 tells us that with continuous strategy sets and differentiable payoff functions, as long as there is some possibility for a joint improvement in payoffs around the stage game Nash equilibrium, some cooperation can be sustained.

Going further, examination of condition (12.AA.2) tells us that cooperation becomes easier as δ grows.

Proposition 12.AA.2: Suppose that outcome path Q can be sustained as an SPNE outcome path using Nash reversion when the discount rate is δ. Then it can be so sustained for any $\delta' \geq \delta$.

In fact, as δ gets very large, a great number of outcomes become sustainable. The result presented in Proposition 12.AA.3, a version of the *Nash reversion folk theorem* [originally due to Friedman (1971)], shows that *any* stationary outcome path that gives each player a discounted payoff that exceeds that arising from infinite repetition of the stage game Nash equilibrium $q^* = (q_1^*, q_2^*)$ can be sustained as an SPNE if δ is sufficiently close to 1.

Proposition 12.AA.3: For any pair of actions $q = (q_1, q_2)$ such that $\pi_i(q_1, q_2) > \pi_i(q_1^*, q_2^*)$ for $i = 1, 2$, there exists a $\underline{\delta} < 1$ such that, for all $\delta > \underline{\delta}$, infinite repetition of $q = (q_1, q_2)$ is the outcome path of an SPNE using Nash reversion strategies.

The proof of Proposition 12.AA.3 follows immediately from condition (12.AA.3) letting $\delta \to 1$. In fact, with a more sophisticated argument, the logic of Proposition 12.AA.3 can be extended to nonstationary outcome paths. By doing so, it is possible to convexify the set of possible payoffs identified in Proposition 12.AA.3 by alternating between various action pairs (q_1, q_2). In this way, we can support any payoffs in the shaded region of Figure 12.AA.1 as the average payoffs of an SPNE.[32]

Exercise 12.AA.1: Argue that no pair of actions q such that $\pi_i(q_1, q_2) < \pi_i(q_1^*, q_2^*)$ for some i can be sustained as a stationary SPNE outcome path using Nash reversion.

More Severe Punishments and the Folk Theorem

It is intuitively clear that, for a given level of $\delta < 1$, the more severe the punishments that can be credibly threatened in response to a deviation, the easier it is to prevent

32. See Fudenberg and Maskin (1991) for details.

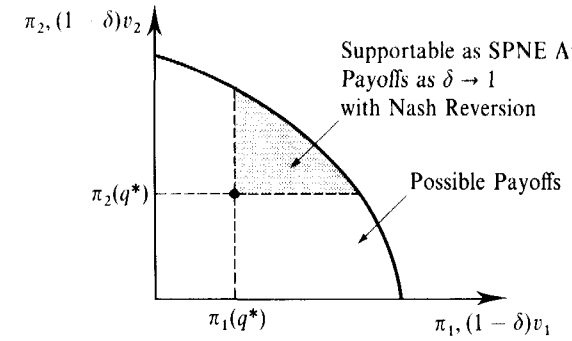


Figure 12.AA.1
The Nash reversion folk theorem.

players from deviating from any given outcome path. In general, Nash reversion is not the most severe credible punishment that is possible. Just as players can be induced to cooperate through the use of threatened punishments, they can also be induced to punish each other.

To consider this issue, it is useful to let $\underline{\pi}_i = \text{Min}_{q_j} [\text{Max}_{q_i} \pi_i(q_i, q_j)]$ denote player i's *minimax payoff*.[33] Payoff $\underline{\pi}_i$ is the lowest payoff that player i's rival can hold him to in the stage game if player i anticipates the action that his rival will play. Note, first, that player i's payoff in the stage game Nash equilibrium $q^* = (q_1^*, q_2^*)$ cannot be below $\underline{\pi}_i$. More importantly, regardless of the strategies played by his rival, player i's average payoff in the infinitely repeated game or in any subgame within it cannot be below $\underline{\pi}_i$. Thus, no punishment following a deviation can give player i an average payoff below $\underline{\pi}_i$. Payoffs that strictly exceed $\underline{\pi}_i$ for each player i are known as *individually rational payoffs*.

Note that for a punishment to be credible we must be sure that after an initial deviation occurs and the punishment is called for, no player wants to deviate from the prescribed punishment path. This means that a punishment is credible if and only if it itself constitutes an SPNE outcome path. Proposition 12.AA.4 tells us that as long as $\delta > 0$ and conditions similar to those in Proposition 12.AA.1 hold, SPNEs that yield more severe punishments than Nash reversion can be constructed whenever each player i's stage game Nash equilibrium payoff strictly exceeds $\underline{\pi}_i$. (You are asked to prove this result in Exercise 12.AA.2.)

Proposition 12.AA.4: Consider an infinitely repeated game with $\delta > 0$ and $S_i \subset \mathbb{R}$ for $i = 1, 2$. Suppose also that $\pi_i(q)$ is differentiable at $q^* = (q_1^*, q_2^*)$, with $\partial \pi_i(q_i^*, q_j^*)/\partial q_j \neq 0$ for $j \neq i$ and $i = 1, 2$, and that $\pi_i(q_1^*, q_2^*) > \underline{\pi}_i$ for $i = 1, 2$. Then there is some SPNE with discounted payoffs to the two players of (v_1', v_2') such that $(1 - \delta)v_i' < \pi_i(q_1^*, q_2^*)$ for $i = 1, 2$.

Under the conditions of Proposition 12.AA.4, for any $\delta \in (0, 1)$, more severe punishments than Nash reversion can credibly be threatened. We should therefore expect that more cooperative outcomes can be sustained than those sustainable through the threat of Nash reversion whenever a fully cooperative outcome is not already achievable using Nash reversion strategies.

33. In general, a player's minimax payoff will be lower if mixed strategies are allowed. In this case, the statement of the folk theorem given in Proposition 12.AA.5 remains unchanged, but with these (potentially) lower levels of $\underline{\pi}_i$.

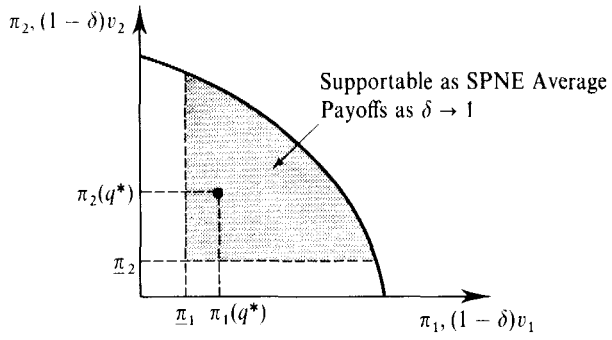


Figure 12.AA.2
The folk theorem.

For arbitrary $\delta < 1$, constructing the full set of SPNEs is a delicate process. Each SPNE, whether collusive or punishing, uses other SPNEs as threatened punishments. For details on how this is done, see the original contributions by Abreu (1986) and (1988) and the presentation in Fudenberg and Tirole (1992). As with SPNEs using Nash reversion strategies, the full set of SPNEs grows as δ increases, making possible both more cooperation and more severe punishments. In fact, the result presented in Proposition 12.AA.5, known as the *folk theorem*, tells us that *any* feasible individually rational payoffs can be supported as the average payoffs in an SPNE as long as players discount the future to a sufficiently small degree.[34] (Feasibility simply means that there is some outcome path Q that generates these average payoffs.)

Proposition 12.AA.5: (*The Folk Theorem*) For any feasible pair of individually rational payoffs $(\pi_1, \pi_2) \gg (\underline{\pi}_1, \underline{\pi}_2)$, there exists a $\underline{\delta} < 1$ such that, for all $\delta > \underline{\delta}$, (π_1, π_2) are the average payoffs arising in an SPNE.

In comparison with Proposition 12.AA.3, Proposition 12.AA.5 tells us that as $\delta \to 1$ we can support any average payoffs that exceed each player's minimax payoff.[35] This limiting set of SPNE average payoffs is shown in Figure 12.AA.2.

Example 12.AA.1 gives some idea of how this can be done.

Example 12.AA.1: *Sustaining an Average Payoff of Zero in the Infinitely Repeated Cournot Game.* In this example, we construct an SPNE in which both firms earn an average payoff of zero in an infinitely repeated Cournot game. In particular, let the stage game be a symmetric Cournot duopoly game with cost function $c(q) = cq$, where $c > 0$, and a continuous inverse demand function $p(\cdot)$ such that $p(x) \to 0$ as $x \to \infty$. It will be convenient to write a firm's profit when both firms choose quantity q as $\pi(q) = [p(2q) - c]q$ and, as before, a firm's best-response profits when its rival

34. The theorem's name refers to the fact that some version of the result was known in game theory "folk wisdom" well before its formal appearance in the literature. See Fudenberg and Maskin (1986) and (1991) for a proof of the result. When there are more than two players, the result requires that the set of feasible payoffs satisfy an additional "dimensionality" condition. The original appearances of the result in the literature actually analyzed infinitely repeated games *without* discounting [see, for example, Rubinstein (1979)].

35. We may also be able in some cases to give each player exactly his minimax payoff. This is the case, for example, in the repeated Bertrand game, where the stage game's Nash equilibrium yields the minimax payoffs. In Example 12.AA.1, we show that we can also do this for large enough δ in the repeated Cournot duopoly game.

chooses quantity q as $\hat{\pi}(q)$.[36] Note that $\underline{\pi}_j = 0$ for $j = 1, 2$ here; if firm j's rival chooses a quantity at least as large as the competitive quantity q_c satisfying $p(q_c) = c$, then the best firm j can do is to produce nothing and earn zero, and firm j can never be forced to a payoff worse than zero.

Consider strategies for the players that take the following form:

(i) Both firms play quantity $\tilde{q}$ in period 1 followed by the monopoly quantity q^m in every period $t > 1$ as long as no one deviates, where quantity $\tilde{q}$ satisfies

$$\pi(\tilde{q}) + \frac{\delta}{1-\delta}\pi(q^m) = 0. \tag{12.AA.7}$$

(ii) If anyone deviates when $\tilde{q}$ is meant to be played, the outcome path described in (i) is restarted.

(iii) If anyone deviates when q^m is meant to be played, Nash reversion occurs.

Note that the outcome path described in (i), if followed by both players, gives both players an average payoff of zero by construction [recall (12.AA.7)].

By Proposition 12.AA.3, we know that for some $\underline{\delta} < 1$ we can sustain infinite repetition of q^m through Nash reversion for all $\delta > \underline{\delta}$. Thus, for $\delta > \underline{\delta}$, neither firm will deviate from the above strategies when q^m is supposed to be played. Will they deviate when $\tilde{q}$ is supposed to be played? Consider firm j's payoff from deviating from $\tilde{q}$ in a single period and conforming with the prescribed strategy thereafter. Firm j earns $\hat{\pi}(\tilde{q}) + (\delta)(0)$ because it plays a best response when deviating, and then the original path is restarted. Thus, this deviation does not improve firm j's payoff if $\hat{\pi}(\tilde{q}) = 0$ (it cannot be less than zero because $\underline{\pi}_i = 0$). This is so if $\tilde{q} \geq q_c$. But examining condition (12.AA.7), we see that as δ approaches 1, $\pi(\tilde{q})$ must get increasingly negative for (12.AA.7) to hold and, in particular, that there exists a $\delta_c < 1$ such that $\tilde{q}$ will exceed q_c for all $\delta > \delta_c$. Thus, for $\delta > \text{Max}\{\delta_c, \underline{\delta}\}$, these strategies constitute an SPNE that gives both firms an average payoff of 0.[37] ■

APPENDIX B: STRATEGIC ENTRY DETERRENCE AND ACCOMMODATION

In this appendix, we discuss an important example of credible precommitments to affect future market conditions in which an incumbent firm engages in pre-entry capacity expansion to gain a strategic advantage over a potential entrant and possibly to deter this firm's entry altogether [the original analyses of this issue are due to Spence (1977) and Dixit (1980)]. In what follows, we study the following three-stage game that is adapted from Dixit (1980).

36. We can make the strategy sets compact by noting that in no period will any firm ever choose a quantity larger than the level $\bar{q}$ such that $\pi(\bar{q}) + [\delta/(1-\delta)](\text{Max}_q\, \pi(q)) = 0$, because it would do better setting its quantity equal to zero forever. Then, without loss, we can let each firm choose its output from the compact set $[0, \bar{q}]$.

37. We have not considered any multiperiod deviations, but it can be shown that if no single-period deviation followed by conformity with the strategies is worthwhile, then neither is any multiperiod deviation (this is a general principle of dynamic programming).

Stage 1: An incumbent, firm I, chooses the capacity level of its plant, denoted by k_I. Capacity costs r per unit.

Stage 2: A potential entrant, firm E, decides whether to enter the market. If it does, it pays an entry cost of F.

Stage 3: If firm E enters, the two firms choose their output levels, q_I and q_E, simultaneously. The resulting price is $p(q_I + q_E)$. For firm E, output costs $(w + r)$ per unit: for each unit of output produced, firm E incurs both a capacity cost of r and a labor cost of w. For firm I, production must not exceed its previously chosen capacity level. Its production cost, however, is only w per unit because it has already built its capacity. If, on the other hand, firm E does not enter, then firm I acts as a monopolist who can produce up to k_I units of output at cost w per unit.

To determine the subgame perfect Nash equilibrium (SPNE) of this game, we begin by analyzing behavior in the stage 3 subgames and then work backward.

Stage 3: Quantity Competition

The subgames in stage 3 are distinguished by two previous events: whether firm E has entered and the previous capacity choice of firm I. We first consider the outcome of stage 3 competition following entry and then discuss firm I's behavior in stage 3 if entry does not occur. For simplicity, we assume throughout that firms' profit functions are strictly concave in own quantity; a sufficient condition for this is for $p(\cdot)$ to be concave. The concavity of $p(\cdot)$ also implies that firms' best-response functions are downward sloping.

Stage 3 competition after entry. Figure 12.BB.1 depicts firm E's best-response function in stage 3, which we denote by $b(q|w + r)$ to emphasize that it is the best-response function for a firm with marginal cost $w + r$. Firm E's stage 3 profits decline as we move along this curve to the right (involving higher levels of q_I) and, at some point, denoted Z in the figure, they fall below the entry cost F.

Now consider firm I's optimal behavior. The key difference between firm I and firm E is that firm I has already built its capacity. Hence, firm I's expenditure on this capacity is sunk (it cannot recover it by reducing its capacity), its capacity level is fixed, and its marginal cost is only w. Suppose we let $b(q|w)$ denote the best-response function of a firm with marginal cost w. Then firm I's best-response function in stage 3 is

$$b_I(q_E|k_I) = \text{Min}\{b(q_E|w), k_I\}.$$

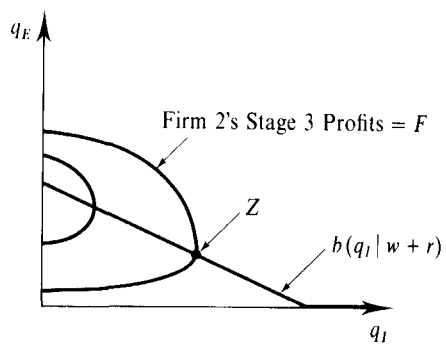
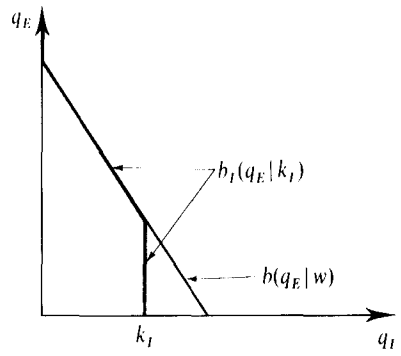


Figure 12.BB.1
Firm E's stage 3 best-response function after entry.



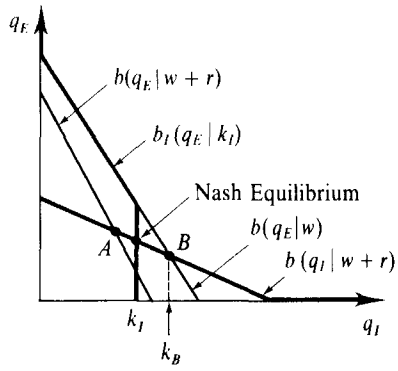


Figure 12.BB.2 (left)
Firm I's stage 3 best-response function after entry.

Figure 12.BB.3 (right)
Stage 3 Nash equilibrium after entry.

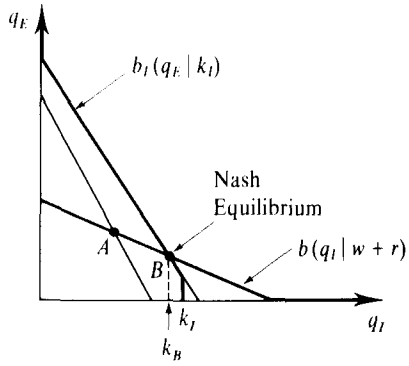


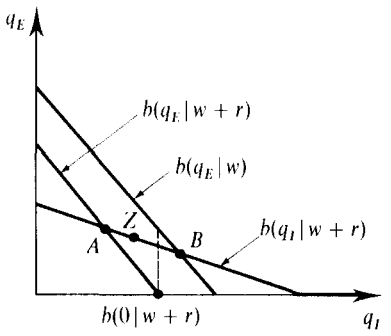
Figure 12.BB.4
A stage 3 equilibrium in which firm I does not use all of its capacity.

That is, firm I's best response to an output choice of q_E by firm E is the same as that for a firm with marginal cost level w as long as this output level does not exceed its previously chosen capacity. Figure 12.BB.2 illustrates firm I's best-response function.

We can now put together the best-response functions for the two firms to determine the equilibrium in stage 3 following firm E's decision to enter, for any given level of k_I. This equilibrium is shown in Figure 12.BB.3.

In Figure 12.BB.3, point A is the outcome that would arise if there were no first-mover advantage for firm I, that is, if the two firms chose both their capacity and output levels simultaneously. However, when firm I is able to choose its capacity level first, by choosing an appropriate level of k_I, it can get the post-entry equilibrium to lie anywhere on firm E's best-response function up to point B. Firm I is able to induce points to the right of point A because its ability to incur its capacity costs prior to stage 3 competition allows it to have a marginal cost in stage 3 of only w, rather than $w + r$. Note, however, that firm I cannot induce a point on firm 2's best-response function beyond point B, even though it might want to; if it built a capacity greater than level k_B, it would not have an incentive to actually use all of it. Figure 12.BB.4 depicts this situation. A threat to produce up to capacity following entry would in this case not be credible.

Stage 3 outcomes if firm E does not enter. If firm E decides not to enter, then firm I will be a monopolist in stage 3. Its optimal monopoly output is then the point where its best-response function hits the $q_E = 0$ axis, $b_I(0|k_I)$.



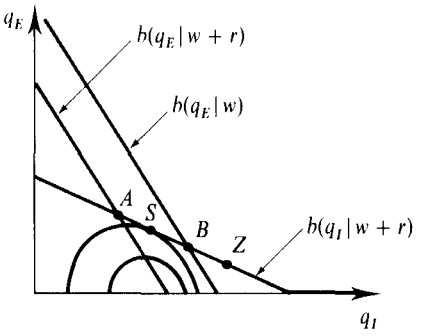


Figure 12.BB.5 (left)
Blockaded entry.

Figure 12.BB.6 (right)
Strategic entry accommodation when entry is inevitable.

Stage 2: Firm E's Entry Decision

Firm E's entry decision is straightforward: Given the level of capacity k_I chosen by firm I in stage 1, firm E will enter if it expects nonnegative profits net of its entry cost F. This means that firm E will enter when it expects that the post-entry equilibrium will lie to the left of point Z on its best-response function in Figure 12.BB.1.

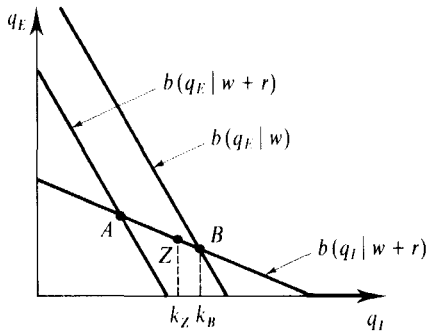
Stage 1: Firm I's Stage 1 Capacity Investment

Now consider firm I's optimal capacity choice in stage 1. There are three situations in which firm I could find itself: Entry could be blockaded, entry could be inevitable, or entry deterrence could be possible but not inevitable. Let us consider each in turn.

Entry is blockaded. One possibility is that the entry cost F is large enough that firm E does not find it worthwhile to enter even if firm I ignores the possibility of entry and simply builds the same capacity that it would if it were an uncontested monopolist, $b(0|w+r)$. This situation, in which we say that *entry is blockaded*, is shown in Figure 12.BB.5. In this case, firm I achieves its best possible outcome: it builds a capacity of $b(0|w+r)$, no entry occurs, and then it sells $b(0|w+r)$ units of output.

Entry deterrence is impossible: strategic entry accommodation. Suppose that point Z is to the right of point B. In this case, entry deterrence is impossible; firm E will find it profitable to enter regardless of k_I. What is firm I's optimal choice of k_I in this case? In Figure 12.BB.6, we have drawn isoprofit curves for firm I; note that because these include the cost of capacity, they are the isoprofit curves corresponding to those of a firm with marginal cost $(w+r)$. Now recall that firm I can induce any point on firm E's best-response function up to point B through an appropriate choice of capacity. It will choose the point that maximizes its profit. In Figure 12.BB.6, this point, which involves a tangency between firm E's best-response function and firm I's isoprofit curves, is denoted as point S. This outcome corresponds to exactly the outcome that would emerge in a model of sequential quantity choice, known as a *Stackleberg leadership model* (see Exercise 12.C.18). Note that firm I's first-mover advantage allows it to earn higher profits than the otherwise identical firm E.

The point of tangency, S, could also lie to the right of point B. In this case, the optimal capacity choice will be $k_I = k_B$, and the outcome will not be as desirable for firm I as the Stackleberg point. Here firm I is unable to credibly



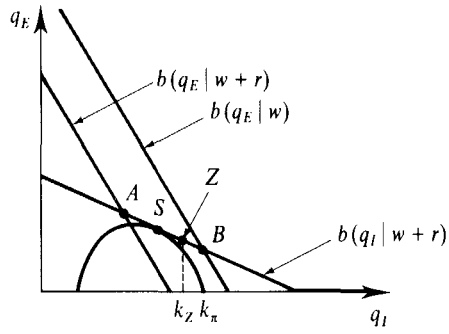


Figure 12.BB.7 (left)
Entry deterrence is possible but not inevitable.

Figure 12.BB.8 (right)
Entry deterrence versus entry accommodation.

commit to produce the output associated with point S, even if it builds sufficient capacity in stage 1.

Entry deterrence is possible but not inevitable. Suppose now that point Z lies to the left of point B but not so far that entry is blockaded, as shown in Figure 12.BB.7. Firm I can deter firm E's entry by picking a capacity level at least as large as point k_Z in the figure. The only question is whether this will be optimal for firm I, or whether firm I is better off accommodating firm E's entry. To judge this, firm I will compare its profits at point $(k_Z, 0)$ to those at point S (or at point B if point S lies to the right of B). This can be done by comparing the capacity level k_π in Figure 12.BB.8, the output level under monopoly that gives the same profit as the optimal accommodation point S, with k_Z. If $k_\pi > k_Z$, then firm I prefers to deter entry because its profits are higher in this case; but if $k_\pi < k_Z$, then it will prefer accommodation. Note that if deterrence is optimal, then even though entry does not occur its *threat* nevertheless has an effect on the market outcome, raising the level of output and welfare relative to a situation in which no entry is possible.

Exercise 12.BB.1: Show that when entry deterrence is possible but not inevitable, if point S lies to the right of point Z, then entry deterrence is better than entry accommodation.

REFERENCES

Abreu, D. (1986). Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory* **39**: 191–225.

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica* **56**: 383–96.

Abreu, D., D. Pearce, and E. Stachetti. (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica* **58**: 1041–64.

Baumol, W., J. Panzar, and R. Willig. (1982). *Contestable Markets and the Theory of Industry Structure.* San Diego: Harcourt, Brace, Jovanovich.

Bertrand, J. (1883). Théorie mathématique de la richesse sociale. *Journal des Savants* **67**: 499–508.

Bulow, J., J. Geanakoplos, and P. Klemperer. (1985). Multimarket oligopoly: strategic substitutes and complements. *Journal of Political Economy* **93**: 488–511.

Chamberlin, E. (1933). *The Theory of Monopolistic Competition.* Cambridge, Mass.: Harvard University Press.

Cournot, A. (1838). *Recherches sur les Principes Mathématiques de la Théorie des Richesses.* [English edition: *Researches into the Mathematical Principles of the Theory of Wealth*, edited by N. Bacon. London: Macmillan, 1897.]

Dixit, A. (1980). The role of investment in entry deterrence. *Economic Journal* **90**: 95–106.

Dixit, A., and J. E. Stiglitz. (1977). Monopolistic competition and optimal product diversity. *American Economic Review* **67**: 297–308.

Edgeworth, F. (1897). Me teoria pura del monopolio. *Giornale degli Economisti* **40**: 13 31. [English translation: The pure theory of monopoly. In *Papers Relating to Political Economy*, Vol. I, edited by F. Edgeworth. London: Macmillan, 1925.]

Friedman, J. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies* **28**: 1–12.

Fudenberg, D., and E. Maskin. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* **52**: 533–54.

Fudenberg, D., and E. Maskin. (1991). On the dispensibility of public randomizaton in discounted repeated games. *Journal of Economic Theory* **53**: 428–38.

Fudenberg, D., and J. Tirole. (1984). The fat cat effect, the puppy dog ploy, and the lean and hungry look. *American Economic Review, Papers and Proceedings* **74**: 361–68.

Fudenberg, D., and J. Tirole. (1992). *Game Theory.* Cambridge, Mass.: MIT Press.

Green, E., and R. Porter. (1984). Noncooperative collusion under imperfect price information. *Econometrica* **52**: 87–100.

Hart, O. D. (1985). Monopolistic competition in the spirit of Chamberlin: A general model. *Review of Economic Studies* **52**: 529–46.

Kreps, D. M., and J. Scheinkman. (1983). Quantity precommitment and Bertrand competition yield Cournot outcomes. *Rand Journal of Economics* **14**: 326–37.

Mankiw, N. G., and M. D. Whinston. (1986). Free entry and social inefficiency. *Rand Journal of Economics* **17**: 48 58.

Osborne, M. J., and A. Rubinstein. (1994). *A Course in Game Theory.* Cambridge, Mass.: MIT Press.

Rotemberg, J., and G. Saloner. (1986). A supergame-theoretic model of business cycles and price wars during booms. *American Economic Review* **76**: 390–407.

Rubinstein, A. (1979). Equilibrium in supergames with the overtaking criterion. *Journal of Economic Theory* **21**: 1 9.

Salop, S. (1979). Monopolistic competition with outside goods. *Bell Journal of Economics* **10**: 141–56.

Shapiro, C. (1989). Theories of oligopoly behavior. In *Handbook of Industrial Organization*, edited by R. Schmalensee and R. D. Willig. Amsterdam: North-Holland.

Spence, A. M. (1976). Product selection, fixed costs, and monopolistic competition. *Review of Economic Studies* **43**: 217 35.

Spence, A. M. (1977). Entry, capacity investment, and oligopolistic pricing. *Bell Journal of Economics* **8**: 534 44.

Stigler, G. (1960). A theory of oligopoly. *Journal of Political Economy* **72**: 44–61.

Tirole, J. (1988). *The Theory of Industrial Organization.* Cambridge, Mass.: MIT Press.

EXERCISES

12.B.1[A] The expression $[p^m - c'(q^m)]/p^m$, where p^m and q^m are the monopolist's price and output level, respectively, is known as the monopolist's *price–cost margin* (or as the *Lerner index of monopoly power*). It measures the distortion of the monopolist's price above its marginal cost as a proportion of its price.

(a) Show the monopolist's price–cost margin is always equal to the inverse of the price elasticity of demand at price p^m.

(b) Also argue that if the monopolist's marginal cost is positive at every output level, then demand must be *elastic* (i.e., the price elasticity of demand is greater than 1) at the monopolist's optimal price.

12.B.2[B] Consider a monopolist with cost function $c(q) = cq$, with $c > 0$, facing demand function $x(p) = \alpha p^{-\varepsilon}$, where $\varepsilon > 0$.

(a) Show that if $\varepsilon \leq 1$, then the monopolist's optimal price is not well defined.

(b) Assume that $\varepsilon > 1$. Derive the monopolist's optimal price, quantity, and price–cost margin $(p^m - c)/p^m$. Calculate the resulting deadweight welfare loss.

(c) (Harder) Consider a sequence of demand functions that differ in their levels of ε and α but that all involve the same competitive quantity $x(c)$ [i.e., for each level of ε, α is adjusted to keep $x(c)$ the same]. How does the deadweight loss vary with ε? (If you cannot derive an analytic answer, try calculating some values on a computer.)

12.B.3[B] Suppose that we consider a monopolist facing demand function $x(p, \theta)$ with cost function $c(q, \phi)$, where θ and ϕ are parameters. Use the implicit function theorem to compute the changes in the monopolist's price and quantity as a function of a differential change in either θ or ϕ. When will each lead to a price increase?

12.B.4[B] Consider a monopolist with a cost of c per unit. Use a "revealed preference" proof to show that the monopoly price is nondecreasing in c. Then extend your argument to the case in which the monopolist's cost function is $c(q, \phi)$, with $[c(q'', \phi) - c(q', \phi)]$ increasing in ϕ for all $q'' > q'$, by showing that the monopoly price is nondecreasing in ϕ. (If you did Exercise 12.B.3, also relate this condition to the one you derived there.)

12.B.5[B] Suppose that a monopolist faces many consumers. Argue that in each of the following two cases, the monopolist can do no better than it does by restricting itself to simply charging a price per unit, say p.

(a) Suppose that each consumer i wants either one or no units of the monopolist's good and that the monopolist is unable to discern any particular consumer's preferences.

(b) Suppose that consumers may desire to consume multiple units of the good. The monopolist cannot discern any particular consumer's preferences. In addition, resale of the good is costless and after the monopolist has made its sales to consumers a competitive market develops among consumers for the good.

12.B.6[A] Suppose that the government can tax or subsidize a monopolist who faces inverse demand function $p(q)$ and has cost function $c(q)$ [assume both are differentiable and that $p(q)q - c(q)$ is concave in q]. What tax or subsidy per unit of output would lead the monopolist to act efficiently?

12.B.7[B] Consider the widget market. The total demand by men for widgets is given by $x_m(p) = a - \theta_m p$, and the total demand by women is given by $x_w(p) = a - \theta_w p$, where $\theta_w < \theta_m$. The cost of production is c per widget.

(a) Suppose the widget market is competitive. Find the equilibrium price and quantity sold.

(b) Suppose, instead, that firm A is a monopolist of widgets [also make this assumption in **(c)** and **(d)**]. If firm A is prohibited from "discriminating" (i.e., charging different prices to men and women), what is its profit-maximizing price? Under what conditions do both men and women consume a positive level of widgets in this solution?

(c) If firm A has produced some total level of output X, what is the welfare-maximizing way to distribute it between the men and the women? (Assume here and below that Marshallian aggregate surplus is a valid measure of welfare.)

(d) Suppose that firm A is allowed to discriminate. What prices does it charge? In the case where the nondiscriminatory solution in **(b)** has positive consumption of widgets by both men and women, does aggregate welfare as measured by the Marshallian aggregate surplus rise or

fall relative to when discrimination is allowed? Relate your conclusion to your answer in **(c)**. What if the nondiscriminatory solution in **(b)** has only one type of consumers being served?

12.B.8[B] Consider the following two-period model: A firm is a monopolist in a market with an inverse demand function (in each period) of $p(q) = a - bq$. The cost per unit in period 1 is c_1. In period 2, however, the monopolist has "learned by doing," and so its constant cost per unit of output is $c_2 = c_1 - mq_1$, where q_1 is the monopolist's period 1 output level. Assume $a > c$ and $b > m$. Also assume that the monopolist does not discount future earnings.

(a) What is the monopolist's level of output in each of the periods?

(b) What outcome would be implemented by a benevolent social planner who fully controlled the monopolist? Is there any sense in which the planner's period 1 output is selected so that "price equals marginal cost"?

(c) Given that the monopolist will be selecting the period 2 output level, would the planner like the monopolist to slightly increase the level of period 1 output above that identified in **(a)**? Can you give any intuition for this?

12.B.9[C] Consider a situation in which there is a monopolist in a market with inverse demand function $p(q)$. The monopolist makes two choices: How much to invest in cost reduction, I, and how much to sell, q. If the monopolist invests I in cost reduction, his (constant) per-unit cost of production is $c(I)$. Assume that $c'(I) < 0$ and that $c''(I) > 0$. Assume throughout that the monopolist's objective function is concave in q and I.

(a) Derive the first-order conditions for the monopolist's choices.

(b) Compare the monopolist's choices with those of a benevolent social planner who can control both q and I (a "first-best" comparison).

(c) Compare the monopolist's choices with those of a benevolent social planner who can control I but not q (a "second-best" comparison). Suppose that the planner chooses I and then the monopolist chooses q.

12.B.10[B] Consider a monopolist that can choose both its product's price p and its quality q. The demand for its product is given by $x(p, q)$, which is increasing in q and decreasing in p. Given the price chosen by the monopolist, does the monopolist choose the socially efficient quality level?

12.C.1[A] In text.

12.C.2[C] Extend the argument of Proposition 12.C.1 to show that under the assumptions made in the text [in particular, the assumption that there is a price $\bar{p} < \infty$ such that $x(p) = 0$ for all $p \geq \bar{p}$], both firms setting their price equal to c with certainty is the unique Nash equilibrium of the Bertrand duopoly model even when we allow for mixed strategies.

12.C.3[B] Note that the unique Nash equilibrium of the Bertrand duopoly model has each firm playing a weakly dominated strategy. Consider an alteration of the model in which prices must be named in some discrete unit of account (e.g., pennies) of size Δ.

(a) Show that both firms naming prices equal to the smallest multiple of Δ that is strictly greater than c is a pure strategy equilibrium of this game. Argue that it does not involve either firm playing a weakly dominated strategy.

(b) Argue that as $\Delta \to 0$, this equilibrium converges to both firms charging prices equal to c.

12.C.4[B] Consider altering the Bertrand duopoly model to a case in which each firm j's cost per unit is c_j and $c_1 < c_2$.

(a) What are the pure strategy Nash equilibria of this game?

(b) Examine a model in which prices must be named in discrete units, as in Exercise 12.C.3. What are the pure strategy Nash equilibria of such a game? Which do not involve the play of weakly dominated strategies? As the grid becomes finer, what is the limit of these equilibria in undominated strategies?

12.C.5[B] Suppose that we have a market with I buyers, each of whom wants at most one unit of the good. Buyer i is willing to pay up to v_i for his unit, and $v_1 > v_2 > \cdots > v_I$. There are a total of $q < I$ units available. Suppose that buyers simultaneously submit bids for a unit of the output and that the output goes to the q highest bidders, who pay the amounts of their bids. Show that every buyer making a bid of v_{q+1} and the good being assigned to buyers $1, \ldots, q$ is a Nash equilibrium of this game. Argue that this is a competitive equilibrium price. Also show that in *any* pure strategy Nash equilibrium of this game, buyers 1 through q receive a unit and buyers $q + 1$ through I do not.

12.C.6[A] In text.

12.C.7[B] In text.

12.C.8[C] Consider a homogeneous-good J-firm Cournot model in which the demand function $x(p)$ is downward sloping but otherwise arbitrary. The firms all have an identical cost function $c(q)$ that is increasing in q and convex. Denote by Q the aggregate output of the J firms, and let $Q_{-j} = \sum_{k \neq j} q_k$.

(a) Show that firm j's best response can be written as $b(Q_{-j})$.

(b) Show that $b(Q_{-j})$ need not be unique (i.e., that it is in general a correspondence, not a function).

(c) Show that if $\hat{Q}_{-j} > Q_{-j}$, $q_j \in b(Q_{-j})$, and $\hat{q}_j \in b(\hat{Q}_{-j})$, then $(\hat{q}_j + \hat{Q}_{-j}) \geq (q_j + Q_{-j})$. Deduce from this that $b(\cdot)$ can jump only upward and that $b'(Q_{-j}) \geq -1$ whenever this derivative is defined.

(d) Use you result in **(c)** to prove that a symmetric pure strategy Nash equilibrium exists in this model.

(e) Show that multiple equilibria are possible.

(f) Give sufficient conditions (they are very weak) for the symmetric equilibrium to be the only equilibrium in pure strategies.

12.C.9[B] Consider a two-firm Cournot model with constant returns to scale but in which firms' costs may differ. Let c_j denote firm j's cost per unit of output produced, and assume that $c_1 > c_2$. Assume also that the inverse demand function is $p(q) = a - bq$, with $a > c_1$.

(a) Derive the Nash equilibrium of this model. Under what conditions does it involve only one firm producing? Which will this be?

(b) When the equilibrium involves both firms producing, how do equilibrium outputs and profits vary when firm 1's cost changes?

(c) Now consider the general case of J firms. Show that the ratio of industry profits divided by industry revenue in any (pure strategy) Nash equilibrium is exactly H/ε, where ε is the elasticity of the market demand curve at the equilibrium price and H, the *Herfindahl index of concentration*, is equal to the sum of the firms' squared market shares $\sum_j (q_j^*/Q^*)^2$. (*Note:* This result depends on the assumption of constant returns to scale.)

12.C.10[B] Consider a J-firm Cournot model in which firms' costs differ. Let $c_j(q_j) = \alpha_j \tilde{c}(q_j)$ denote firm j's cost function, and assume that $\tilde{c}(\cdot)$ is strictly increasing and convex. Assume that $\alpha_1 > \cdots > \alpha_J$.

(a) Show that if more than one firm is making positive sales in a Nash equilibrium of this model, then we cannot have productive efficiency; that is, the equilibrium aggregate output Q^* is produced inefficiently.

(b) If so, what is the correct measure of welfare loss relative to a fully efficient (competitive) outcome? [*Hint:* Reconsider the discussion in Section 10.E.]

(c) Provide an example in which welfare decreases when a firm becomes more productive (i.e., when α_j falls for some j). [*Hint:* Consider an improvement in cost for firm 1 in the model of Exercise 12.C.9.] Why can this happen?

12.C.11[C] Consider a capacity-constrained duopoly pricing game. Firm j's capacity is q_j for $j = 1, 2$, and it has a constant cost per unit of output of $c \geq 0$ up to this capacity limit. Assume that the market demand function $x(p)$ is continuous and strictly decreasing at all p such that $x(p) > 0$ and that there exists a price $\tilde{p}$ such that $x(\tilde{p}) = q_1 + q_2$. Suppose also that $x(p)$ is concave. Let $p(\cdot) = x^{-1}(\cdot)$ denote the inverse demand function.

Given a pair of prices charged, sales are determined as follows: consumers try to buy at the low-priced firm first. If demand exceeds this firm's capacity, consumers are served in order of their valuations, starting with high-valuation consumers. If prices are the same, demand is split evenly unless one firm's demand exceeds its capacity, in which case the extra demand spills over to the other firm. Formally, the firms' sales are given by the functions $x_1(p_1, p_2)$ and $x_2(p_1, p_2)$ satisfying [$x_i(\cdot)$ gives the amount firm i sells taking account of its capacity limitation in fulfilling demand]

$$\text{If } p_j > p_i\text{:} \qquad x_i(p_1, p_2) = \text{Min}\{q_i, x(p_i)\}$$
$$x_j(p_1, p_2) = \text{Min}\{q_j, \text{Max}\{x(p_j) - q_i, 0\}\}$$
$$\text{If } p_2 = p_1 = p\text{:} \quad x_i(p_1, p_2) = \text{Min}\{q_i, \text{Max}\{x(p)/2, x(p) - q_j\}\} \quad \text{for } i = 1, 2.$$

(a) Suppose that $q_1 < b_c(q_2)$ and $q_2 < b_c(q_1)$, where $b_c(\cdot)$ is the best-response function for a firm with constant marginal costs of c. Show that $p_1^* = p_2^* = p(q_1 + q_2)$ is a Nash equilibrium of this game.

(b) Argue that if either $q_1 > b_c(q_2)$ or $q_2 > b_c(q_1)$, then no pure strategy Nash equilibrium exists.

12.C.12[B] Consider two strictly concave and differentiable profit functions $\pi_j(q_j, q_k)$, $j = 1, 2$, defined on $q_j \in [0, q]$.

(a) Give sufficient conditions for the best-response functions $b_j(q_j)$ to be increasing or decreasing.

(b) Specialize to the Cournot model. Argue that a decreasing (downward-sloping) best-response function is the "normal" case.

12.C.13[B] Show that when $v > c + 3t$ in the linear city model discussed in Example 12.C.2, a firm j's best response to any price of its rival p_{-j} always results in all consumers purchasing from one of the two firms.

12.C.14[C] Consider the linear city model discussed in Example 12.C.2.

(a) Derive the best-response functions when $v \in (c + 2t, c + 3t)$. Show that the unique Nash equilibrium in this case is $p_1^* = p_2^* = c + t$.

(b) Repeat **(a)** for the case in which $v \in (c + \frac{3}{2}t, c + 2t)$.

(c) Show that when $v < c + t$, the unique Nash equilibrium involves prices of $p_1^* = p_2^* = (v + c)/2$ and some consumers not purchasing from either firm.

(d) Show that when $v \in (c + t, c + \frac{3}{2}t)$, the unique symmetric equilibrium is $p_1^* = p_2^* = v - t/2$. Are there asymmetric equilibria in this case?

(e) Compare the change in equilibrium prices and profits from a reduction in t in the case studied in **(d)** with that in the equilibria of **(a)** and **(b)**.

12.C.15[B] Derive the Nash equilibrium prices of the linear city model where a consumer's travel cost is quadratic in distance, that is, where the total cost of purchasing from firm j is $p_j + td^2$, where d is the consumer's distance from firm j. Restrict attention to the case in which v is large enough that the possibility of nonpurchase can be ignored.

12.C.16[B] Derive the Nash equilibrium prices and profits in the circular city model with J firms when travel costs are quadratic, as in Exercise 12.C.15. Restrict attention to the case in which v is large enough that the possibility of nonpurchase can be ignored. What happens as J grows large? As t falls?

12.C.17[B] Consider the linear city model in which the two firms may have different constant unit production costs $c_1 > 0$ and $c_2 > 0$. Without loss of generality, take $c_1 \leq c_2$ and suppose that v is large enough that nonpurchase can be ignored. Determine the Nash equilibrium prices and sales levels for equilibria in which both firms make strictly positive sales. How do local changes in c_1 affect the equilibrium prices and profits of firms 1 and 2? For what values of c_1 and c_2 does the equilibrium involve one firm making no sales?

12.C.18[B] (*The Stackleberg leadership model*) There are two firms in a market. Firm 1 is the "leader" and picks its quantity first. Firm 2, the "follower," observes firm 1's choice and then chooses its quantity. Profits for each firm i given quantity choices q_1 and q_2 are $p(q_1 + q_2)q_i - cq_i$, where $p'(q) < 0$ and $p'(q) + p''(q)q < 0$ at all $q \geq 0$.

(a) Prove formally that firm 1's quantity choice is larger than its quantity choice would be if the firms chose quantities simultaneously and that its profits are larger as well. Also show that aggregate output is larger and that firm 2's profits are smaller.

(b) Draw a picture of this outcome using best-response functions and isoprofit contours.

12.C.19[C] Do Exercise 8.B.5.

12.C.20[B] Prove Proposition 12.C.2 for the case of a general convex cost function $c(q)$.

12.D.1[B] Consider an infinitely repeated Bertrand duopoly with discount factor $\delta < 1$. Determine the conditions under which strategies of the form in (12.D.1) sustain the monopoly price in each of the following cases:

(a) Market demand in period t is $x_t(p) = \gamma^t x(p)$ where $\gamma > 0$.

(b) At the end of each period, the market ceases to exist with probability γ.

(c) It takes K periods to respond to a deviation.

12.D.2[B] In text.

12.D.3[B] Consider an infinitely repeated Cournot duopoly with discount factor $\delta < 1$, unit costs of $c > 0$, and inverse demand function $p(q) = a - bq$, with $a > c$ and $b > 0$.

(a) Under what conditions can the symmetric joint monopoly outputs $(q_1, q_2) = (q^m/2, q^m/2)$ be sustained with strategies that call for $(q^m/2, q^m/2)$ to be played if no one has yet deviated and for the single-period Cournot (Nash) equilibrium to be played otherwise?

(b) Derive the minimal level of δ such that output levels $(q_1, q_2) = (q, q)$ with $q \in [((a - c)/2b), ((a - c)/b)]$ are sustainable through Nash reversion strategies. Show that this level of δ, $\delta(q)$, is an increasing, differentiable function of q.

12.D.4[B] Consider an infinitely repeated Bertrand oligopoly with discount factor $\delta \in [\frac{1}{2}, 1)$.

(a) If the cost of production changes, what happens to the most profitable price that can be sustained?

(b) Suppose, instead, that the cost of production will increase permanently in period 2 (i.e., from period 2 on, it will be higher than in period 1). What effect does this have on the maximal price that can be sustained in period 1?

12.D.5[C] [Based on Rotemberg and Saloner (1986)] Consider a model of infinitely repeated Bertrand interaction where in each period there is a probability $\lambda \in (0, 1)$ of a "high-demand" state in which demand is $x(p)$ and a probability $(1 - \lambda)$ of a "low-demand" state in which demand is $\alpha x(p)$, where $\alpha \in (0, 1)$. The cost of production is $c > 0$ per unit. Consider Nash reversion strategies of the following form: charge price p_H in a high-demand state if no previous deviation has occurred, charge p_L in a low-demand state if no previous deviation has occurred, and set price equal to c if a deviation has previously occurred.

(a) Show that if δ is sufficiently high, then there is an SPNE in which the firms set $p_H = p_L = p^m$, the monopoly price.

(b) Show that for some $\underline{\delta}$ above $\frac{1}{2}$, a firm will want to deviate from price p^m in the high-demand state whenever $\delta < \underline{\delta}$. Identify the highest price p_H that the firms can sustain when $\delta \in [\frac{1}{2}, \underline{\delta})$ (verify that they can still sustain price $p_L = p^m$ in the low-demand state). Notice that this equilibrium may involve "countercyclical" pricing; that is, $p_L > p_H$. Intuitively, what drives this result?

(c) Show that when $\delta < \frac{1}{2}$ we must have $p_H = p_L = c$.

12.E.1[B] Suppose that we have a two-stage model of entry into a homogeneous-good market characterized by price competition. If potential firms differ in efficiency, need the equilibrium have the most efficient firm being active?

12.E.2[B] Prove that π_J is decreasing in J under assumptions (A1) to (A3) of Proposition 12.E.1.

12.E.3[B] Calculate the welfare loss from the free-entry equilibrium number of firms relative to the socially optimal number of firms in the models discussed in Examples 12.E.1 and 12.E.2. What happens to this loss as $K \to 0$?

12.E.4[B] Consider a two-stage model of entry in which all potential entrants have a cost per unit of c (in additional to an entry cost of K) and in which, whatever number of firms enter, a perfect cartel is formed. What is the socially optimal number of firms for a planner who cannot control this cartel behavior? What are the welfare consequences if the planner cannot control entry?

12.E.5[C] Consider a two-stage entry model with a market that looks like the market in Exercise 12.C.16. The entry cost is K. Compare the equilibrium number of firms to the number that a planner would pick who can control (a) entry and pricing and (b) only entry.

12.E.6[B] Compare a one-stage and a two-stage model of entry with Cournot competition [all potential entrants are identical and production costs are $c(q) = cq$]. Argue that any (SPNE) equilibrium outcome of the two-stage game is also an outcome of the one-stage game. Show by example that the reverse is not true. Argue that we cannot, however, have more firms active in the one-stage game than in the two-stage game.

12.E.7[B] Consider a one-stage entry model in which firms announce prices and all potential firms have average costs of $AC(q)$ (including their fixed setup costs) with a minimum average

cost of c reached at q. Show that if there exists a J^* such that $J^*\bar{q} = x(\bar{c})$, then any equilibrium of this model produces the perfectly competitive outcome and, hence, the outcome is (first-best) efficient.

12.F.1[B] Show that in the Cournot model discussed in Section 12.F with demand function $\alpha x(p)$, a firm's best-response function $b(Q_{-j})$ is (weakly) decreasing in Q_{-j} provided α is large enough.

12.F.2[B] Suppose each of the I consumers in the economy has quasilinear preferences and a demand function for good ℓ of $x_{\ell i}(p) = a - bp$.

(a) Derive the market inverse demand function.

(b) Now consider a Cournot entry model with this market inverse demand function, technology $c(q) = cq$, and entry cost K. Analyze what happens to the equilibrium prices and output levels, as well as what happens to consumer welfare (measured by consumer surplus), as $I \to \infty$ for both a one-stage and a two-stage entry model.

12.F.3[B] Analyze the two-stage Cournot entry model discussed in Section 12.F when α remains fixed but $K \to 0$. Show, in particular, that the welfare loss goes to zero.

12.F.4[B] Consider the following two-stage entry model with differentiated products and price competition following entry: All potential firms have zero marginal costs and an entry cost of $K > 0$. In stage 2, the demand function for firm j as a function of the price vector $p = (p_1, \ldots, p_J)$ of the J active firms is $x_j(p) = \alpha[\gamma - \beta(Jp_j/\sum_k p_k)]$. Analyze the welfare properties as the size (α) and the substitution (β) parameters change.

12.G.1[B] Consider the linear inverse demand Cournot duopoly model and the linear city differentiated-price duopoly model with differing unit costs that you examined in Exercises 12.C.9 and 12.C.17. Find the derivative, with respect to a change in firm 1's unit cost, of firm 2's equilibrium quantity in the Cournot model and equilibrium price in the linear city model. In which model is this change in firm 2's behavior beneficial to firm 1?

12.AA.1[A] In text.

12.AA.2[C] Prove Proposition 12.AA.4. [*Hint:* Consider a strategy profile of the following form: the players are to play an outcome path involving some pair (q_1, q_2) in period 1 and (q_1^*, q_2^*) in every period thereafter. If either player deviates, this outcome path is restarted.]

12.BB.1[A] In text.

12.BB.2[B] Show that if the incumbent in the entry deterrence model discussed in Appendix B is indifferent between deterring entry and accommodating it, social welfare is strictly greater if he chooses deterrence. Discuss generally why we might not be too surprised if entry deterrence could in some cases raise social welfare.

12.BB.3[C] Consider the linear city model of Exercise 12.C.2 with $v > c + 3t$. Suppose that firm 1 enters the market first and can choose to set up either one plant at one end of the city or two plants, one at each end. Each plant costs F. Then firm E decides whether to enter (for simplicity, restrict it to building one plant) and at which end it wants to locate its plant. Determine the equilibrium of this model. How is it affected by the underlying parameter values? Compare the welfare of this outcome with the welfare if there were no entrant. Compare with the case where there is an entrant but firm 1 is allowed to build only one plant.

CHAPTER

13

Adverse Selection, Signaling, and Screening

13.A Introduction

One of the implicit assumptions of the fundamental welfare theorems is that the characteristics of all commodities are observable to all market participants. Without this condition, distinct markets cannot exist for goods having differing characteristics, and so the complete markets assumption cannot hold. In reality, however, this kind of information is often asymmetrically held by market participants. Consider the following three examples:

(i) When a firm hires a worker, the firm may know less than the worker does about the worker's innate ability.
(ii) When an automobile insurance company insures an individual, the individual may know more than the company about her inherent driving skill and hence about her probability of having an accident.
(iii) In the used-car market, the seller of a car may have much better information about her car's quality than a prospective buyer does.

A number of questions immediately arise about these settings of *asymmetric information*: How do we characterize market equilibria in the presence of asymmetric information? What are the properties of these equilibria? Are there possibilities for welfare-improving market intervention? In this chapter, we study these questions, which have been among the most active areas of research in microeconomic theory during the last twenty years.

We begin, in Section 13.B, by introducing asymmetric information into a simple competitive market model. We see that in the presence of asymmetric information, market equilibria often fail to be Pareto optimal. The tendency for inefficiency in these settings can be strikingly exacerbated by the phenomenon known as *adverse selection*. Adverse selection arises when an informed individual's trading decisions depend on her privately held information in a manner that adversely affects uninformed market participants. In the used-car market, for example, an individual is more likely to decide to sell her car when she knows that it is not very good. When adverse selection is present, uninformed traders will be wary of any informed trader who wishes to trade with them, and their willingness to pay for the product offered

will be low. Moreover, this fact may even further exacerbate the adverse selection problem: If the price that can be received by selling a used car is very low, only sellers with *really* bad cars will offer them for sale. As a result, we may see little trade in markets in which adverse selection is present, even if a great deal of trade would occur were information symmetrically held by all market participants.

We also introduce and study in Section 13.B an important concept for the analysis of market intervention in settings of asymmetric information: the notion of a *constrained Pareto optimal allocation.* These are allocations that cannot be Pareto improved upon by a central authority who, like market participants, cannot observe individuals' privately held information. A Pareto-improving market intervention can be achieved by such an authority only when the equilibrium allocation fails to be a constrained Pareto optimum. In general, the central authority's inability to observe individuals' privately held information leads to a more stringent test for Pareto-improving market intervention.

In Sections 13.C and 13.D, we study how market behavior may adapt in response to these informational asymmetries. In Section 13.C, we consider the possibility that informed individuals may find ways to *signal* information about their unobservable knowledge through observable actions. For example, a seller of a used car could offer to allow a prospective buyer to take the car to a mechanic. Because sellers who have good cars are more likely to be willing to take such an action, this offer can serve as a signal of quality. In Section 13.D, we consider the possibility that uninformed parties may develop mechanisms to distinguish, or *screen*, informed individuals who have differing information. For example, an insurance company may offer two policies: one with no deductible at a high premium and another with a significant deductible at a much lower premium. Potential insureds then *self-select*, with high-ability drivers choosing the policy with a deductible and low-ability drivers choosing the no-deductible policy. In both sections, we consider the welfare characteristics of the resulting market equilibria and the potential for Pareto-improving market intervention.

For expositional purposes, we present all the analysis that follows in terms of the labor market example (i). We should nevertheless emphasize the wide range of settings and fields within economics in which these issues arise. Some of these examples are developed in the exercises at the end of the chapter.

13.B Informational Asymmetries and Adverse Selection

Consider the following simple labor market model adapted from Akerlof's (1970) pioneering work:[1] there are many identical potential firms that can hire workers. Each produces the same output using an identical constant returns to scale technology in which labor is the only input. The firms are risk neutral, seek to maximize their expected profits, and act as price takers. For simplicity, we take the price of the firms' output to equal 1 (in units of a numeraire good).

Workers differ in the number of units of output they produce if hired by a firm,

1. Akerlof (1970) used the example of a used-car market in which only the seller of a used car knows if the car is a "lemon." For this reason, this type of model is sometimes referred to as a *lemons* model.

which we denote by θ.[2] We let $[\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$ denote the set of possible worker productivity levels, where $0 \leq \underline{\theta} < \bar{\theta} < \infty$. The proportion of workers with productivity of θ or less is given by the distribution function $F(\theta)$, and we assume that $F(\cdot)$ is nondegenerate, so that there are at least two types of workers. The total number (or, more precisely, measure) of workers is N.

Workers seek to maximize the amount that they earn from their labor (in units of the numeraire good). A worker can choose to work either at a firm or at home, and we suppose that a worker of type θ can earn $r(\theta)$ on her own through home production. Thus, $r(\theta)$ is the opportunity cost to a worker of type θ of accepting employment; she will accept employment at a firm if and only if she receives a wage of at least $r(\theta)$ (for convenience, we assume that she accepts if she is indifferent).[3]

As a point of comparison, consider first the competitive equilibrium arising in this model when workers' productivity levels are *publicly observable*. Because the labor of each different type of worker is a distinct good, there is a distinct equilibrium wage $w^*(\theta)$ for each type θ. Given the competitive, constant returns nature of the firms, in a competitive equilibrium we have $w^*(\theta) = \theta$ for all θ (recall that the price of their output is 1), and the set of workers accepting employment in a firm is $\{\theta: r(\theta) \leq \theta\}$.[4]

As would be expected from the first fundamental welfare theorem, this competitive outcome is Pareto optimal. To verify this, recall that any Pareto optimal allocation of labor must maximize aggregate surplus (see Section 10.E). Letting $I(\theta)$ be a binary variable that equals 1 if a worker of type θ works for a firm and 0 otherwise, the sum of the aggregate surplus in these labor markets is equal to

$$\int_{\underline{\theta}}^{\bar{\theta}} N[I(\theta)\theta + (1 - I(\theta))r(\theta)]\, dF(\theta). \tag{13.B.1}$$

(This is simply the total revenue generated by the workers' labor.)[5] Aggregate surplus is therefore maximized by setting $I(\theta) = 1$ for those θ with $r(\theta) \leq \theta$ and $I(\theta) = 0$ otherwise (we again resolve indifference in favor of working at a firm). Put simply,

2. A worker's productivity could be random without requiring any change in the analysis that follows; in this case, θ is her *expected* (in a statistical sense) level of productivity.

3. An equivalent model arises from instead specifying $r(\theta)$ as the disutility of labor. In this alternative model, a worker of type θ has quasilinear preferences of the form $u(m, I) = m - r(\theta)I$, where m is the worker's consumption of the numeraire good and $I \in \{0, 1\}$ is a binary variable with $I = 1$ if the worker works and $I = 0$ if not. With these preferences, a worker again accepts employment if and only if she receives a wage of at least $r(\theta)$, and the rest of our analysis remains unaltered.

4. More precisely, there are also competitive equilibria in which $w^*(\theta) = \theta$ for all types of workers who are employed in the equilibrium [those with $r(\theta) \leq \theta$] and $w^*(\theta) \geq \theta$ for those types who are not [those with $r(\theta) > \theta$]. However, for the sake of expositional simplicity, when discussing competitive equilibria that involve no trade in this section we shall restrict attention to equilibrium wages that are equal to workers' (expected) productivity.

5. In Section 10.E, the aggregate surplus from an allocation in a product market (where firms produce output) was written as consumers' direct benefits from consumption of the good less firms' total costs of production. Here, in a labor market setting, a firm's "cost" of employing a worker is the positive revenue it earns, and a worker receives a direct utility (exclusive of any wage payments) of 0 if she works for a firm and $r(\theta)$ if she does not. Hence, aggregate surplus in these markets is equal to firms' total revenues, $\int NI(\theta)\theta\, dF(\theta)$, plus consumers' total revenue from home production, $\int N(1 - I(\theta))r(\theta)\, dF(\theta)$.

since a type θ worker produces at least as much at a firm as at home if and only if $r(\theta) \leq \theta$, in any Pareto optimal allocation the set of workers who are employed by the firms must be $\{\theta: r(\theta) \leq \theta\}$.

We now investigate the nature of competitive equilibrium when workers' productivity levels are *unobservable* by the firms. We begin by developing a notion of competitive equilibrium for this environment with asymmetric information.

To do so, note first that when workers' types are not observable, the wage rate must be independent of a worker's type, and so we will have a single wage rate w for all workers. Consider, then, the supply of labor as a function of the wage rate w. A worker of type θ is willing to work for a firm if and only if $r(\theta) \leq w$. Hence, the set of worker types who are willing to accept employment at wage rate w is

$$\Theta(w) = \{\theta: r(\theta) \leq w\}. \tag{13.B.2}$$

Consider, next, the demand for labor as a function of w. If a firm believes that the average productivity of workers who accept employment is μ, its demand for labor is given by

$$z(w) = \begin{cases} 0 & \text{if } \mu < w \\ [0, \infty] & \text{if } \mu = w \\ \infty & \text{if } \mu > w. \end{cases} \tag{13.B.3}$$

Now, if worker types in set Θ^* are accepting employment offers in a competitive equilibrium, and if firms' beliefs about the productivity of potential employees correctly reflect the actual average productivity of the workers hired in this equilibrium, then we must have $\mu = E[\theta \mid \theta \in \Theta^*]$. Hence, (13.B.3) implies that the demand for labor can equal its supply in an equilibrium with a positive level of employment if and only if $w = E[\theta \mid \theta \in \Theta^*]$. This leads to the notion of a competitive equilibrium presented in Definition 13.B.1.

Definition 13.B.1: In the competitive labor market model with unobservable worker productivity levels, a *competitive equilibrium* is a wage rate w^* and a set Θ^* of worker types who accept employment such that

$$\Theta^* = \{\theta: r(\theta) \leq w^*\} \tag{13.B.4}$$

and

$$w^* = E[\theta \mid \theta \in \Theta^*]. \tag{13.B.5}$$

Condition (13.B.5) involves *rational expectations* on the part of the firms. That is, firms correctly anticipate the average productivity of those workers who accept employment in the equilibrium.

Note, however, that the expectation in (13.B.5) is not well defined when *no* workers are accepting employment in an equilibrium (i.e., when $\Theta^* = \varnothing$). In the discussion that follows, we assume for simplicity that in this circumstance each firm's expectation of potential employees' average productivity is simply the unconditional expectation $E[\theta]$, and we take $w^* = E[\theta]$ in any such equilibrium. (As discussed in footnote 4, we restrict attention to wages that equal workers' expected productivity in any no-trade equilibrium. See Exercise 13.B.5 for the consequences of altering the assumption that expected productivity is $E[\theta]$ when $\Theta^* = \varnothing$.)

Asymmetric Information and Pareto Inefficiency

Typically, a competitive equilibrium as defined in Definition 13.B.1 will fail to be Pareto optimal. To see this point in the simplest-possible setting, consider the case where $r(\theta) = r$ for all θ (every worker is equally productive at home) and suppose that $F(r) \in (0, 1)$, so that there are some workers with $\theta > r$ and some with $\theta < r$. In this setting, the Pareto optimal allocation of labor has workers with $\theta \geq r$ accepting employment at a firm and those with $\theta < r$ not doing so.

Now consider the competitive equilibrium. When $r(\theta) = r$ for all θ, the set of workers who are willing to accept employment at a given wage, $\Theta(w)$, is either $[\underline{\theta}, \bar{\theta}]$ (if $w \geq r$) or $\varnothing$ (if $w < r$). Thus, $E[\theta \mid \theta \in \Theta(w)] = E[\theta]$ for all w and so by (13.B.5) the equilibrium wage rate must be $w^* = E[\theta]$. If $E[\theta] \geq r$, then *all* workers accept employment at a firm; if $E[\theta] < r$, then none do. Which type of equilibrium arises depends on the relative fractions of good and bad workers. For example, if there is a high fraction of low-productivity workers then, because firms cannot distinguish good workers from bad, they will be unwilling to hire any workers at a wage rate that is sufficient to have them accept employment (i.e., a wage of at least r). On the other hand, if there are very few low-productivity workers, then the average productivity of the workforce will be above r, and so the firms will be willing to hire workers at a wage that they are willing to accept. In one case, too many workers are employed relative to the Pareto optimal allocation, and in the other too few.

The cause of this failure of the competitive allocation to be Pareto optimal is simple to see: because firms are unable to distinguish among workers of differing productivities, the market is unable to allocate workers efficiently between firms and home production.[6]

Adverse Selection and Market Unraveling

A particularly striking breakdown in efficiency can arise when $r(\theta)$ varies with θ. In this case, the average productivity of those workers who are willing to accept employment in a firm depends on the wage, and a phenomenon known as *adverse selection* may arise. Adverse selection is said to occur when an informed individual's trading decision depends on her unobservable characteristics in a manner that adversely affects the uninformed agents in the market. In the present context, adverse selection arises when only relatively less capable workers are willing to accept a firm's employment offer at any given wage.

Adverse selection can have a striking effect on market equilibrium. For example, it may seem from our discussion of the case in which $r(\theta) = r$ for all θ that problems arise for the Pareto optimality of competitive equilibrium in the presence of asymmetric information only if there are some workers who should work for a firm and some who should not (since when either $\bar{\theta} < r$ or $\underline{\theta} > r$ the competitive equilibrium outcome is Pareto optimal). In fact, because of adverse selection, this is

6. Another way to understand the difficulty here is that asymmetric information leads to a situation with missing markets and thereby creates externalities (recall Chapter 11). When a worker of type $\theta > E[\theta] = w$ marginally reduces her supply of labor to a firm here, the firm is made worse off, in contrast with the situation in a competitive market with perfect information, where the wage exactly equals a worker's marginal productivity.

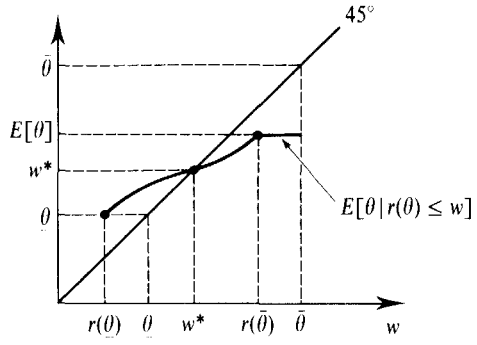


Figure 13.B.1
A competitive equilibrium with adverse selection.

not so; indeed, the market may fail completely despite the fact that *every* worker type should work at a firm.

To see the power of adverse selection, suppose that $r(\theta) \le \theta$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and that $r(\cdot)$ is a strictly increasing function. The first of these assumptions implies that the Pareto optimal labor allocation has every worker type employed by a firm. The second assumption says that workers who are more productive at a firm are also more productive at home. It is this assumption that generates adverse selection: Because the payoff of home production is greater for more capable workers, only less capable workers accept employment at any given wage w [i.e., those with $r(\theta) \le w$].

The expected value of worker productivity in condition (13.B.5) now depends on the wage rate. As the wage rate increases, more productive workers become willing to accept employment at a firm, and the average productivity of those workers accepting employment rises. For simplicity, from this point on, we assume that $F(\cdot)$ has an associated density function $f(\cdot)$, with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. This insures that the average productivity of those workers willing to accept employment, $E[\theta \mid r(\theta) \le w]$, varies continuously with the wage rate on the set $w \in [r(\underline{\theta}), \infty]$.

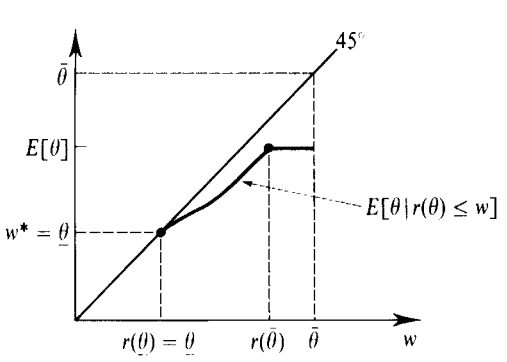
To determine the equilibrium wage, we use conditions (13.B.4) and (13.B.5). Together they imply that the competitive equilibrium wage w^* must satisfy

$$w^* = E[\theta \mid r(\theta) \le w^*]. \tag{13.B.6}$$

We can use Figure 13.B.1 to study the determination of the equilibrium wage w^*. There we graph the values of $E[\theta \mid r(\theta) \le w]$ as a function of w. This function gives the expected value of θ for workers who would choose to work for a firm when the prevailing wage is w. It is increasing in the level w for wages between $r(\underline{\theta})$ and $r(\bar{\theta})$, has a minimum value of $\underline{\theta}$ when $w = r(\underline{\theta})$, and attains a maximum value of $E[\theta]$ for $w \ge r(\bar{\theta})$.[7] The competitive equilibrium wage w^* is found by locating the wage rate at which this function crosses the 45-degree line; at this point, condition (13.B.6) is satisfied. The set of workers accepting employment at a firm is then $\Theta^* = \{\theta: r(\theta) \le w^*\}$. Their average productivity is exactly w^*.[8]

7. The figure does not depict this function for wages below $r(\underline{\theta})$. Because $E[\theta] > r(\underline{\theta})$ in this model, no wage below $r(\underline{\theta})$ can be an equilibrium wage under our assumption that $E[\theta \mid \Theta(w) = \varnothing] = E[\theta]$.

8. For another diagrammatic determination of equilibrium, see Exercise 13.B.1.



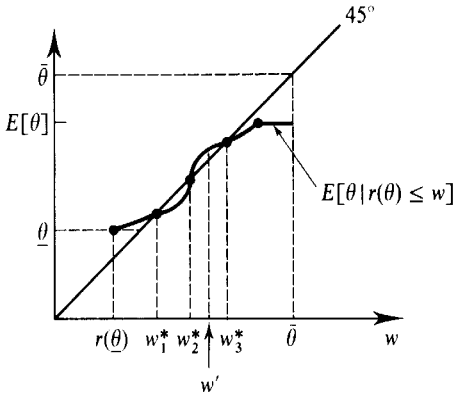


Figure 13.B.2 (left)
Complete market failure.

Figure 13.B.3 (right)
Multiple competitive equilibria.

We can see immediately from Figure 13.B.1 that the market equilibrium need not be efficient. The problem is that to get the best workers to accept employment at a firm, we need the wage to be at least $r(\bar{\theta})$. But in the case depicted, firms cannot break even at this wage because their inability to distinguish among different types of workers leaves them receiving only an expected output of $E[\theta] < r(\bar{\theta})$ from each worker that they hire. The presence of enough low-productivity workers therefore forces the wage down below $r(\bar{\theta})$, which in turn drives the best workers out of the market. But once the best workers are driven out of the market, the average productivity of the workforce falls, thereby further lowering the wage that firms are willing to pay. As a result, once the best workers are driven out of the market, the next-best may follow; the good may then be driven out by the mediocre.

How far can this process go? Potentially *very* far. To see this, consider the case depicted in Figure 13.B.2, where we have $r(\underline{\theta}) = \underline{\theta}$ and $r(\theta) < \theta$ for all other θ. There the equilibrium wage rate is $w^* = \underline{\theta}$, and only type $\underline{\theta}$ workers accept employment in the equilibrium. Because of adverse selection, essentially *no* workers are hired by firms (more precisely, a set of measure zero) even though the social optimum calls for *all* to be hired![9]

Example 13.B.1: To see an explicit example in which the market completely unravels let $r(\theta) = \alpha\theta$, where $\alpha < 1$, and let θ be distributed uniformly on $[0, 2]$. Thus, $r(\underline{\theta}) = \underline{\theta}$ (since $\underline{\theta} = 0$), and $r(\theta) < \theta$ for $\theta > 0$. In this case, $E[\theta \mid r(\theta) \leq w] = (w/2\alpha)$. For $\alpha > \frac{1}{2}$, $E[\theta \mid r(\theta) \leq 0] = 0$ and $E[\theta \mid r(\theta) \leq w] < w$ for all $w > 0$, as in Figure 13.B.2.[10]

The competitive equilibrium defined in Definition 13.B.1 need not be unique. Figure 13.B.3, for example, depicts a case in which there are three equilibria with strictly positive employment levels. Multiple competitive equilibria can arise because there is virtually no restriction on the slope of the function $E[\theta \mid r(\theta) \leq w]$. At any wage w, this slope depends on the density of workers who are just indifferent about accepting employment and so it can vary greatly if this density varies.

9. In this equilibrium, every agent receives the same payoff as if the market were abolished: every firm earns zero and a worker of type θ earns $r(\theta)$ for all θ (including $\theta = \underline{\theta}$).

10. This example is essentially the one developed in Akerlof (1970). His example corresponds to the case $\alpha = \frac{2}{3}$.

Note that the equilibria in Figure 13.B.3 can be *Pareto ranked.* Firms earn zero profits in any equilibrium, and workers are better off if the wage rate is higher (those workers who do not accept employment are indifferent; all other workers are strictly better off). Thus, the equilibrium with the highest wage Pareto dominates all the others. The low-wage, Pareto-dominated equilibria arise because of a *coordination failure*: the wage is too low because firms expect that the productivity of workers accepting employment is poor and, at the same time, only bad workers accept employment precisely because the wage is low.

A Game-Theoretic Approach

The notion of competitive equilibrium that we have employed above is that used by Akerlof (1970). We might ask whether these competitive equilibria can be viewed as the outcome of a richer model in which firms *could* change their offered wages but choose not to in equilibrium.

The situation depicted in Figure 13.B.3 might give you some concern in this regard. For example, consider the equilibrium with wage rate w_2^*. In this equilibrium, a firm that experimented with small changes in its wage offer would find that a small increase in its wage, say to the level $w' > w_2^*$ depicted in the figure, would raise its profits because it would then attract workers with an average productivity of $E[\theta \mid r(\theta) \leq w'] > w'$. Hence, it seems unlikely that a model in which firms could change their offered wages would ever lead to this equilibrium outcome. Similarly, at the equilibrium involving wage w_1^*, a firm that understood the structure of the market would realize that it could earn a strictly positive profit by raising its offered wage to w'.

To be more formal about this idea, consider the following game-theoretic model: The underlying structure of the market [e.g., the distribution of worker productivities $F(\cdot)$ and the reservation wage function $r(\cdot)$] is assumed to be common knowledge. Market behavior is captured in the following two-stage game: In stage 1, two firms simultaneously announce their wage offers (the restriction to two firms is without loss of generality). Then, in stage 2, workers decide whether to work for a firm and, if so, which one. (We suppose that if they are indifferent among some set of firms, then they randomize among them with equal probabilities.)[11]

Proposition 13.B.1 characterizes the subgame perfect Nash equilibria (SPNEs) of this game for the adverse selection model in which $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and $F(\cdot)$ has an associated density $f(\cdot)$ with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.

Proposition 13.B.1: Let W^* denote the set of competitive equilibrium wages for the adverse selection labor market model, and let $w^* = \text{Max}\,\{w: w \in W^*\}$.

(i) If $w^* > r(\underline{\theta})$ and there is an $\varepsilon > 0$ such that $E[\theta \mid r(\theta) \leq w'] > w'$ for all $w' \in (w^* - \varepsilon, w^*)$, then there is a unique pure strategy SPNE of the two-stage game-theoretic model. In this SPNE, employed workers receive

11. Note that if there is a single type of worker with productivity θ, this model is simply the labor market version of the Bertrand model of Section 12.C and has an equilibrium wage equal to θ, the competitive wage.

a wage of w^*, and workers with types in the set $\Theta(w^*) = \{\theta: r(\theta) \leq w^*\}$ accept employment in firms.

(ii) If $w^* = r(\underline{\theta})$, then there are multiple pure strategy SPNEs. However, in every pure strategy SPNE each agent's payoff exactly equals her payoff in the highest-wage competitive equilibrium.

Proof: To begin, note that in any SPNE a worker of type θ must follow the strategy of accepting employment only at one of the highest-wage firms, and of doing so if and only if its wage is at least $r(\theta)$.[12] Using this fact, we can determine the equilibrium behavior of the firms. We do so for each of the two cases in turn.

(i) $w^* > r(\underline{\theta})$: Note, first, that in any SPNE both firms must earn exactly zero. To see this, suppose that there is an SPNE in which a total of M workers are hired at a wage $\bar{w}$ and in which the aggregate profits of the two firms are

$$\Pi = M(E[\theta \mid r(\theta) \leq \bar{w}] - \bar{w}) > 0.$$

Note that $\Pi > 0$ implies that $M > 0$, which in turn implies that $\bar{w} \geq r(\underline{\theta})$. In this case, the (weakly) less-profitable firm, say firm j, must be earning no more than $\Pi/2$. But firm j can earn profits of at least $M(E[\theta \mid r(\theta) \leq \bar{w} + \alpha] - \bar{w} - \alpha)$ by instead offering wage $\bar{w} + \alpha$ for $\alpha > 0$. Since $E[\theta \mid r(\theta) \leq w]$ is continuous in w, these profits can be made arbitrarily close to Π by choosing α small enough. Thus, firm j would be better off deviating, which yields a contradiction: we must therefore have $\Pi \leq 0$. Because neither firm can have strictly negative profits in an SPNE (a firm can always offer a wage of zero), we conclude that both firms must be earning exactly zero in any SPNE.

From this fact, we know that if $\bar{w}$ is the highest wage rate offered by either of the two firms in an SPNE, then either $\bar{w} \in W^*$ (i.e., it must be a competitive equilibrium wage rate) or $\bar{w} < r(\underline{\theta})$ (it must be so low that no workers accept employment). But suppose that $\bar{w} < w^* = \text{Max}\{w: w \in W^*\}$. Then either firm can earn strictly positive expected profits by deviating and offering any wage rate $w' \in (w^* - \varepsilon, w^*)$. We conclude that the highest wage rate offered must equal w^* in any SPNE.

Finally, we argue that both firms naming w^* as their wage, plus the strategies for workers described above, constitute an SPNE. With these strategies, both firms earn zero. Neither firm can earn a positive profit by unilaterally lowering its wage because it gets no workers if it does so. To complete the argument, we show that $E[\theta \mid r(\theta) \leq w] < w$ at every $w > w^*$, so that no unilateral deviation to a higher wage can yield a firm positive profits either. By hypothesis, w^* is the highest competitive wage. Hence, there is no $w > w^*$ at which $E[\theta \mid r(\theta) \leq w] = w$. Therefore, because $E[\theta \mid r(\theta) \leq w]$ is continuous in w, $E[\theta \mid r(\theta) \leq w] - w$ must have the same sign for all $w > w^*$. But we cannot have $E[\theta \mid r(\theta) \leq w] > w$ for all $w > w^*$ because, as $w \to \infty$, $E[\theta \mid r(\theta) \leq w] \to E[\theta]$, which, under our assumptions, is finite. We must therefore have $E[\theta \mid r(\theta) \leq w] < w$ at all $w > w^*$. This completes the argument for case (i).

The assumption that there exists an $\varepsilon > 0$ such that $E[\theta \mid r(\theta) \leq w'] > w'$ for all $w' \in (w^* - \varepsilon, w^*)$ rules out pathological cases such as that depicted in Figure 13.B.4.

(ii) $w^* = r(\underline{\theta})$: In this case, $E[\theta \mid r(\theta) \leq w] < w$ for all $w > w^*$, so that any firm attracting workers at a wage in excess of w^* incurs losses. Moreover, a firm must

12. Recall that we assume that a worker accepts employment whenever she is indifferent.

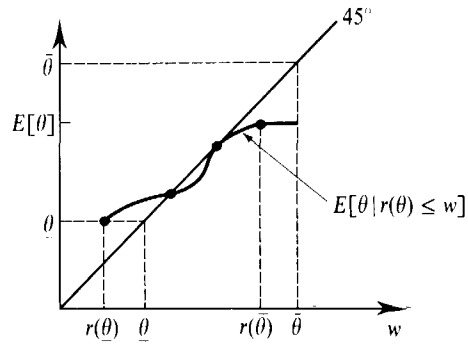


Figure 13.B.4
A pathological example.

earn exactly zero by announcing any $w \leq w^*$. Hence, the set of wage offers (w_1, w_2) that can arise in an SPNE is $\{(w_1, w_2)\colon w_j \leq w^* \text{ for } j = 1, 2\}$. In every one of these SPNEs, all agents earn exactly what they earn at the competitive equilibrium involving wage rate w^*: both firms earn zero, and a worker of type θ earns $r(\theta)$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. ■

One difference between this game-theoretical model and the notion of competitive equilibrium specified in Definition 13.B.1 involves the level of firms' sophistication. In the competitive equilibria of Definition 13.B.1, firms can be fairly unsophisticated. They need know only the average productivity level of the workers who accept employment at the going equilibrium wage; they need not have any idea of the underlying market mechanism. In contrast, in the game-theoretic model, firms understand the entire structure of the market, including the full relationship that exists between the wage rate and the quality of employed workers. The game-theoretic model tells us that if sophisticated firms have the ability to make wage offers, then we break the coordination problem described above. If the wage is too low, some firm will find it in its interest to offer a higher wage and attract better workers; the highest-wage competitive outcome must then arise.[13]

Constrained Pareto Optima and Market Intervention

We have seen that the presence of asymmetric information often results in market equilibria that fail to be Pareto optimal. As a consequence, a central authority who knows all agents' private information (e.g., worker types in the models above), and can engage in lump-sum transfers among agents in the economy, can achieve a Pareto improvement over these outcomes.

In practice, however, a central authority may be no more able to observe agents' private information than are market participants. Without this information, the authority will face additional constraints in trying to achieve a Pareto improvement. For example, arranging lump-sum transfers among workers of different types will be impossible because the authority cannot observe workers' types directly. For Pareto-improving market intervention to be possible in this case, a more stringent test must therefore be passed. An allocation that cannot be Pareto improved by an

13. See Exercise 13.B.6, however, for an example of a model of adverse selection in which, for some parameter values, the highest-wage competitive equilibrium is *not* an SPNE of our game-theoretic model.

authority who is unable to observe agents' private information is known as a *constrained* (or *second-best*) *Pareto optimum.* Because it is more difficult to generate a Pareto improvement in the absence of an ability to observe agents' types, a constrained Pareto optimal allocation need not be (fully) Pareto optimal [however, a (full) Pareto optimum is necessarily a constrained Pareto optimum].

Here, as an example, we shall study whether Pareto-improving market intervention is possible in the context of our adverse selection model (where $r(\cdot)$ is strictly increasing with $r(\theta) \leq 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and $F(\cdot)$ has an associated density $f(\cdot)$ with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$) when the central authority cannot observe worker types. That is, we study whether the competitive equilibria of this adverse selection model are constrained Pareto optima.

In general, the formal analysis of this problem uses tools that we develop in Section 14.C in our study of principal-agent models with hidden information (see, in particular, the discussion of monopolistic screening). As these techniques have yet to be introduced, we shall not analyze this problem fully here. (Once you have studied Section 14.C, however, refer back to the discussion in small type at the end of this section.) Nevertheless, we can convey much of the analysis here.

By way of motivation, note first that in examining whether a Pareto improvement relative to a market equilibrium is possible, we might as well simply think of intervention schemes in which the authority runs the firms herself and tries to achieve a Pareto improvement for the workers (the firms' owners will then earn exactly what they were earning in the equilibrium, namely zero profits). Second, because the authority cannot distinguish directly among different types of workers, any differences in lump-sum transfers to or from a worker can depend only on whether the worker is employed (the workers otherwise appear identical). Thus, intuitively, there should be no loss of generality in restricting attention to interventions in which the authority runs the firms herself, offers a wage of w_e to those accepting employment, an unemployment benefit of w_u to those who do not [these workers also receive $r(\theta)$], leaves the workers free to choose whether to accept employment in a firm, and balances her budget. (In the small-type discussion at the end of this section, we show formally that this is the case.)

Given this background, can the competitive equilibria of our adverse selection model be Pareto-improved upon in this way? Consider, first, dominated competitive equilibria, that is, competitive equilibria that are Pareto dominated by some other competitive equilibrium (e.g., the equilibrium with wage rate w_1^* shown in Figure 13.B.3). A central authority who is unable to observe worker types can always implement the best (highest-wage) competitive equilibrium outcome. She need only set $w_e = w^*$, the highest competitive equilibrium wage, and $w_u = 0$. All workers in set $\Theta(w^*)$ then accept employment in a firm and, since $w^* = E[\theta \mid r(\theta) \leq w^*]$, the authority exactly balances her budget.[14] Thus, the outcome in such an equilibrium is *not* a constrained Pareto optimum. In this case, the planner is essentially able to step in and solve the coordination failure that is keeping the market at the low-wage equilibrium.

14. An equivalent but less heavy-handed intervention would have the authority simply require any operating firm to pay a wage rate equal to w^*. Firms will be willing to remain operational because they break even at this wage rate, and a Pareto improvement results.

What about the highest-wage competitive equilibrium (i.e., the SPNE outcome in the game-theoretic model of Proposition 13.B.1)? As Proposition 13.B.2 shows, any such equilibrium *is* a constrained Pareto optimum in this model.

Proposition 13.B.2: In the adverse selection labor market model (where $r(\cdot)$ is strictly increasing with $r(\theta) \leq \theta$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and $F(\cdot)$ has an associated density $f(\cdot)$ with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$), the highest-wage competitive equilibrium is a constrained Pareto optimum.

Proof: If all workers are employed in the highest wage competitive equilibrium then the outcome is fully (and, hence, constrained) Pareto optimal. So suppose some are not employed. Note, first, that for any wage w_e and unemployment benefit w_u offered by the central authority the set of worker types accepting employment has the form $[\underline{\theta}, \hat{\theta}]$ for some $\hat{\theta}$ [it is $\{\theta: w_u + r(\theta) \leq w_e\}$]. Suppose, then, that the authority attempts to implement an outcome in which worker types $\theta \leq \hat{\theta}$ for $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ accept employment. To do so, she must choose w_e and w_u so that

$$w_u + r(\hat{\theta}) = w_e. \tag{13.B.7}$$

In addition, to balance her budget, w_u and w_e must also satisfy[15]

$$w_e F(\hat{\theta}) + w_u(1 - F(\hat{\theta})) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta)\, d\theta. \tag{13.B.8}$$

Substituting into (13.B.7) from (13.B.8), we find that, given the choice of $\hat{\theta}$, the values of w_u and w_e must be

$$w_u(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta)\, d\theta - r(\hat{\theta})F(\hat{\theta}) \tag{13.B.9}$$

and

$$w_e(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta)\, d\theta + r(\hat{\theta})(1 - F(\hat{\theta})), \tag{13.B.10}$$

or, equivalently,

$$w_u(\hat{\theta}) = F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta})) \tag{13.B.11}$$

$$w_e(\hat{\theta}) = F(\hat{\theta})(E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta})) + r(\hat{\theta}). \tag{13.B.12}$$

Now, let θ^* denote the highest worker type who accepts employment in the highest-wage competitive equilibrium. We know that $r(\theta^*) = E[\theta \mid \theta \leq \theta^*]$. Hence, from conditions (13.B.11) and (13.B.12), we see that $w_u(\theta^*) = 0$ and $w_e(\theta^*) = r(\theta^*)$. Thus, the outcome when the authority sets $\hat{\theta} = \theta^*$ is exactly the same as in the highest-wage competitive equilibrium.

We now examine whether a Pareto improvement can be achieved by setting $\hat{\theta} \neq \theta^*$. Note that for any $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$ with $\hat{\theta} \neq \theta^*$, type $\underline{\theta}$ workers are worse off than in the equilibrium if $w_e(\hat{\theta}) < r(\theta^*)$ [$r(\theta^*)$ is their wage in the equilibrium] and type $\bar{\theta}$ workers are worse off if $w_u(\hat{\theta}) < 0$.

Consider $\hat{\theta} < \theta^*$ first. Since $r(\theta^*) > r(\hat{\theta})$, condition (13.B.10) implies that

$$w_e(\hat{\theta}) \leq \int_{\underline{\theta}}^{\hat{\theta}} \theta f(\theta)\, d\theta + r(\theta^*)(1 - F(\hat{\theta})),$$

15. The authority will never wish to run a budget surplus. If w_u and w_e lead to a budget surplus, then setting $\hat{w}_u = w_u + \varepsilon$ and $\hat{w}_e = w_e + \varepsilon$ for some $\varepsilon > 0$ is budget feasible and is Pareto superior. (Note that the set of workers accepting employment would be unchanged.)

and so

$$\begin{aligned} w_e(\hat{\theta}) - r(\theta^*) &\leq F(\hat{\theta})\big(E[\theta \mid \theta \leq \hat{\theta}] - r(\theta^*)\big) \\ &= F(\hat{\theta})\big(E[\theta \mid \theta \leq \hat{\theta}] - E[\theta \mid \theta \leq \theta^*]\big) \\ &< 0. \end{aligned}$$

Thus, type $\underline{\theta}$ workers must be made worse off by any such intervention.

Now consider $\hat{\theta} > \theta^*$. We know that $E[\theta \mid r(\theta) \leq w] < w$ for all $w > w^*$ (see the proof of Proposition 13.B.1). Thus, since $r(\theta^*) = w^*$ and $r(\cdot)$ is strictly increasing, we have $E[\theta \mid r(\theta) \leq r(\hat{\theta})] < r(\hat{\theta})$ for all $\hat{\theta} > \theta^*$. Moreover,

$$E[\theta \mid r(\theta) \leq r(\hat{\theta})] = E[\theta \mid \theta \leq \hat{\theta}],$$

and so $E[\theta \mid \theta \leq \hat{\theta}] - r(\hat{\theta}) < 0$ for all $\hat{\theta} > \theta^*$. But condition (13.B.11) then implies that $w_u(\hat{\theta}) < 0$ for all $\hat{\theta} > \theta^*$, and so type $\bar{\theta}$ workers are made worse off by any such intervention. ■

Hence, when a central authority cannot observe worker types, her options may be severely limited. Indeed, in the adverse selection model just considered, the authority is unable to create a Pareto improvement as long as the highest-wage competitive equilibrium (the SPNE outcome of the game-theoretic model of Proposition 13.B.1) is the market outcome.[16] More generally, whether Pareto-improving market intervention is possible in situations of asymmetric information depends on the specifics of the market under study (and as we have already seen, possibly on which equilibria result). Exercises 13.B.8 and 13.B.9 provide two examples of models in which the highest-wage competitive equilibrium may fail to be a constrained Pareto optimum.

Although it is impossible to Pareto improve a constrained Pareto optimal allocation, market intervention could still be justified in the pursuit of distributional aims. For example, if social welfare is given by the sum of weighted worker utilities

$$\int_{\underline{\theta}}^{\bar{\theta}} [I(\theta)\theta + (1 - I(\theta))r(\theta)]\lambda(\theta)\, dF(\theta), \tag{13.B.13}$$

where $\lambda(\theta) > 0$ for all θ, then social welfare may be increased even though some worker types end up worse off. In the applied literature, for example, it is common to see aggregate surplus used as the social welfare function, which is equivalent to the choice of $\lambda(\theta) = N$ for all θ.[17] When society has this social welfare function, social welfare can be raised relative to the competitive equilibrium in Figure 13.B.1 (which, by Proposition 13.B.2, is a constrained Pareto optimum) simply by mandating that all workers must work for a firm and that all firms must

16. Proposition 13.B.2 can also be readily generalized to allow $r(\theta) > \theta$ for some θ. (See Exercise 13.B.10.)

17. Note that when types cannot be observed, aggregate surplus is no longer a valid welfare measure for *any* social welfare function because, unlike the case of perfect information, lump-sum transfers across worker types are infeasible. (See Section 10.E for a discussion of the need for lump-sum transfers to justify aggregate surplus as a welfare measure for any social welfare function.)

pay workers a wage of $E(\theta)$. Although workers of type $\bar{\theta}$ are made worse off by this intervention, welfare as measured by aggregate surplus increases.[18]

An interesting interpretation of the choice of aggregate surplus as a social welfare function is in terms of an unborn worker's ex ante expected utility. In particular, imagine that each worker originally has a probability $f(\theta)$ of ending up a type θ worker. If this unborn worker is risk neutral, then her ex ante expected utility is exactly equal to expression (13.B.13) with $\lambda(\theta) = 1$ for all θ. Thus, maximization of aggregate surplus is equivalent to maximization of this unborn worker's expected utility. We might then say that an allocation is an *ex ante constrained Pareto optimum* in this model if, in the absence of an ability to observe worker types, it is impossible to devise a market intervention that raises aggregate surplus. We see, therefore, that whether an allocation is a constrained optimum (and, thus, whether a planned intervention leads to a Pareto improvement) can depend on the point at which the welfare evaluation is conducted (i.e., before the workers know their types, or after).[19]

Let us now use the techniques of Section 14.C to show formally that we can restrict attention in searching for a Pareto improvement to interventions of the type considered above. We shall look for a Pareto improvement for the workers keeping the profits of the firms' owners nonnegative. For notational simplicity, we shall treat the firms as a single aggregate firm.

By the revelation principle (see Section 14.C), we know that we can restrict attention to direct revelation mechanisms in which every worker type tells the truth. Here a direct revelation mechanism assigns, for each worker type $\theta \in [\underline{\theta}, \bar{\theta}]$, a payment from the authority to the worker of $w(\theta) \in \mathbb{R}$, a tax $t(\theta)$ paid by the firm to the authority, and an employment decision $I(\theta) \in \{0, 1\}$. The set of feasible mechanisms here are those that satisfy the *individual rationality constraint* for the firm,

$$\int_{\underline{\theta}}^{\bar{\theta}} [I(\theta)\theta - t(\theta)]\, dF(\theta) \geq 0, \tag{13.B.14}$$

the *budget balance condition* for the central authority,

$$\int_{\underline{\theta}}^{\bar{\theta}} [t(\theta) - w(\theta)]\, dF(\theta) \geq 0, \tag{13.B.15}$$

and the *truth-telling* (or *incentive compatibility*, or *self-selection*) *constraints* that say that for all θ and $\hat{\theta}$

$$w(\theta) + (1 - I(\theta))r(\theta) \geq w(\hat{\theta}) + (1 - I(\hat{\theta}))r(\theta). \tag{13.B.16}$$

Note, first, that mechanism $[w(\cdot), t(\cdot), I(\cdot)]$ is feasible only if $[w(\cdot), I(\cdot)]$ satisfies both condition (13.B.16) and

$$\int_{\underline{\theta}}^{\bar{\theta}} [I(\theta)\theta - w(\theta)]\, dF(\theta) \geq 0. \tag{13.B.17}$$

18. Moreover, because lump-sum transfers among different types of workers are not possible in the absence of an ability to observe worker types, the achievement of these distributional aims actually *requires* direct intervention in the labor market, in contrast with the case of perfect information.

19. Holmstrom and Myerson (1983) call this ex ante notion of constrained Pareto optimality *ex ante incentive efficiency*. Their terminology refers to the fact that we are taking an ex ante perspective in evaluating welfare (before the realization of worker types) and that a central authority who cannot observe worker types faces *incentive constraints* if she wants to induce workers to reveal their types. Holmstrom and Myerson call our previous notion of constrained Pareto efficiency *interim incentive efficiency* because the perspective used to assess Pareto optimality is that of workers who already know their types. See Section 23.F for a more general discussion of these concepts.

Moreover, if $[w(\cdot), I(\cdot)]$ satisfies (13.B.16) and (13.B.17), then there exists a $t(\cdot)$ such that $[w(\cdot), t(\cdot), I(\cdot)]$ satisfies (13.B.14) (13.B.16). Condition (13.B.17), however, is exactly the budget constraint faced by a central authority who runs the firms herself. Hence, we can restrict attention to schemes in which the authority runs the firms herself and uses a direct revelation mechanism $[w(\cdot), I(\cdot)]$ satisfying (13.B.16) and (13.B.17).

Now consider any two types θ' and θ'' for which $I(\theta') = I(\theta'')$. Setting $\theta = \theta'$ and $\hat{\theta} = \theta''$ in condition (13.B.16), we see that we must have $w(\theta') \geq w(\theta'')$. Likewise, letting $\theta = \theta''$ and $\hat{\theta} = \theta'$, we must have $w(\theta'') \geq w(\theta')$. Together, this implies that $w(\theta') = w(\theta'')$. Since $I(\theta) \in \{0, 1\}$, we see that any feasible mechanism $[w(\cdot), I(\cdot)]$ can be viewed as a scheme that gives each worker a choice between two outcomes, $(w_e, I = 1)$ and $(w_u, I = 0)$ and satisfies the budget balance condition (13.B.17). This is exactly the class of mechanisms studied above.

13.C Signaling

Given the problems observed in Section 13.B, one might expect mechanisms to develop in the marketplace to help firms distinguish among workers. This seems plausible because both the firms and the high-ability workers have incentives to try to accomplish this objective. The mechanism that we examine in this section is that of *signaling*, which was first investigated by Spence (1973, 1974). The basic idea is that high-ability workers may have actions they can take to distinguish themselves from their low-ability counterparts.

The simplest example of such a signal occurs when workers can submit to some costless test that reliably reveals their type. It is relatively straightforward to show that in any subgame perfect Nash equilibrium all workers with ability greater than $\underline{\theta}$ will submit to the test and the market will achieve the full information outcome (see Exercise 13.C.1). Any worker who chooses not to take the test will be correctly treated as being no better than the worst type of worker.

However, in many instances, no procedure exists that directly reveals a worker's type. Nevertheless, as the analysis in this section reveals, the potential for signaling may still exist.

Consider the following adaptation of the model discussed in Section 13.B. For simplicity, we restrict attention to the case of two types of workers with productivities θ_H and θ_L, where $\theta_H > \theta_L > 0$ and $\lambda = \text{Prob}\,(\theta = \theta_H) \in (0, 1)$. The important extension of our previous model is that before entering the job market a worker can get some education, and the amount of education that a worker receives is observable. To make matters particularly stark, we assume that education does *nothing* for a worker's productivity (see Exercise 13.C.2 for the case of productive signaling). The cost of obtaining education level e for a type θ worker (the cost may be of either monetary or psychic origin) is given by the twice continuously differentiable function $c(e, \theta)$, with $c(0, \theta) = 0$, $c_e(e, \theta) > 0$, $c_{ee}(e, \theta) > 0$, $c_\theta(e, \theta) < 0$ for all $e > 0$, and $c_{e\theta}(e, \theta) < 0$ (subscripts denote partial derivatives). Thus, both the cost and the marginal cost of education are assumed to be lower for high-ability workers; for example, the work required to obtain a degree might be easier for a high-ability individual. Letting $u(w, e \mid \theta)$ denote the utility of a type θ worker who chooses education level e and receives wage w, we take $u(w, e \mid \theta)$ to equal her wage less any educational costs incurred: $u(w, e \mid \theta) = w - c(e, \theta)$. As in Section 13.B, a worker of type θ can earn $r(\theta)$ by working at home.

In the analysis that follows, we shall see that this otherwise useless education may serve as a signal of unobservable worker productivity. In particular, equilibria emerge in which high-productivity workers choose to get more education than low-productivity workers and firms correctly take differences in education levels as a signal of ability. The welfare effects of signaling activities are generally ambiguous. By revealing information about worker types, signaling can lead to a more efficient allocation of workers' labor, and in some instances to a Pareto improvement. At the same time, because signaling activity is costly, workers' welfare may be reduced if they are compelled to engage in a high level of signaling activity to distinguish themselves.

To keep things simple, throughout most of this section we concentrate on the special case in which $r(\theta_H) = r(\theta_L) = 0$. Note that under this assumption the unique equilibrium that arises in the absence of the ability to signal (analyzed in Section 13.B) has all workers employed by firms at a wage of $w^* = E[\theta]$ and is Pareto efficient. Hence, our study of this case emphasizes the potential inefficiencies created by signaling. After studying this case in detail, we briefly illustrate (in small type) how, with alternative assumptions about the function $r(\cdot)$, signaling may instead generate a Pareto improvement.

A portion of the game tree for this model is shown in Figure 13.C.1. Initially, a random move of nature determines whether a worker is of high or low ability. Then, conditional on her type, the worker chooses how much education to obtain. After obtaining her chosen education level, the worker enters the job market. Conditional on the observed education level of the worker, two firms simultaneously make wage offers to her. Finally, the worker decides whether to work for a firm and, if so, which one.

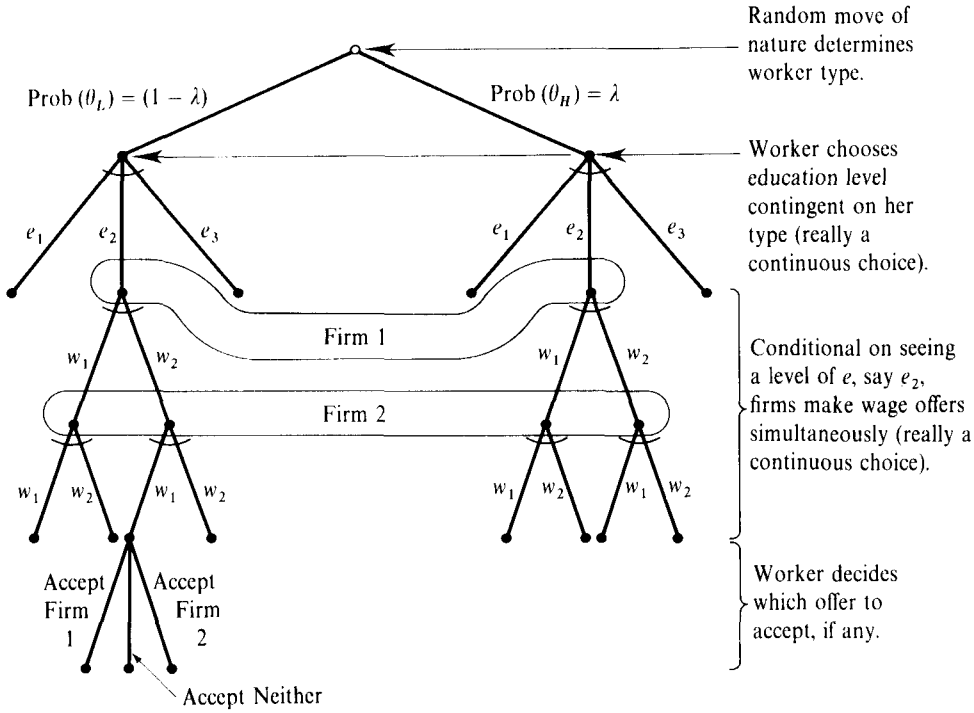


Figure 13.C.1
The extensive form of the education signaling game.

Note that, in contrast with the model of Section 13.B, here we explicitly model only a single worker of unknown type; the model with many workers can be thought of as simply having many of these single-worker games going on simultaneously, with the fraction of high-ability workers in the market being λ. In discussing the equilibria of this game, we often speak of the "high-ability workers" and "low-ability workers," having the many-workers case in mind.

The equilibrium concept we employ is that of a weak perfect Bayesian equilibrium (see Definition 9.C.3), but with an added condition. Put formally, we require that, in the game tree depicted in Figure 13.C.1, the firms' beliefs have the property that, for each possible choice of e, there exists a number $\mu(e) \in [0, 1]$ such that: (i) firm 1's belief that the worker is of type θ_H after seeing her choose e is $\mu(e)$ and (ii) after the worker has chosen e, firm 2's belief that the worker is of type θ_H and that firm 1 has chosen wage offer w is precisely $\mu(e)\sigma_1^*(w \mid e)$, where $\sigma_1^*(w \mid e)$ is firm 1's equilibrium probability of choosing wage offer w after observing education level e. This extra condition adds an element of commonality to the firms' beliefs about the type of worker who has chosen e, and requires that the firms' beliefs about each others' wage offers following e are consistent with the equilibrium strategies both on and off the equilibrium path.

We refer to a weak perfect Bayesian equilibrium satisfying this extra condition on beliefs as a *perfect Bayesian equilibrium* (PBE). Fortunately, this PBE notion can more easily, and equivalently, be stated as follows: A set of strategies and a belief function $\mu(e) \in [0, 1]$ giving the firms' common probability assessment that the worker is of high ability after observing education level e is a PBE if

(i) The worker's strategy is optimal given the firm's strategies.
(ii) The belief function $\mu(e)$ is derived from the worker's strategy using Bayes' rule where possible.
(iii) The firms' wage offers following each choice e constitute a Nash equilibrium of the simultaneous-move wage offer game in which the probability that the worker is of high ability is $\mu(e)$.[20]

In the context of the model studied here, this notion of a PBE is equivalent to the sequential equilibrium concept discussed in Section 9.C. We also restrict our attention throughout to pure strategy equilibria.

We begin our analysis at the end of the game. Suppose that after seeing some education level e, the firms attach a probability of $\mu(e)$ that the worker is type θ_H. If so, the expected productivity of the worker is $\mu(e)\theta_H + (1 - \mu(e))\theta_L$. In a simultaneous-move wage offer game, the firms' (pure strategy) Nash equilibrium wage offers equal the worker's expected productivity (this game is very much like the Bertrand pricing game discussed in Section 12.C). Thus, in any (pure strategy) PBE, we must have both firms offering a wage exactly equal to the worker's expected productivity, $\mu(e)\theta_H + (1 - \mu(e))\theta_L$.

20. Thus, the extra condition we add imposes equilibrium-like play in parts of the tree off the equilibrium path. See Section 9.C for a discussion of the need to augment the weak perfect Bayesian equilibrium concept to achieve this end.

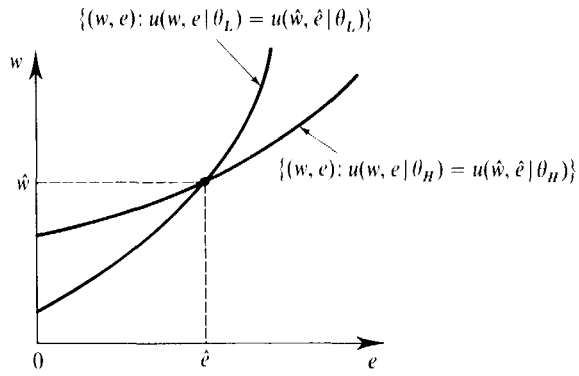


Figure 13.C.2 (left)
Indifference curves for high- and low-ability workers: the single-crossing property.

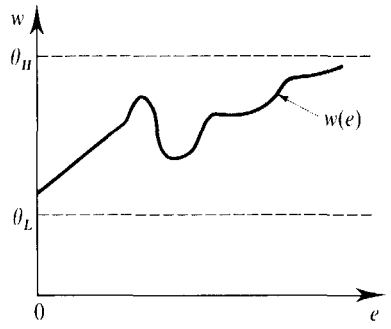


Figure 13.C.3 (right)
A wage schedule.

Knowing this fact, we turn to the issue of the worker's equilibrium strategy, her choice of an education level contingent on her type. As a first step in this analysis, it is useful to examine the worker's preferences over (wage rate, education level) pairs. Figure 13.C.2 depicts an indifference curve for each of the two types of workers (with wages measured on the vertical axis and education levels measured on the horizontal axis). Note that these indifference curves cross only once and that, where they do, the indifference curve of the high-ability worker has a smaller slope. This property of preferences, known as the *single-crossing property*, plays an important role in the analysis of signaling models and in models of asymmetric information more generally. It arises here because the worker's marginal rate of substitution between wages and education at any given (w, e) pair is $(dw/de)_{\bar{u}} = c_e(e, \theta)$, which is decreasing in θ because $c_{e\theta}(e, \theta) < 0$.

We can also graph a function giving the equilibrium wage offer that results for each education level, which we denote by $w(e)$. Note that since in any PBE $w(e) = \mu(e)\theta_H + (1 - \mu(e))\theta_L$ for the equilibrium belief function $\mu(e)$, the equilibrium wage offer resulting from any choice of e must lie in the interval $[\theta_L, \theta_H]$. A possible wage offer function $w(e)$ is shown in Figure 13.C.3.

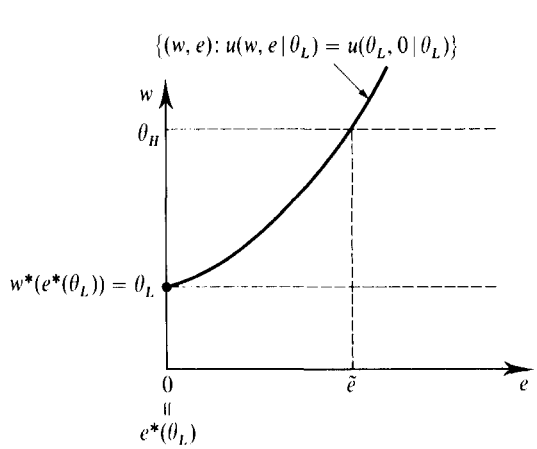
We are now ready to determine the equilibrium education choices for the two types of workers. It is useful to consider separately two different types of equilibria that might arise: *separating equilibria*, in which the two types of workers choose different education levels, and *pooling equilibria*, in which the two types choose the same education level.

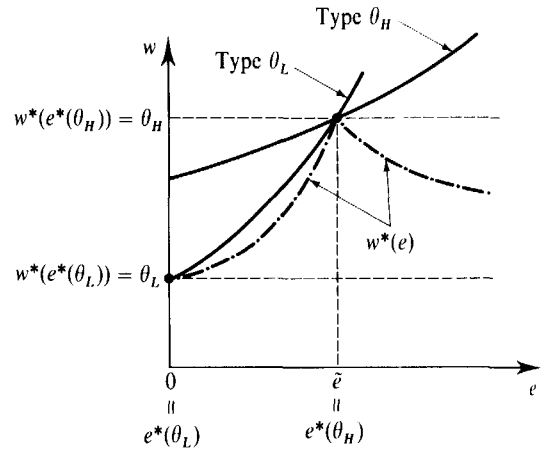
Separating Equilibria

To analyze separating equilibria, let $e^*(\theta)$ be the worker's equilibrium education choice as a function of her type, and let $w^*(e)$ be the firms' equilibrium wage offer as a function of the worker's education level. We first establish two useful lemmas.

Lemma 13.C.1: In any separating perfect Bayesian equilibrium, $w^*(e^*(\theta_H)) = \theta_H$ and $w^*(e^*(\theta_L)) = \theta_L$; that is, each worker type receives a wage equal to her productivity level.

Proof: In any PBE, beliefs on the equilibrium path must be correctly derived from the equilibrium strategies using Bayes' rule. Here this implies that upon seeing education level $e^*(\theta_L)$, firms must assign probability one to the worker being type θ_L. Likewise, upon seeing education level $e^*(\theta_H)$, firms must assign probability one





to the worker being type θ_H. The resulting wages are then exactly θ_L and θ_H, respectively. ■

Figure 13.C.4 (left) Low-ability worker's outcome in a separating equilibrium.

Lemma 13.C.2: In any separating perfect Bayesian equilibrium, $e^*(\theta_L) = 0$; that is, a low-ability worker chooses to get no education.

Proof: Suppose not, that is, that when the worker is type θ_L, she chooses some strictly positive education level $\hat{e} > 0$. According to Lemma 13.C.1, by doing so, the worker receives a wage equal to θ_L. However, she would receive a wage of at least θ_L if she instead chose $e = 0$. Since choosing $e = 0$ would have save her the cost of education, she would be strictly better off by doing so, which is a contradiction to the assumption that $\hat{e} > 0$ is her equilibrium education level. ■

Figure 13.C.5 (right) A separating equilibrium: Type is inferred from education level.

Lemma 13.C.2 implies that, in any separating equilibrium, type θ_L's indifference curve through her equilibrium level of education and wage must look as depicted in Figure 13.C.4.

Using Figure 13.C.4, we can construct a separating equilibrium as follows: Let $e^*(\theta_H) = \tilde{e}$, let $e^*(\theta_L) = 0$, and let the schedule $w^*(e)$ be as drawn in Figure 13.C.5. The firms' equilibrium beliefs following education choice e are $\mu^*(e) = (w^*(e) - \theta_L)/(\theta_H - \theta_L)$. Note that they satisfy $\mu^*(e) \in [0, 1]$ for all $e \geq 0$, since $w^*(e) \in [\theta_L, \theta_H]$.

To verify that this is indeed a PBE, note that we are completely free to let firms have any beliefs when e is neither 0 nor $\tilde{e}$. On the other hand, we must have $\mu(0) = 0$ and $\mu(\tilde{e}) = 1$. The wage offers drawn, which have $w^*(0) = \theta_L$ and $w^*(\tilde{e}) = \theta_H$, reflect exactly these beliefs.

What about the worker's strategy? It is not hard to see that, given the wage function $w^*(e)$, the worker is maximizing her utility by choosing $e = 0$ when she is type θ_L and by choosing $e = \tilde{e}$ when she is type θ_H. This can be seen in Figure 13.C.5 by noting that, for each type that she may be, the worker's indifference curve is at its highest-possible level along the schedule $w^*(e)$. Thus, strategies $[e^*(\theta), w^*(e)]$ and the associated beliefs $\mu(e)$ of the firms do in fact constitute a PBE.

Note that this is not the only PBE involving these education choices by the two types of workers. Because we have so much freedom to choose the firms' beliefs off the equilibrium path, many wage schedules can arise that support these education

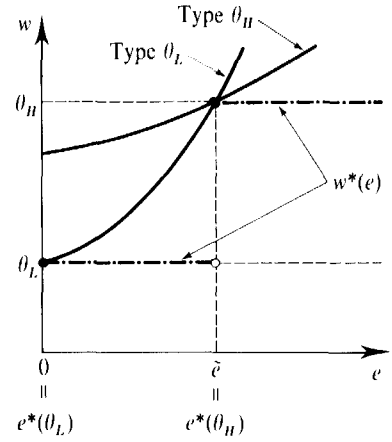


Figure 13.C.6 (left)
A separating equilibrium with the same education choices as in Figure 13.C.5 but different off-equilibrium-path beliefs.

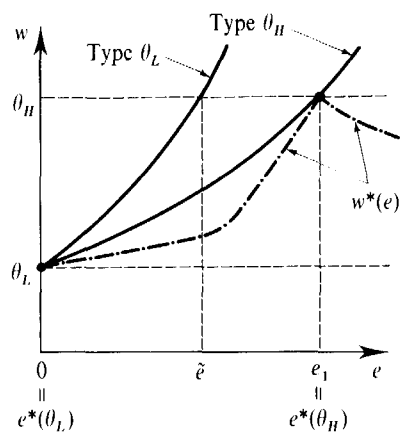


Figure 13.C.7 (right)
A separating equilibrium with an education choice $e^*(\theta_H) > \tilde{e}$ by high-ability workers.

choices. Figure 13.C.6 depicts another one; in this PBE, firms believe that the worker is certain to be of high quality if $e \geq \tilde{e}$ and is certain to be of low quality if $e < \tilde{e}$. The resulting wage schedule has $w^*(e) = \theta_H$ if $e \geq \tilde{e}$ and $w^*(e) = \theta_L$ if $e < \tilde{e}$.

In these separating equilibria, high-ability workers are willing to get otherwise useless education simply because it allows them to distinguish themselves from low-ability workers and receive higher wages. The fundamental reason that education can serve as a signal here is that the marginal cost of education depends on a worker's type. Because the marginal cost of education is higher for a low-ability worker [since $c_{e\theta}(e, \theta) < 0$], a type θ_H worker may find it worthwhile to get some positive level of education $e' > 0$ to raise her wage by some amount $\Delta w > 0$, whereas a type θ_L worker may be unwilling to get this same level of education in return for the same wage increase. As a result, firms can reasonably come to regard education level as a signal of worker quality.

The education level for the high-ability type observed above is not the only one that can arise in a separating equilibrium in this model. Indeed, many education levels for the high-ability type are possible. In particular, any education level between $\tilde{e}$ and e_1 in Figure 13.C.7 can be the equilibrium education level of the high-ability workers. A wage schedule that supports education level $e^*(\theta_H) = e_1$ is depicted in the figure. Note that the education level of the high-ability worker cannot be below $\tilde{e}$ in a separating equilibrium because, if it were, the low-ability worker would deviate and pretend to be of high ability by choosing the high-ability education level. On the other hand, the education level of the high-ability worker cannot be above e_1 because, if it were, the high-ability worker would prefer to get no education, even if this resulted in her being thought to be of low ability.

Note that these various separating equilibria can be Pareto ranked. In all of them, firms earn zero profits, and a low-ability worker's utility is θ_L. However, a high-ability worker does strictly better in equilibria in which she gets a lower level of education. Thus, separating equilibria in which the high-ability worker gets education level $\tilde{e}$ (e.g., the equilibria depicted in Figures 13.C.5 and 13.C.6) Pareto dominate all the others. The Pareto-dominated equilibria are sustained because of the high-ability worker's fear that if she chooses a lower level of education than that prescribed in the equilibrium firms will believe that she is not a high-ability worker. These beliefs can be maintained because in equilibrium they are never disconfirmed.

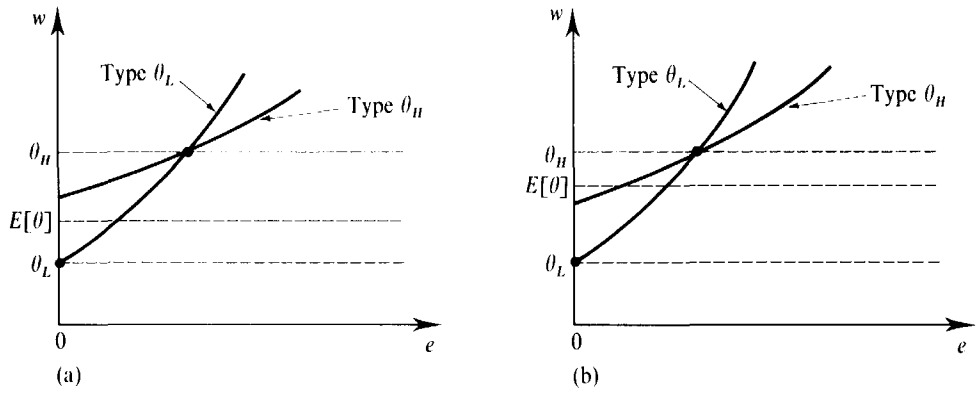


Figure 13.C.8
Separating equilibria may be Pareto dominated by the no-signaling outcome. (a) A separating equilibrium that is not Pareto dominated by the no-signaling outcome. (b) A separating equilibrium that is Pareto dominated by the no-signaling outcome.

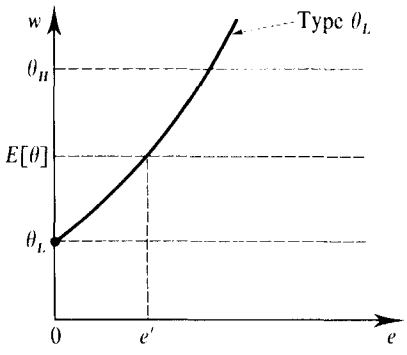
It is of interest to compare welfare in these equilibria with that arising when worker types are unobservable but no opportunity for signaling is available. When education is not available as a signal (so workers also incur no education costs), we are back in the situation studied in Section 13.B. In both cases, firms earn expected profits of zero. However, low-ability workers are strictly worse off when signaling is possible. In both cases they incur no education costs, but when signaling is possible they receive a wage of θ_L rather than $E(\theta)$.

What about high-ability workers? The somewhat surprising answer is that high-ability workers may be either better or worse off when signaling is possible. In Figure 13.C.8(a), the high-ability workers are better off because of the increase in their wages arising through signaling. However, in Figure 13.C.8(b), even though high-ability workers seek to take advantage of the signaling mechanism to distinguish themselves, they are *worse* off than when signaling is impossible! Although this may seem paradoxical (if high-ability workers choose to signal, how can they be worse off?), its cause lies in the fact that in a separating signaling equilibrium firms' expectations are such that the wage–education outcome from the no-signaling situation, $(w, e) = (E[\theta], 0)$, is no longer available to the high-ability workers; if they get no education in the separating signaling equilibrium, they are thought to be of low ability and offered a wage of θ_L. Thus, they can be worse off when signaling is possible, even though they are choosing to signal.

Note that because the set of separating equilibria is completely unaffected by the fraction λ of high-ability workers, as this fraction grows it becomes more likely that the high-ability workers are made worse off by the possibility of signaling [compare Figures 13.C.8(a) and 13.C.8(b)]. In fact, as this fraction gets close to 1, nearly every worker is getting costly education just to avoid being thought to be one of the handful of bad workers!

Pooling Equilibria

Consider now pooling equilibria, in which the two types of workers choose the same level of education, $e^*(\theta_L) = e^*(\theta_H) = e^*$. Since the firms' beliefs must be correctly derived from the equilibrium strategies and Bayes' rule when possible, their beliefs when they see education level e^* must assign probability λ to the worker being type θ_H. Thus, in any pooling equilibrium, we must have $w^*(e^*) = \lambda\theta_H + (1 - \lambda)\theta_L = E[\theta]$.



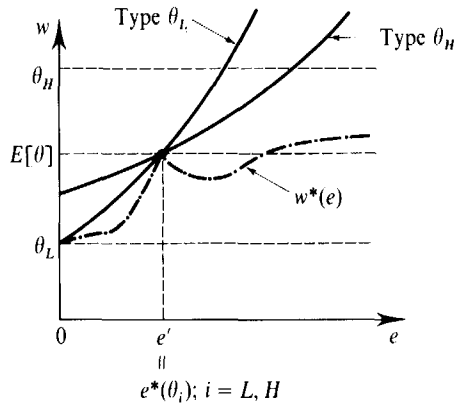


Figure 13.C.9 (left)
The highest-possible education level in a pooling equilibrium.

Figure 13.C.10 (right)
A pooling equilibrium.

The only remaining issue therefore concerns what levels of education can arise in a pooling equilibrium. It turns out that any education level between 0 and the level e' depicted in Figure 13.C.9 can be sustained.

Figure 13.C.10 shows an equilibrium supporting education level e'. Given the wage schedule depicted, each type of worker maximizes her payoff by choosing education level e'. This wage schedule is consistent with Bayesian updating on the equilibrium path because it gives a wage offer of $E[\theta]$ when education level e' is observed.

Education levels between 0 and e' can be supported in a similar manner. Education levels greater than e' cannot be sustained because a low-ability worker would rather set $e = 0$ than $e > e'$ even if this results in a wage payment of θ_L. Note that a pooling equilibrium in which both types of worker get no education Pareto dominates any pooling equilibrium with a positive education level. Once again, the Pareto-dominated pooling equilibria are sustained by the worker's fear that a deviation will lead firms to have an unfavorable impression of her ability. Note also that a pooling equilibrium in which both types of worker obtain no education results in exactly the same outcome as that which arises in the absence of an ability to signal. Thus, pooling equilibria are (weakly) Pareto dominated by the no-signaling outcome.

Multiple Equilibria and Equilibrium Refinement

The multiplicity of equilibria observed here is somewhat disconcerting. As we have seen, we can have separating equilibria in which firms learn the worker's type, but we can also have pooling equilibria where they do not; and within each type of equilibrium, many different equilibrium levels of education can arise. In large part, this multiplicity stems from the great freedom that we have to choose beliefs off the equilibrium path. Recently, a great deal of research has investigated the implications of putting "reasonable" restrictions on such beliefs along the lines we discussed in Section 9.D.

To see a simple example of this kind of reasoning, consider the separating equilibrium depicted in Figure 13.C.7. To sustain e_1 as the equilibrium education level of high-ability workers, firms must believe that any worker with an education level below e_1 has a positive probability of being of type θ_L. But consider any education level $\hat{e} \in (\tilde{e}, e_1)$. A type θ_L worker could never be made better off choosing such an education level than she is getting education level $e = 0$ *regardless of what*

firms believe about her as a result. Hence, any belief by firms upon seeing education level $\hat{e} > \tilde{e}$ other than $\mu(\hat{e}) = 1$ seems unreasonable. But if this is so, then we must have $w(\hat{e}) = \theta_H$, and so the high-ability worker would deviate to $\hat{e}$. In fact, by this logic, the only education level that can be chosen by type θ_H workers in a separating equilibrium involving reasonable beliefs is $\tilde{e}$.

In Appendix A we discuss in greater detail the use of these types of reasonable-beliefs refinements. One refinement proposed by Cho and Kreps (1987), known as the *intuitive criterion*, extends the idea discussed in the previous paragraph to rule out not only the dominated separating equilibria but also all pooling equilibria. Thus, if we accept the Cho and Kreps (1987) argument, we predict a *unique* outcome to this two-type signaling game: the best separating equilibrium outcome, which is shown in Figures 13.C.5 and 13.C.6.

Second-Best Market Intervention

In contrast with the market outcome predicted by the game-theoretic model studied in Section 13.B (the highest-wage competitive equilibrium), in the presence of signaling a central authority who cannot observe worker types may be able to achieve a Pareto improvement relative to the market outcome. To see this in the simplest manner, suppose that the Cho and Kreps (1987) argument predicting the best separating equilibrium outcome is correct. We have already seen that the best separating equilibrium can be Pareto dominated by the outcome that arises when signaling is impossible. When it is, a Pareto improvement can be achieved simply by banning the signaling activity.

In fact, it may be possible to achieve a Pareto improvement even when the no-signaling outcome does not Pareto dominate the best separating equilibrium. To see how, consider Figure 13.C.11. In the figure, the best separating equilibrium has low-ability workers at point $(\theta_L, 0)$ and high-ability workers at point $(\theta_H, \tilde{e})$. Note that the high-ability workers would be worse off if signaling were banned, since the point $(E[\theta], 0)$ gives them less than their equilibrium level of utility. Nevertheless, note that if we gave the low- and high-ability workers outcomes of $(\hat{w}_L, 0)$ and $(\hat{w}_H, \hat{e}_H)$, respectively, both types would be better off. The central authority can achieve this outcome by mandating that workers with education levels below $\hat{e}_H$ receive a wage of $\hat{w}_L$ and that workers with education levels of at least $\hat{e}_H$ receive a

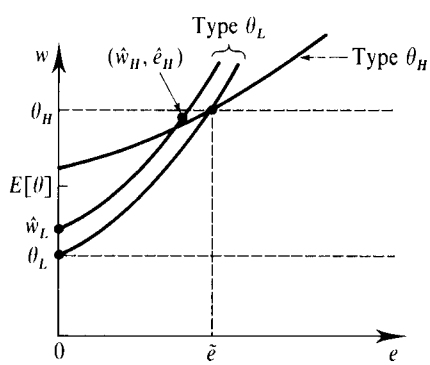


Figure 13.C.11
Achieving a Pareto improvement through cross-subsidization.

wage of $\hat{w}_H$. If so, low-ability workers would choose $e = 0$ and high-ability workers would choose $e = \hat{e}_H$. This alternative outcome involves firms incurring losses on low-ability workers and making profits on high-ability workers. However, as long as the firms break even on *average*, they are no worse off than before and a Pareto improvement has been achieved. The key to this Pareto improvement is that the central authority introduces *cross-subsidization*, where high-ability workers are paid less than their productivity level while low-ability workers are paid more than theirs, an outcome that cannot occur in a separating signaling equilibrium. (Note that the outcome when signaling is banned is an extreme case of cross-subsidization.)

Exercise 13.C.3: In the signaling model discussed in Section 13.C with $r(\theta_H) = r(\theta_L) = 0$, construct an example in which a central authority who does not observe worker types can achieve a Pareto improvement over the best separating equilibrium through a policy that involves cross-subsidization, but cannot achieve a Pareto improvement by simply banning the signaling activity. [*Hint*: Consider first a case with linear indifference curves.]

The case with $r(\theta_H) = r(\theta_L) = 0$ studied above, in which the market outcome in the absence of signaling is Pareto optimal, illustrates how the use of costly signaling can reduce welfare. Yet, when the market outcome in the absence of signaling is not efficient, signaling's ability to reveal information about worker types may instead create a Pareto improvement by leading to a more efficient allocation of labor. To see this point, suppose that we have $r = r(\theta_L) = r(\theta_H)$, with $\theta_L < r < \theta_H$ and $E[\theta] < r$. In this case, the equilibrium outcome without signaling has no workers employed. In contrast, any Pareto efficient outcome must have the high-ability workers employed by firms.

We now study the equilibrium outcome when signaling is possible. Consider, first, the wage and employment outcome that results after educational choice e by the worker. Following the worker's choice of educational level e, equilibrium behavior involves a wage of $w^*(e) = \mu(e)\theta_H + (1 - \mu(e))\theta_L$. If $w^*(e) \geq r$, then both types of workers would accept employment; if $w^*(e) < r$, then neither type would do so.

We now determine the equilibrium education choices of the two types of workers. Note first that any pooling equilibrium must have both types choosing $e = 0$ and neither type accepting employment. To see this, suppose that both types are choosing education level $\hat{e}$. Then $\mu(\hat{e}) = \lambda$ and $w^*(e) = E[\theta] < r$, and so neither type accepts employment. Hence, if $\hat{e} > 0$, both types would be better off choosing $e = 0$ instead. Thus, only an education level of zero is possible in a pooling equilibrium. In this zero education pooling equilibrium, the outcome is identical to the equilibrium outcome arising in the absence of the opportunity to signal.

The set of separating equilibria, on the other hand, is illustrated in Figure 13.C.12. In any separating equilibrium, a low-ability worker sets $e = 0$, is offered a wage of θ_L, and chooses to work at home, thereby achieving a utility of r. High-ability workers, on the other hand, select an education level in the interval $[\hat{e}, e_2]$ depicted in the figure, are offered a wage of θ_H, and accept employment. Note that no separating equilibrium can have $e^*(\theta_H) < \hat{e}$, since then low-ability workers would deviate and set $e = e^*(\theta_H)$; also, no separating equilibrium can have $e^*(\theta_H) > e_2$, since high-ability workers would then be better off setting $e = 0$ and working at home.

Note that in all these equilibria, both pooling and separating, the high-ability workers are weakly better off compared with the equilibrium arising without signaling opportunities and are strictly better off in separating equilibria with $e^*(\theta_H) < e_2$. Moreover, both the low-ability workers and the firms are equally well off. Thus, in the case with $\theta_L < r < \theta_H$ and $E[\theta] < r$, any pooling or separating signaling equilibrium weakly *Pareto dominates* the outcome arising

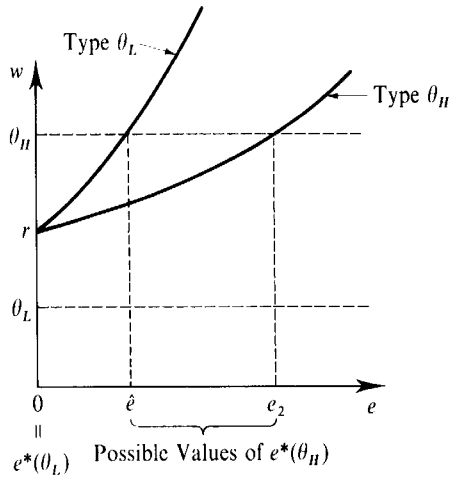


Figure 13.C.12
Separating equilibria when $r(\theta_L) = r(\theta_H) = r \in (\theta_L, \theta_H)$.

in the absence of signaling, and this Pareto dominance is *strict* for (essentially) all separating equilibria.

13.D Screening

In Section 13.C, we considered how signaling may develop in the marketplace as a response to the problem of asymmetric information about a good to be traded. There, individuals on the *more informed* side of the market (workers) chose their level of education in an attempt to signal information about their abilities to uninformed parties (the firms). In this section, we consider an alternative market response to the problem of unobservable worker productivity in which the *uninformed* parties take steps to try to distinguish, or *screen*, the various types of individuals on the other side of the market.[21] This possibility was first studied by Rothschild and Stiglitz (1976) and Wilson (1977) in the context of insurance markets (see Exercise 13.D.2).

As in Section 13.C, we focus on the case in which there are two types of workers, θ_L and θ_H, with $\theta_H > \theta_L > 0$ and where the fraction of workers who are of type θ_H is $\lambda \in (0, 1)$. In addition, workers earn nothing if they do not accept employment in a firm [in the notation used in Section 13.B, $r(\theta_L) = r(\theta_H) = 0$]. However, we now suppose that jobs may differ in the "task level" required of the worker. For example, jobs could differ in the number of hours per week that the worker is required to work. Or the task level might represent the speed at which a production line is run in a factory.

To make matters particularly simple, and to make the model parallel that in Section 13.C, we suppose that higher task levels add *nothing* to the output of the worker; rather, their *only* effect is to lower the utility of the worker.[22] The output of a type θ worker is therefore θ regardless of the worker's task level.

21. The setting analyzed here is one of *competitive screening* of workers, since we assume that there are several competing firms. See Section 14.C for a discussion of the *monopolistic screening* case, where a single firm screens workers.

22. As was true in the case of educational signaling, the assumption that higher task levels do not raise productivity is made purely for expositional purposes. Exercise 13.D.1 considers the case in which the firms' profits are increasing in the task level.

We assume that the utility of a type θ worker who receives wage w and faces task level $t \geq 0$ is

$$u(w, t \mid \theta) = w - c(t, \theta),$$

where $c(t, \theta)$ has all the properties assumed of the function $c(e, \theta)$ in Section 13.C. In particular, $c(0, \theta) = 0$, $c_t(t, \theta) > 0$, $c_{tt}(t, \theta) > 0$, $c_\theta(t, \theta) < 0$ for all $t > 0$, and $c_{t\theta}(t, \theta) < 0$. As will be clear shortly, the task level t serves to distinguish among types here in a manner that parallels the role of education in the signaling model discussed in Section 13.C.

Here we study the pure strategy subgame perfect Nash equilibria (SPNEs) of the following two-stage game:[23]

Stage 1: Two firms simultaneously announce sets of offered contracts. A contract is a pair (w, t). Each firm may announce any finite number of contracts.

Stage 2: Given the offers made by the firms, workers of each type choose whether to accept a contract and, if so, which one. For simplicity, we assume that if a worker is indifferent between two contracts, she always chooses the one with the lower task level and that she accepts employment if she is indifferent about doing so. If a worker's most preferred contract is offered by both firms, she accepts each firm's offer with probability $\frac{1}{2}$.

Thus, a firm can offer a variety of contracts; for example, it might have several production lines, each running at a different speed. Different types of workers may then end up choosing different contracts.[24]

It is helpful to start by considering what the outcome of this game would be if worker types were *observable*. To address this case, we allow firms to condition their offer on a worker's type (so that a firm can offer a contract (w_L, t_L) solely to type θ_L workers and another contract (w_H, t_H) solely to type θ_H workers).

Proposition 13.D.1: In any SPNE of the screening game with observable worker types, a type θ_i worker accepts contract $(w_i^*, t_i^*) = (\theta_i, 0)$, and firms earn zero profits.

Proof: We first argue that any contract (w_i^*, t_i^*) that workers of type θ_i accept in equilibrium must produce exactly zero profits; that is, it must involve a wage $w_i^* = \theta_i$. To see this, note that if $w_i^* > \theta_i$, then some firm is making a loss offering this contract and it would do better by not offering any contract to type θ_i workers. Suppose, on the other hand, that $w_i^* < \theta_i$, and let $\Pi > 0$ be the aggregate profits earned by the two firms on type θ_i workers. One of the two firms must be earning no more than $\Pi/2$ from these workers. If it deviates by offering a contract $(w_i^* + \varepsilon, t_i^*)$ for any

23. For this game, the set of subgame perfect Nash equilibria is identical to the sets of strategy profiles in weak perfect Bayesian equilibria or sequential equilibria.

24. The models in the original Rothschild and Stiglitz (1976) and Wilson (1977) analyses differ from our model in two respects. First, firms in those papers were restricted to offering only a single contract. This could make sense in the production line interpretation, for example, if each firm had only a single production line. Second, those authors allowed for "free entry," so that an additional firm could always enter if a profitable contracting opportunity existed. In fact, making these two changes has little effect on our conclusions. The only difference is in the precise conditions under which an equilibrium exists. (For more on this, see Exercise 13.D.4.)

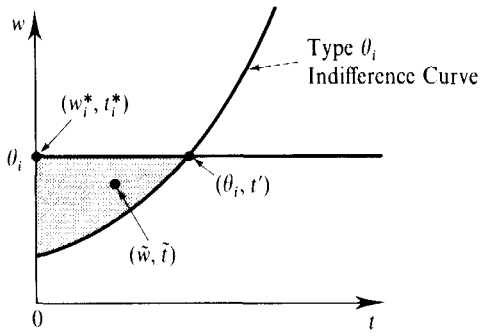


Figure 13.D.1 (left)
The equilibrium contract (w_i^*, t_i^*) for type θ_i with perfect observability.

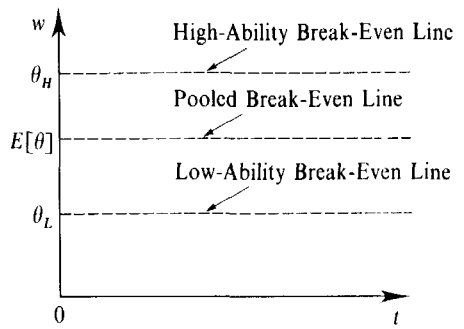


Figure 13.D.2 (right)
Break-even lines.

$\varepsilon > 0$, it will attract all type θ_i workers. Since ε can be made arbitrarily small, its profits from type θ_i workers can be made arbitrarily close to Π, and so this deviation will increase its profits. Thus, we must have $w_i^* = \theta_i$.

Now suppose that $(w_i^*, t_i^*) = (\theta_i, t')$ for some $t' > 0$. Then, as shown in Figure 13.D.1 (where the wage is measured on the vertical axis and the task level is measured on the horizontal axis), either firm could deviate and earn strictly positive profits by offering a contract in the shaded area of the figure, such as $(\tilde{w}, \tilde{t})$. The only contract at which there are no profitable deviations is $(w_i^*, t_i^*) = (\theta_i, 0)$, the contract that maximizes a type θ_i worker's utility subject to the constraint that the firms offering the contract break even. ■

We now turn to the situation in which worker types are *not observable*. In this case, each contract offered by a firm may in principle be accepted by either type of worker. We can note immediately that the complete information outcome identified in Proposition 13.D.1 cannot arise when worker types are unobservable: Because every low-ability worker prefers the high-ability contract $(\theta_H, 0)$ to contract $(\theta_L, 0)$, if these were the two contracts being offered by the firms then *all* workers would accept contract $(\theta_H, 0)$ and the firms would end up losing money.

To determine the equilibrium outcome with unobservable worker types, it is useful to begin by drawing three break-even lines: the zero-profit lines for productivity levels θ_L, $E[\theta]$, and θ_H, respectively. These three break-even lines are depicted by the dashed lines in Figure 13.D.2. The middle line represents the break-even line for a contract that attracts both types of workers, and we therefore refer to it as the *pooled* break-even line.

As in Section 13.C, we can in principle have two types of (pure strategy) equilibria: *separating* equilibria, in which the two types of workers accept different contracts, and *pooling* equilibria, in which both types of workers sign the same contract. (It can be shown that in any equilibrium both types of workers will accept some contract; we assume that this is so in the discussion that follows.) We proceed with a series of lemmas. Lemma 13.D.1 applies to both pooling and separating equilibria.

Lemma 13.D.1: In any equilibrium, whether pooling or separating, both firms must earn zero profits.

Proof: Let (w_L, t_L) and (w_H, t_H) be the contracts chosen by the low- and high-ability workers, respectively (these could be the same contract), and suppose that the two firms' aggregate profits are $\Pi > 0$. Then one firm must be making no more than $\Pi/2$. Consider a deviation by this firm in which it offers contracts $(w_L + \varepsilon, t_L)$ and

$(w_H + \varepsilon, t_H)$ for $\varepsilon > 0$. Contract $(w_L + \varepsilon, t_L)$ will attract all type θ_L workers, and contract $(w_H + \varepsilon, t_H)$ will attract all type θ_H workers. [Note that since type θ_i initially prefers contract (w_i, t_i) to (w_j, t_j), we have $w_i - c(t_i, \theta_i) \geq w_j - c(t_j, \theta_i)$, and so $(w_i + \varepsilon) - c(t_i, \theta_i) \geq (w_j + \varepsilon) - c(t_j, \theta_i)$.] Since ε can be chosen to be arbitrarily small, this deviation yields this firm profits arbitrarily close to Π, and so the firm has a profitable deviation. Thus, we must have $\Pi \leq 0$. Because no firm can incur a loss in any equilibrium (it could always earn zero by offering no contracts), both firms must in fact earn a profit of zero. ■

An important implication of Lemma 13.D.1 is that, in any equilibrium, no firm can have a deviation that allows it to earn strictly positive profits. We shall use this fact repeatedly in the discussion that follows. Using it, we immediately get the result given in Lemma 13.D.2 regarding pooling equilibria.

Lemma 13.D.2: No pooling equilibria exist.

Proof: Suppose that there is a pooling equilibrium contract (w^p, t^p). By Lemma 13.D.1, it lies on the pooled break-even line, as shown in Figure 13.D.3. Suppose that firm j is offering contract (w^p, t^p). Then firm $k \neq j$ has a deviation that yields it a strictly positive profit: It offers a single contract $(\tilde{w}, \tilde{t})$ that lies somewhere in the shaded region in Figure 13.D.3 and has $\tilde{w} < \theta_H$. This contract attracts all the type θ_H workers and none of the type θ_L workers, who prefer (w^p, t^p) over $(\tilde{w}, \tilde{t})$. Moreover, since $\tilde{w} < \theta_H$, firm k makes strictly positive profits from this contract when the high-ability workers accept it. ■

We now consider the possibilities for separating equilibria. Lemma 13.D.3 shows that all contracts accepted in a separating equilibrium must yield zero profits.

Lemma 13.D.3: If (w_L, t_L) and (w_H, t_H) are the contracts signed by the low- and high-ability workers in a separating equilibrium, then both contracts yield zero profits; that is, $w_L = \theta_L$ and $w_H = \theta_H$.

Proof: Suppose first that $w_L < \theta_L$. Then either firm could earn strictly positive profits by instead offering only contract $(\tilde{w}_L, t_L)$, where $\theta_L > \tilde{w}_L > w_L$. All low-ability workers would accept this contract; moreover, the deviating firm earns strictly positive profits from any worker (of low or high ability) who accepts it. Since Lemma 13.D.1 implies that no such deviation can exist in an equilibrium, we must have $w_L \geq \theta_L$ in any separating equilibrium.

Suppose, instead, that $w_H < \theta_H$, as in Figure 13.D.4. If we have a separating

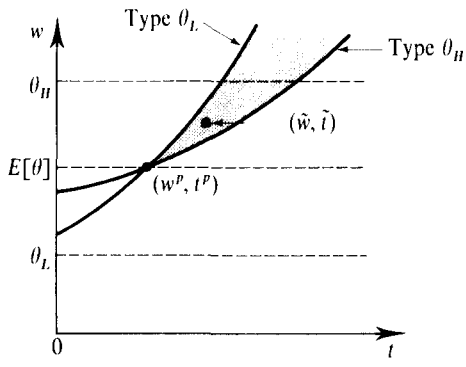


Figure 13.D.3 (left)
No pooling equilibria exist.

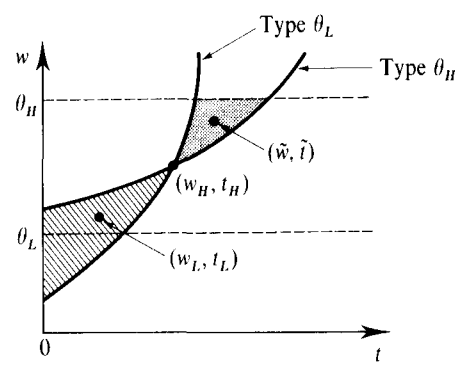


Figure 13.D.4 (right)
The high-ability contract in a separating equilibrium cannot have $w_H < \theta_H$.

equilibrium, then the type θ_L contract (w_L, t_L) must lie in the hatched region of the figure (by Lemma 13.D.1, it must also have $w_L > \theta_L$). To see this, note that since type θ_H workers choose contract (w_H, t_H), contract (w_L, t_L) must lie on or below the type θ_H indifference curve through (w_H, t_H), and since type θ_L workers choose (w_L, t_L) over (w_H, t_H), contract (w_L, t_L) must lie on or above the type θ_L indifference curve through (w_H, t_H). Suppose that firm j is offering the low-ability contract (w_L, t_L). Then firm $k \neq j$ could earn strictly positive profits by deviating and offering only a contract lying in the shaded region of the figure with a wage strictly less than θ_H, such as $(\tilde{w}, \tilde{t})$. This contract, which has $w_H < \theta_H$, will be accepted by all the type θ_H workers and by none of the type θ_L workers [since firm j will still be offering contract (w_L, t_L)]. So we must have $w_H \geq \theta_H$ in any separating equilibrium.

Since, by Lemma 13.D.1, firms break even in any equilibrium, we must in fact have $w_L = \theta_L$ and $w_H = \theta_H$. ■

Lemma 13.D.4 identifies the contract that must be accepted by low-ability workers in any separating equilibrium.

Lemma 13.D.4: In any separating equilibrium, the low-ability workers accept contract $(\theta_L, 0)$; that is, they receive the same contract as when no informational imperfections are present in the market.

Proof: By Lemma 13.D.3, $w_L = \theta_L$ in any separating equilibrium. Suppose that the low-ability workers' contract is instead some point (θ_L, t'_L) with $t'_L > 0$, as in Figure 13.D.5. (Although it is not important for the proof, the high-ability contract must then lie on the segment of the high-ability break-even line lying in the hatched region of the figure, as shown.) If so, then a firm can make strictly positive profits by offering only a contract lying in the shaded region of the figure, such as $(\tilde{w}, \tilde{t})$. All low-ability workers accept this contract, and the contract yields the firm strictly positive profits from any worker (of low or high ability) who accepts it. ■

We can now derive the high-ability workers' contract.

Lemma 13.D.5: In any separating equilibrium, the high-ability workers accept contract $(\theta_H, \hat{t}_H)$, where $\hat{t}_H$ satisfies $\theta_H - c(\hat{t}_H, \theta_L) = \theta_L - c(0, \theta_L)$.

Proof: Consider Figure 13.D.6. By Lemmas 13.D.3 and 13.D.4, we know that $(w_L, t_L) = (\theta_L, 0)$ and that $w_H = \theta_H$. In addition, if the type θ_L workers are willing to accept contract $(\theta_L, 0)$, t_H must be at least as large as the level $\hat{t}_H$ depicted in the

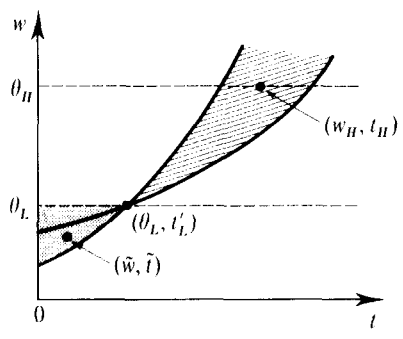


Figure 13.D.5 (left)
The low-ability workers must receive contract $(\theta_L, 0)$ in any separating equilibrium.

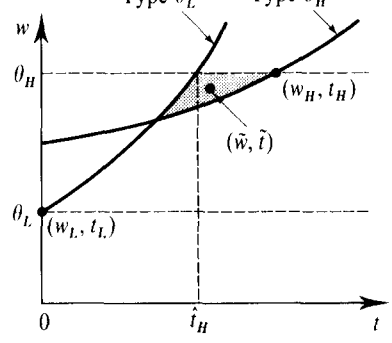


Figure 13.D.6 (right)
The high-ability workers must receive contract $(\theta_H, \hat{t}_H)$ in any separating equilibrium.

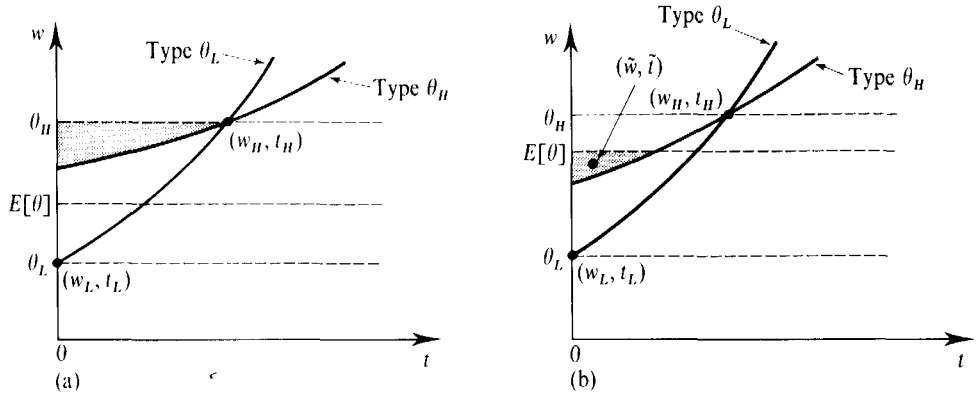


Figure 13.D.7
An equilibrium may not exist. (a) No pooling contract breaks the separating equilibrium. (b) The pooling contract $(\tilde{w}, \tilde{t})$ breaks the separating equilibrium.

figure. Note that low-ability workers are indifferent between contracts $(\theta_L, 0)$ and $(\theta_H, \hat{t}_H)$, and so $\theta_H - c(\hat{t}_H, \theta_L) = \theta_L - c(0, \theta_L)$. Suppose, then, that the high-ability contract (θ_H, t_H) has $t_H > \hat{t}_H$, as in the figure. Then either firm can earn a strictly positive profit by also offering, in addition to its current contracts, a contract lying in the shaded region of the figure with $w_H < \theta_H$, such as $(\tilde{w}, \tilde{t})$. This contract attracts all the high-ability workers and does not change the choice of the low-ability workers. Thus, in any separating equilibrium, the high-ability contract must be $(\theta_H, \hat{t}_H)$. ■

Proposition 13.D.2 summarizes the discussion so far.

Proposition 13.D.2: In any subgame perfect Nash equilibrium of the screening game, low-ability workers accept contract $(\theta_L, 0)$, and high-ability workers accept contract $(\theta_H, \hat{t}_H)$, where $\hat{t}_H$ satisfies $\theta_H - c(\hat{t}_H, \theta_L) = \theta_L - c(0, \theta_L)$.

Proposition 13.D.2 does not complete our analysis, however. Although we have established what any equilibrium must look like, we have not established that one exists. In fact, we now show that *one may not exist.*

Suppose that both firms are offering the two contracts identified in Proposition 13.D.2 and illustrated in Figure 13.D.7(a). Does either firm have an incentive to deviate? No firm can earn strictly positive profits by deviating in a manner that attracts either only high-ability or only low-ability workers (just try to find such a deviation). But what about a deviation that attracts *all* workers? Consider a deviation in which the deviating firm attracts all workers to a single pooling contract. In Figure 13.D.7(a), a contract can attract both types of workers if and only if it lies in the shaded region. There is no profitable deviation of this type if, as depicted in the figure, this shaded area lies completely above the pooled break-even line. However, when some of the shaded area lies strictly below the pooled break-even line, as in Figure 13.D.7(b), a profitable deviation to a pooling contract such as $(\tilde{w}, \tilde{t})$ exists. In this case, *no equilibrium exists.*

Even when no single pooling contract breaks the separating equilibrium, it is possible that a profitable deviation involving a pair of contracts may do so. For example, a firm can attract both types of workers by offering the contracts $(\tilde{w}_L, \tilde{t}_L)$ and $(\tilde{w}_H, \tilde{t}_H)$ depicted in Figure 13.D.8. When it does so, type θ_L workers accept contract $(\tilde{w}_L, \tilde{t}_L)$ and type θ_H workers accept $(\tilde{w}_H, \tilde{t}_H)$. If this pair of contracts yields the firm a positive profit, then this deviation breaks the separating contracts identified

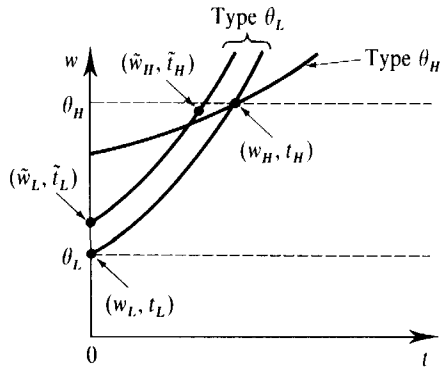


Figure 13.D.8
A profitable deviation using a pair of contracts may exist that breaks the separating equilibrium.

in Proposition 13.D.2 and no equilibrium exists. More generally, an equilibrium exists only if there is no such profitable deviation.

Welfare Properties of Screening Equilibria

Restricting attention to cases in which an equilibrium does exist, the screening equilibrium has welfare properties parallel to those of the signaling model's best separating equilibrium [with $r(\theta_L) = r(\theta_H) = 0$]. First, as in the earlier model, asymmetric information leads to Pareto inefficient outcomes. Here high-ability workers end up signing contracts that make them engage in completely unproductive and disutility-producing tasks merely to distinguish themselves from their less able counterparts. As in the signaling model, the low-ability workers are always worse off here when screening is possible than when it is not. One difference from the signaling model, however, is that in cases where an equilibrium exists, screening must make the high-ability workers better off; it is precisely in those cases where it would not that a move to a pooling contract breaks the separating equilibrium [see Figure 13.D.7(b)]. Indeed, when an equilibrium does exist, it is a constrained Pareto optimal outcome; if no firm has a deviation that can attract both types of workers and yield it a positive profit, then a central authority who is unable to observe worker types cannot achieve a Pareto improvement either.[25]

What can be said about the potential nonexistence of equilibrium in this model? Two paths have been followed in the literature. One approach is to establish existence of equilibria in the larger strategy space that allows for mixed strategies; on this, see Dasgupta and Maskin (1986). The other is to take the position that the lack of equilibria indicates that, in some important way, the model is incompletely specified. The aspect the literature has emphasized in this regard is the lack of any dynamic reactions to new contract offers [see Wilson (1977), Riley (1979), and Hellwig (1986)]. Wilson (1977), for example, uses a definition of equilibrium that captures the idea that firms are able to withdraw unprofitable contracts from the market. A set of contracts is a *Wilson equilibrium* if no firm has a profitable deviation that remains profitable once existing contracts that lose money after the deviation are withdrawn. This extra requirement may make deviations less attractive. In the deviation considered in Figure 13.D.3, for example, once contract $(\tilde{w}, \tilde{t})$ is introduced, the original contract (w^p, t^p) loses

25. Actually, there is a small gap: An equilibrium may exist when there is another pair of contracts that would give higher utility to both types of workers and that would yield the firm deviating to it exactly zero profits. In this case, the equilibrium is not a constrained Pareto optimum.

money. But if (w^p, t^p) is withdrawn as a result, then low-ability workers will accept $(\tilde{w}, \tilde{t})$ and this deviation ends up being unprofitable. Hellwig (1986) examines sequential equilibria and their refinements in a game that explicitly allows for such withdrawals.

By introducing such reactions, these papers establish the existence of pure strategy equilibria. Introducing reactions of this sort does not simply eliminate the nonexistence problem, however, but also yields somewhat different predictions regarding the characteristics of market equilibria and their welfare properties. For example, when firms can make multiple offers as we have allowed here, cross-subsidization can arise in Wilson equilibria. Indeed, Miyazaki (1977) shows that in the case in which multiple offers are possible, a Wilson equilibrium always exists and is necessarily a constrained Pareto optimum.

In the screening model examined above, we took the view that the uninformed firms made employment offers to the informed workers. Yet we could equally well imagine a model in which informed workers instead make contract offers to the firms. For example, each worker might propose a task level at which she is willing to work, and firms might then offer a wage for that task level. Note, however, that this alternative model exactly parallels the signaling model in Section 13.C and, as we have seen, yields quite different predictions. For example, the signaling model has numerous equilibria, but here we have at most a single equilibrium. This is somewhat disturbing. Given that our models are inevitably simplifications of actual market processes, if market outcomes are really very sensitive to issues such as this our models may provide us with little predictive ability.

One approach to this problem is offered by Maskin and Tirole (1992). They note that contracts like those we have allowed firms to offer in the screening model discussed in this section are still somewhat restricted. In particular, we could imagine a firm offering a worker a contract that involved an ex post (after signing) choice among a set of wage–task pairs (you will see more about contracts of this type in Section 14.C). Similarly, in considering the counterpart model in which workers make offers, we could allow a worker to propose such a contract. Maskin and Tirole (1992) show that with this enrichment of the allowed contracts (and a weak additional assumption) the sets of sequential equilibria of the two models coincide (there may be multiple equilibria in both cases).

APPENDIX A: REASONABLE-BELIEFS REFINEMENTS IN SIGNALING GAMES

In this appendix, we describe several commonly used reasonable-beliefs refinements of the perfect Bayesian and sequential equilibrium concepts for signaling games, and we apply them to the education signaling model discussed in Section 13.C. Excellent sources for further details and discussion are Cho and Kreps (1987) and Fudenberg and Tirole (1992).

Consider the following class of signaling games: There are I players plus nature. The first move of the game is nature's, who picks a "type" for player 1, $\theta \in \Theta = \{\theta_1, \ldots, \theta_N\}$. The probability of type θ is $f(\theta)$, and this is common knowledge among the players. However, only player 1 observes θ. The second move is player 1's, who picks an action a from set A after observing θ. Then, after seeing player 1's action choice (but not her type), each player $i = 2, \ldots, I$ simultaneously chooses an action s_i from set S_i. We define $S = S_2 \times \cdots \times S_I$. If player 1 is of type θ, her utility from choosing action a and having players $2, \ldots, I$ choose $s = (s_2, \ldots, s_I)$ is $u_1(a, s, \theta)$. Player $i \neq 1$ receives payoff $u_i(a, s, \theta)$ in this event. A perfect Bayesian

equilibrium (PBE) in the sense used in Section 13.C is a profile of strategies $(a(\theta), s_2(a), \ldots, s_I(a))$, combined with a common belief function $\mu(\theta \mid a)$ for players $2, \ldots, I$ that assigns a probability $\mu(\theta \mid a)$ to type θ of player 1 conditional on observing action $a \in A$, such that

(i) Player 1's strategy is optimal given the strategies of players $2, \ldots, I$.
(ii) The belief function $\mu(\theta \mid a)$ is derived from player 1's strategy using Bayes' rule where possible.
(iii) The strategies of players $2, \ldots, I$ specify actions following each choice $a \in A$ that constitute a Nash equilibrium of the simultaneous-move game in which the probability that player 1 is of type θ is $\mu(\theta \mid a)$ for all $\theta \in \Theta$.

In the context of the model under study here, this notion of a PBE is equivalent to the sequential equilibrium notion.

The education signaling model in Section 13.C falls into this category of signaling games if we do not explicitly model the worker's choice between the firms' offers and instead simply incorporate into the payoff functions the implications of her optimal choice (she chooses from among the firms offering the highest wage if this wage is positive and refuses both firms' offers otherwise). In that model, $I = 3$, $\Theta = \{\theta_L, \theta_H\}$, the set $A = \{e: e \geq 0\}$ contains the possible education choices of the worker, and the set $S_i = \{w: w \in \mathbb{R}\}$ contains the possible wage offers by firm i.

Domination-Based Refinements of Beliefs

The simplest reasonable-belief refinement of the PBE notion arises from the idea (discussed in Section 9.D) that reasonable beliefs should not assign positive probability to a player taking an action that is strictly dominated for her. In a signaling game, this problem can arise when players $2, \ldots, I$ (the firms in the education signaling model) assign a probability $\mu(\theta \mid a) > 0$ to player 1 (the worker) being of type θ after observing action a, even though action a is a strictly dominated choice for player 1 when she is of type θ.

Formally, we say that action $a \in A$ is a strictly dominated choice for type θ if there is an action $a' \in A$ such that

$$\min_{s' \in S} u_1(a', s', \theta) > \max_{s \in S} u_1(a, s, \theta).^{26} \tag{13.AA.1}$$

For each action $a \in A$, it is useful to define the set

$$\Theta(a) = \{\theta: \text{there is no } a' \in A \text{ satisfying (13.AA.1)}\}.$$

This is the set of types of player 1 for whom action a is not a strictly dominated choice. We can then say that a PBE has reasonable beliefs if, for all $a \in A$ with $\Theta(a) \neq \varnothing$,

$$\mu(\theta \mid a) > 0 \quad \text{only if} \quad \theta \in \Theta(a)$$

and we consider a PBE to be a sensible prediction only if it has reasonable beliefs.[27]

26. Note that a strategy $a(\theta)$ is strictly dominated for player 1 if and only if it involves play of a strictly dominated action for some type θ.

27. Doing this is equivalent to first eliminating each type θ's dominated actions from the game and then identifying the PBEs of this simplified game.

Unfortunately, in the education signaling model discussed in Section 13.C, this refinement does not narrow down our predictions at all. The set $\Theta(e)$ equals $\{\theta_L, \theta_H\}$ for all education levels e because either worker type will find e to be her optimal choice if the wage offered in response to e is sufficiently in excess of the wage offered at other education levels. Thus, no beliefs are ruled out, and all PBEs of the signaling game pass this test. If we want to narrow down our predictions for this model, we need to go beyond the use of refinements based only on notions of strict dominance.[28]

Recall the argument we made in Section 13.C for eliminating all separating equilibria but the best one. We argued that since, in Figure 13.C.7, a worker of type θ_L would be better off choosing $e = 0$ than she would choosing an education level above $\tilde{e}$ *for any beliefs and resulting equilibrium wage that might follow these two education levels*, no reasonable belief should assign a positive probability to a worker of type θ_L choosing any $e > \tilde{e}$. This is close to an argument that education levels $e > \tilde{e}$ are dominated choices for a type θ_L worker, but with the critical difference reflected in the italicized phrase: Only *equilibrium* responses of the firms are considered, rather than all conceivable responses. That is, we take a backward-induction-like view that the worker should only concern herself with possible equilibrium reactions to her education choices.

To be more formal about this idea, for any nonempty set $\hat{\Theta} \subset \Theta$, let $S^*(\hat{\Theta}, a) \subset S_2 \times \cdots \times S_I$ denote the set of possible equilibrium responses that can arise after action a is observed for *some* beliefs satisfying the property that $\mu(\theta \mid a) > 0$ only if $\theta \in \hat{\Theta}$. The set $S^*(\hat{\Theta}, a)$ contains the set of equilibrium responses by players $2, \ldots, I$ that can follow action choice a for some beliefs that assign positive probability only to types in $\hat{\Theta}$. When $\hat{\Theta} = \Theta$, the set of all conceivable types of player 1, this construction allows for all possible beliefs.[29] We can now say that action $a \in A$ is strictly dominated for type θ in this stronger sense if there exists an action a' with

$$\underset{s' \in S^*(\Theta, a')}{\text{Min}} \; u_1(a', s', \theta) > \underset{s \in S^*(\Theta, a)}{\text{Max}} \; u_1(a, s, \theta). \tag{13.AA.2}$$

Using this stronger notion of dominance, we can define the set

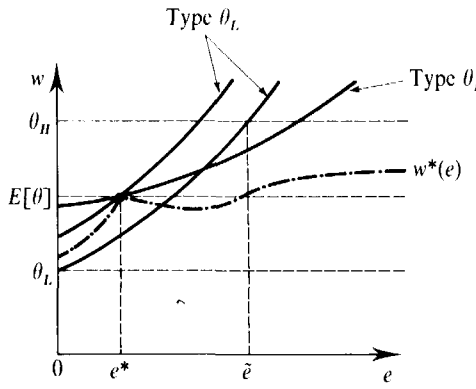
$$\Theta^*(a) = \{\theta\text{: there is no } a' \in A \text{ satisfying (13.AA.2)}\},$$

containing those types of player 1 for whom action a is not strictly dominated in the sense of (13.AA.2). We can now say that a PBE has reasonable beliefs if for all $a \in A$ with $\Theta^*(a) \neq \varnothing$, $\mu(a, \theta) > 0$ only if $\theta \in \Theta^*(a)$.

Using this reasonable-beliefs refinement significantly reduces the set of possible outcomes in the educational signaling model, sometimes even to a unique prediction. In that model, $S^*(\Theta, e) = [\theta_L, \theta_H]$ for all education choices e because, for any belief $\mu \in [0, 1]$, the resulting Nash equilibrium wage must lie between θ_L and θ_H. As a

28. We could, in principle, go further with this identification of strictly dominated strategies for player 1 by also eliminating any strictly dominated strategies for players $2, \ldots, I$, then looking to see whether we have any more strictly dominated actions for any of player 1's types, and so on. However, in the educational signaling model, this does not help us because the firms have no strictly dominated strategies.

29. Note that when there is only one player responding (so $I = 2$), the set $S^*(\Theta, a)$ is exactly the set of responses that are not strictly dominated for player 2 conditional on following action a. Note also that in this case a strategy $s_2(a)$ is weakly dominated for player 2 if, for any $a \in A$, it involves play of some $s \notin S^*(\Theta, a)$.



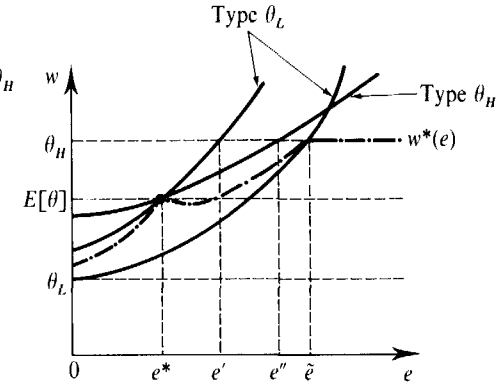


Figure 13.AA.1 (left)
A pooling equilibrium that is eliminated using the dominance test in (13.AA.2).

Figure 13.AA.2 (right)
A pooling equilibrium that is eliminated using the dominance test in (13.AA.3).

consequence, an education choice in excess of $\tilde{e}$ in Figure 13.C.7 is dominated for a type θ_L worker according to the test in (13.AA.2) by the education choice $e = 0$. Hence, in any PBE with reasonable beliefs, $\mu(\theta_H \mid e) = 1$ for all $e > \tilde{e}$. But if this is so, then no separating equilibrium with $e^*(\theta_H) > \tilde{e}$ can survive because, as we argued in Section 13.C, the high-ability worker will do better by deviating to an education level slightly in excess of $\tilde{e}$. Furthermore, we can also eliminate any pooling equilibrium in which the equilibrium outcome is worse for a high-ability worker than outcome $(\theta_H, \tilde{e})$, such as in the equilibrium depicted in Figure 13.AA.1, since any such equilibrium must involve unreasonable beliefs: If $\mu(\theta_H \mid e) = 1$ for all $e > \tilde{e}$, then a type θ_H worker could do better deviating to an education level just above $\tilde{e}$ where she would receive a wage of θ_H. In fact, when the high-ability worker prefers outcome $(\theta_H, \tilde{e})$ to $(E[\theta], 0)$, this argument rules out all pooling equilibria, and so we get the unique prediction of the best separating equilibrium.

Equilibrium Domination and the Intuitive Criterion

We now consider a further strengthening of the notion of dominance, known as *equilibrium dominance*. This leads to a refinement known as the *intuitive criterion* [Cho and Kreps (1987)] that always gives us the unique prediction of the best separating equilibrium in the two-type education signaling model studied in Section 13.C.

The idea behind this refinement can be seen by considering the pooling equilibrium of the education signaling model that is shown in Figure 13.AA.2, an equilibrium that is not eliminated by our previous refinements. Note that, as illustrated in the figure, to support education choice e^* as a pooling equilibrium outcome we must have beliefs for the firms satisfying $\mu(\theta_H \mid e) < 1$ for all $e \in (e', e'')$. Indeed, if $\mu(\theta_H \mid e) = 1$ at any such education level, then the wage offered would be θ_H and the type θ_H worker would find it optimal to deviate.

Suppose, however, that a firm is confronted with a deviation to some education level $\hat{e} \in (e', e'')$ when it was expecting the equilibrium level of education e^* to be chosen. It might reason as follows: "Either type of worker could be sure of getting outcome $(w, e) = (E[\theta], e^*)$ by choosing the equilibrium education level e^*. But a low-ability worker would be worse off deviating to education level e' *regardless* of what beliefs firms have after this choice, while a high-ability worker might be made better off by doing this. Thus, this must not be a low-ability worker." In this case, the choice of e' by the low-ability worker is dominated by her *equilibrium* payoff.

To formalize this idea in terms of our general specification, denote the equilibrium payoff to type θ in PBE $(a^*(\theta), s^*(a), \mu)$ by $u_1^*(\theta) = u_1(a^*(\theta), s^*(a^*(\theta)), \theta)$. We then say that action a is *equilibrium dominated* for type θ in PBE $(a^*(\theta), s^*(a), \mu)$ if

$$u_1^*(\theta) > \underset{s \in S^*(\Theta, a)}{\text{Max}} \; u_1(a, s, \theta). \tag{13.AA.3}$$

Using this notion of dominance, define for each $a \in A$ the set $\Theta^{**}(a) = \{\theta$: condition (13.AA.3) does not hold$\}$. We can now say that a PBE has reasonable beliefs if for all actions a with $\Theta^{**}(a) \neq \varnothing$, $\mu(\theta \mid a) > 0$ only if $\theta \in \Theta^{**}(a)$, and we can restrict attention to those PBEs that have reasonable beliefs.

Note that any action a that is dominated in the sense of (13.AA.2) for type θ must also be equilibrium dominated for this type because $u_1^*(\theta) = u_1^*(a^*(\theta), s^*(a^*(\theta)), \theta) > \text{Min}_{s' \in S^*(\Theta, a')} u_1(a', s', \theta)$ by the definition of a PBE. Thus, this equilibrium dominance-based procedure must rule out all the PBEs that were ruled out by our earlier procedure and may rule out more.

Consider the use of this refinement in the education signaling model of Section 13.C. Since it is stronger than the refinement based on (13.AA.2), this refinement also eliminates all but the best separating equilibrium. However, unlike our earlier dominance-based refinements, the equilibrium dominance-based refinement also eliminates *all* pooling equilibria. For example, in the pooling equilibrium depicted in Figure 13.AA.2, any education choice $\hat{e} \in (e', e'')$ is equilibrium dominated for the low-ability worker. Moreover, once the firms' beliefs following this education choice are restricted to assigning probability 1 to the worker being type θ_H, the high-ability worker wishes to deviate to this education level. Thus, we get a unique prediction for the outcome in this game: the best separating equilibrium.

In signaling games with two types, this equilibrium dominance-based refinement is equivalent to the *intuitive criterion* proposed in Cho and Kreps (1987). Formally, a PBE is said to violate the intuitive criterion if there exists a type θ and an action $a \in A$ such that

$$\underset{s \in S^*(\Theta^{**}(a), a)}{\text{Min}} \; u_1(a, s, \theta) > u_1^*(\theta). \tag{13.AA.4}$$

Thus, we eliminate a PBE using the intuitive criterion if there is some type θ who has a deviation that is *assured* of yielding her a payoff above her equilibrium payoff as long as players $2, \ldots, I$ do not assign a positive probability to the deviation having been made by any type θ for whom this action is equilibrium dominated. We can think of the intuitive criterion as saying that to eliminate a PBE we must find a type of player 1 who wants to deviate even if she is not sure what exact belief of players $2, \ldots, I$ will result, she is only sure that they will not think she is a type who would find the deviation to be an equilibrium-dominated action. In general, the intuitive criterion is a more conservative elimination procedure than just insisting on PBEs involving reasonable beliefs using set $\Theta^{**}(a)$ because any PBE with reasonable beliefs using set $\Theta^{**}(a)$ passes the intuitive criterion's test, but as Example 13.AA.1 illustrates, a PBE could satisfy the intuitive criterion's test but fail to have reasonable beliefs. However, when there are only two types of player 1, the two notions are equivalent.

Example 13.AA.1: Suppose that there are three types of player 1, $\{\theta_1, \theta_2, \theta_3\}$, and

that in some PBE the out-of-equilibrium action $\hat{a}$ is equilibrium dominated for type θ_1 only, so that $\Theta^{**}(\hat{a}) = \{\theta_2, \theta_3\}$. Suppose also that type θ_2 strictly prefers to deviate to action $\hat{a}$ if and only if beliefs over types θ_2 and θ_3 have $\mu(\theta_2 \mid \hat{a}) \geq \frac{1}{4}$ while type θ_3 strictly prefers to deviate to action $\hat{a}$ if and only if $\mu(\theta_2 \mid \hat{a}) \leq \frac{3}{4}$. This situation will not violate the intuitive criterion because condition (13.AA.4) does not hold for either type θ_2 or type θ_3. But in any PBE with reasonable beliefs using set $\Theta^{**}(a)$, one of the two types will deviate to action $\hat{a}$; therefore, this PBE must not have reasonable beliefs in this sense. When there are only two possible types for player 1, say θ_1 and θ_2, this difference disappears because whenever equilibrium domination eliminates a type from consideration, so that $\Theta^{*}(a) = \{\theta_i\}$ for $i = 1$ or 2, there is only one possible belief for players $2, \ldots, I$ to hold. ■

Although the use of either equilibrium domination or the intuitive criterion yields a unique prediction in the education signaling model when there are two types of workers, they do not accomplish this when there are three or more possible worker types (see Exercise 13.AA.1). Stronger refinements such as Banks and Sobel's (1987) notions of *divinity* and *universal divinity*, Cho and Kreps' (1987) related notion called *D1*, and Kohlberg and Mertens' (1986) *stability* do yield the unique prediction of the best separating equilibrium in these games with many worker types. See Cho and Kreps (1987) and Fudenberg and Tirole (1992) for further details.

REFERENCES

Akerlof, G. (1970). The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* **89**: 488–500.

Banks, J., and J. Sobel. (1987). Equilibrium selection in signaling games. *Econometrica* **55**: 647–62.

Cho, I-K., and D. M. Kreps. (1987). Signaling games and stable equilibria. *Quarterly Journal of Economics* **102**: 179–221.

Dasgupta, P., and E. Maskin. (1986). The existence of equilibrium in discontinuous economic games. *Review of Economic Studies* **46**: 1–41.

Fudenberg, D., and J. Tirole. (1992). *Game Theory*. Cambridge, Mass.: MIT Press.

Hellwig, M. (1986). Some recent developments in the theory of competition in markets with adverse selection. (University of Bonn, mimeographed).

Holmstrom, B., and R. B. Myerson. (1983). Efficient and durable decision rules with incomplete information. *Econometrica* **51**: 1799–819.

Kohlberg, E., and J.-F. Mertens. (1986). On the strategic stability of equilibria. *Econometrica* **54**: 1003–38.

Maskin, E., and J. Tirole. (1992). The principal-agent relationship with an informed principal, II: Common values. *Econometrica* **60**: 1–42.

Miyazaki, H. (1977). The rat race and internal labor markets. *Bell Journal of Economics* **8**: 394–418.

Riley, J. (1979). Informational equilibrium. *Econometrica* **47**: 331–59.

Rothschild, M., and J. E. Stiglitz. (1976). Equilibrium in competitive insurance markets: An essay in the economics of imperfect information. *Quarterly Journal of Economics* **80**: 629–49.

Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* **87**: 355–74.

Spence, A. M. (1974). *Market Signaling*. Cambridge, Mass.: Harvard University Press.

Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory* **16**: 167–207.

Wilson, C. (1980). The nature of equilibrium in markets with adverse selection. *Bell Journal of Economics* **11**: 108–30.

EXERCISES

13.B.1[A] Consider three functions of $\hat{\theta}$: $r(\hat{\theta})$, $E[\theta \mid \theta \leq \hat{\theta}]$, and $\hat{\theta}$. Graph these three functions over the domain $[\underline{\theta}, \bar{\theta}]$, assuming that the first two functions are continuous in $\hat{\theta}$ but allowing them to be otherwise quite arbitrary. Identify the competitive equilibria of the adverse selection model of Section 13.B using this diagram. What about the Pareto optimal labor allocation? Now produce a diagram to depict each of the situations in Figures 13.B.1 to 13.B.3.

13.B.2[B] Suppose that $r(\cdot)$ is a continuous and strictly increasing function and that there exists $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ such that $r(\theta) > \theta$ for $\theta > \hat{\theta}$ and $r(\theta) < \theta$ for $\theta < \hat{\theta}$. Let the density of workers of type θ be $f(\theta)$, with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. Show that a competitive equilibrium with unobservable worker types necessarily involves a Pareto inefficient outcome.

13.B.3[B] Consider a *positive selection* version of the model discussed in Section 13.B in which $r(\cdot)$ is a continuous, strictly *decreasing* function of θ. Let the density of workers of type θ be $f(\theta)$, with $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.

(a) Show that the *more capable* workers are the ones choosing to work at any given wage.

(b) Show that if $r(\theta) > \theta$ for all θ, then the resulting competitive equilibrium is Pareto efficient.

(c) Suppose that there exists a $\hat{\theta}$ such that $r(\theta) < \theta$ for $\theta > \hat{\theta}$ and $r(\theta) > \theta$ for $\theta < \hat{\theta}$. Show that any competitive equilibrium with strictly positive employment necessarily involves *too much* employment relative to the Pareto optimal allocation of workers.

13.B.4[B] Suppose two individuals, 1 and 2, are considering a trade at price p of an asset that they both use only as a store of wealth. Ms. 1 is currently the owner. Each individual i has a privately observed signal of the asset's worth y_i. In addition, each cares only about the expected value of the asset one year from now. Assume that a trade at price p takes place only if both parties think they are being made strictly better off. Prove that the probability of trade occurring is zero. [*Hint*: Study the following trading game: The two individuals simultaneously say either "trade" or "no trade," and a trade at price p takes place only if they both say "trade."]

13.B.5[B] Reconsider the case where $r(\theta) = r$ for all θ, but now assume that when the wage is such that no workers are accepting employment firms believe that any worker who might accept would be of the lowest quality, that is, $E[\theta \mid \Theta = \varnothing] = \underline{\theta}$. Maintain the assumption that all workers accept employment when indifferent.

(a) Argue that when $E[\theta] \geq r > \underline{\theta}$, there are now two competitive equilibria: one with $w^* = E[\theta]$ and $\Theta^* = [\underline{\theta}, \bar{\theta}]$ and one with $w^* = \underline{\theta}$ and $\Theta^* = \varnothing$. Also show that when $\underline{\theta} \geq r$ the unique competitive equilibrium is $w^* = E[\theta]$ and $\Theta^* = [\underline{\theta}, \bar{\theta}]$, and when $r > E[\theta]$ the unique competitive equilibrium is $w^* = \underline{\theta}$ and $\Theta^* = \varnothing$.

(b) Show that when $E[\theta] > r$ and there are two equilibria, the full-employment equilibrium Pareto dominates the no-employment one.

(c) Argue that when $E[\theta] \geq r$ the unique SPNE of the game-theoretic model in which two firms simultaneously make wage offers is the competitive equilibrium when this equilibrium is unique, and is the full-employment (highest-wage) competitive equilibrium when the competitive equilibrium is not unique and $E[\theta] > r$. What happens when $E[\theta] = r$? What about the case where $E[\theta] < r$?

(d) Argue that the highest-wage competitive equilibrium is a constrained Pareto optimum.

13.B.6[C] [Based on Wilson (1980)] Consider the following change in the adverse selection model of Section 13.B. Now there are N firms, each of which wants to hire at most 1 worker. The N firms differ in their productivity: In a firm of type γ a worker of type θ produces $\gamma\theta$ units of output. The parameter γ is distributed with density function $g(\cdot)$ on $[0, \infty]$, and $g(\gamma) > 0$ for all $\gamma \in [0, \infty]$.

(a) Let $z(w, \mu)$ denote the aggregate demand for labor when the wage is w and the average productivity of workers accepting employment at that wage is μ. Derive an expression for this function in terms of the density function $g(\cdot)$.

(b) Let $\mu(w) = E[\theta \mid r(\theta) \leq w]$, and define the *aggregate demand function for labor* by $z^*(w) = z(w, \mu(w))$. Show that $z^*(w)$ is strictly increasing in w at wage $\bar{w}$ if and only if the elasticity of μ with respect to w exceeds 1 at wage $\bar{w}$ (assume that all relevant functions are differentiable).

(c) Let $s(w) = \int_0^{r^{-1}(w)} f(\theta)\, d\theta$ denote the *aggregate supply function of labor*, and define a competitive equilibrium wage w^* as one where $z^*(w^*) = s(w^*)$. Show that if there are multiple competitive equilibria, then the one with the highest wage Pareto dominates all the others.

(d) Consider a game-theoretic model in which the firms make simultaneous wage offers, and denote the highest competitive equilibrium wage by w^*. Show that (i) only the highest-wage competitive equilibrium can arise as an SPNE, and (ii) the highest-wage competitive equilibrium is an SPNE if and only if $z^*(w) \leq z^*(w^*)$ for all $w > w^*$.

13.B.7[B] Suppose that it is impossible to observe worker types and consider a competitive equilibrium with wage rate w^*. Show that there is a Pareto-improving market intervention $(\tilde{w}_e, \tilde{w}_u)$ that reduces employment if and only if there is one of the form $(w_e, w_u) = (w^*, \hat{w}_u)$ with $\hat{w}_u > 0$. Similarly, argue that there is a Pareto-improving market intervention $(\tilde{w}_e, \tilde{w}_u)$ that increases employment if and only if there is one of the form $(w_e, w_u) = (\hat{w}_e, 0)$ with $\hat{w}_e > w^*$. Can you use these facts to give a simple proof of Proposition 13.B.2?

13.B.8[B] Consider the following alteration to the adverse selection model in Section 13.B. Imagine that when workers engage in home production, they use product x. Suppose that the amount consumed is related to a worker's type, with the relation given by the increasing function $x(\theta)$. Show that if a central authority can observe purchases of good x but not worker types, then there is a market intervention that results in a Pareto improvement even if the market is at the highest-wage competitive equilibrium.

13.B.9[B] Consider a model of *positive selection* in which $r(\cdot)$ is strictly decreasing and there are two types of workers, θ_H and θ_L, with $\infty > \theta_H > \theta_L > 0$. Let $\lambda = \text{Prob}\,(\theta = \theta_H) \in (0, 1)$. Assume that $r(\theta_H) < \theta_H$ and that $r(\theta_L) > \theta_L$. Show that the highest-wage competitive equilibrium need not be a constrained Pareto optimum. [*Hint*: Consider introducing a small unemployment benefit for a case in which $E[\theta] = r(\theta_L)$. Can you use the result in Exercise 13.B.7 to give an exact condition for when a competitive equilibrium involving full employment is a constrained Pareto optimum?]

13.B.10[B] Show that Proposition 13.B.2 continues to hold when $r(\theta) > \theta$ for some θ.

13.C.1[B] Consider a game in which, first, nature draws a worker's type from some continuous distribution on $[\underline{\theta}, \bar{\theta}]$. Once the worker observes her type, she can choose whether to submit to a costless test that reveals her ability perfectly. Finally, after observing whether the worker has taken the test and its outcome if she has, two firms bid for the worker's services. Prove that in any subgame perfect Nash equilibrium of this model all worker types submit to the test, and firms offer a wage no greater than $\underline{\theta}$ to any worker not doing so.

13.C.2[C] Reconsider the two-type signaling model with $r(\theta_L) = r(\theta_H) = 0$, assuming a worker's productivity is $\theta(1 + \mu e)$ with $\mu > 0$. Identify the separating and pooling perfect Bayesian equilibria, and relate them to the perfect information competitive outcome.

13.C.3[B] In text.

13.C.4[B] Reconsider the signaling model discussed in Section 13.C, now assuming that worker types are drawn from the interval $[\underline{\theta}, \bar{\theta}]$ with a density function $f(\theta)$ that is strictly positive everywhere on this interval. Let the cost function be $c(e, \theta) = (e^2/\theta)$. Derive the (unique) perfect Bayesian equilibrium.

13.C.5[B] Assume a single firm and a single consumer. The firm's product may be either high or low quality and is of high quality with probability λ. The consumer cannot observe quality before purchase and is risk neutral. The consumer's valuation of a high-quality product is v_H; her valuation of a low-quality product is v_L. The costs of production for high (H) and low (L) quality are c_H and c_L, respectively. The consumer desires at most one unit of the product. Finally, the firm's price is regulated and is set at p. Assume that $v_H > p > v_L > c_H > c_L$.

(a) Given the level of p, under what conditions will the consumer buy the product?

(b) Suppose that before the consumer decides whether to buy, the firm (which knows its type) can advertise. Advertising conveys no information directly, but consumers can observe the total amount of money that the firm is spending on advertising, denoted by A. Can there be a separating perfect Bayesian equilibrium, that is, an equilibrium in which the consumer rationally expects firms with different quality levels to pick different levels of advertising?

13.C.6[C] Consider a market for loans to finance investment projects. All investment projects require an outlay of 1 dollar. There are two types of projects: good and bad. A good project has a probability of p_G of yielding profits of $\Pi > 0$ and a probability $(1 - p_G)$ of yielding profits of zero. For a bad project, the relative probabilities are p_B and $(1 - p_B)$, respectively, where $p_G > p_B$. The fraction of projects that are good is $\lambda \in (0, 1)$.

Entrepreneurs go to banks to borrow the cash to make the initial outlay (assume for now that they borrow the entire amount). A loan contract specifies an amount R that is supposed to be repaid to the bank. Entrepreneurs know the type of project they have, but the banks do not. In the event that a project yields profits of zero, the entrepreneur defaults on her loan contract, and the bank receives nothing. Banks are competitive and risk neutral. The risk-free rate of interest (the rate the banks pay to borrow funds) is r. Assume that

$$p_G\Pi - (1 + r) > 0 > p_B\Pi - (1 + r).$$

(a) Find the equilibrium level of R and the set of projects financed. How does this depend on p_G, p_B, λ, Π, and r?

(b) Now suppose that the entrepreneur can offer to contribute some fraction x of the 1 dollar initial outlay from her own funds ($x \in [0, 1]$). The entrepreneur is liquidity constrained, however, so that the effective cost of doing so is $(1 + \rho)x$, where $\rho > r$.

(i) What is an entrepreneur's payoff as a function of her project type, her loan-repayment amount R, and her contribution x?

(ii) Describe the best (from a welfare perspective) separating perfect Bayesian equilibrium of a game in which the entrepreneur first makes an offer that specifies the level of x she is willing to put into a project, banks then respond by making offers specifying the level of R they would require, and finally the entrepreneur accepts a bank's offer or decides not to go ahead with the project. How does the amount contributed by entrepreneurs with good projects change with small changes in p_B, p_G, λ, Π, and r?

(iii) How do the two types of entrepreneurs do in the separating equilibrium of **(b)**(ii) compared with the equilibrium in **(a)**?

13.D.1[B] Extend the screening model to a case in which tasks are productive. Assume that a type θ worker produces $\theta(1 + \mu t)$ units of output when her task level is t where $\mu > 0$. Identify the subgame perfect Nash equilibria of this model.

13.D.2[B] Consider the following model of the insurance market. There are two types of individuals: high risk and low risk. Each starts with initial wealth W but has a chance that an accident (e.g., a fire) will reduce her wealth by L. The probability of this happening is p_L for low-risk types and p_H for high-risk types, where $p_H > p_L$. Both types are expected utility maximizers with a Bernoulli utility function over wealth of $u(w)$, with $u'(w) > 0$ and $u''(w) < 0$ at all w. There are two risk-neutral insurance companies. An insurance policy consists of a premium payment M made by the insured individual to her insurance firm and a payment R from the insurance company to the insured individual in the event of a loss.

(a) Suppose that individuals are prohibited from buying more than one insurance policy. Argue that a policy can be thought of as specifying the wealth levels of the insured individual in the two states "no loss" and "loss."

(b) Assume that the insurance companies simultaneously offer policies; as in Section 13.D, they can each offer any finite number of policies. What are the subgame perfect Nash equilibrium outcomes of the model? Does an equilibrium necessarily exist?

13.D.3[C] Consider the following extension of the model you developed in Exercise 13.D.1. Suppose that there is a fixed task level T that all workers face. The monetary equivalent cost of accepting employment at this task level is $c > 0$, which is independent of worker type. However, now a worker's actual output is observable and verifiable, and so contracts can base compensation on the worker's ex post observed output level.

(a) What is the subgame perfect Nash equilibrium outcome of this model?

(b) Now suppose that the output realization is random. It can be either good (q_G) or bad (q_B). The probability that it is good is p_H for a high-ability worker and p_L for a low-ability worker ($p_H > p_L$). If workers are risk-neutral expected utility maximizers with a Bernoulli utility function over wealth of $u(w) = w$, what is the subgame perfect Nash equilibrium outcome?

(c) What if workers are strictly risk averse with $u''(w) < 0$ at all w?

13.D.4[B] Reconsider the screening model in Section 13.D, but assume that (i) there is an infinite number of firms that could potentially enter the industry and (ii) firms can each offer at most one contract. [The implication of (i) is that, in any SPNE, no firm can have a profitable entry opportunity.] Characterize the equilibria for this case.

13.AA.1[C] Consider the extension of the signaling model discussed in Section 13.C to the case of three types. Assume all three types have $r(\theta) = 0$. Provide an example in which more than one perfect Bayesian equilibrium satisfies the intuitive criterion.

CHAPTER

14

The Principal-Agent Problem

14.A Introduction

In Chapter 13, we considered situations in which asymmetries of information exist between individuals at the time of contracting. In this chapter, we shift our attention to asymmetries of information that develop *subsequent* to the signing of a contract.

Even when informational asymmetries do not exist at the time of contracting, the parties to a contract often anticipate that asymmetries will develop sometime after the contract is signed. For example, after an owner of a firm hires a manager, the owner may be unable to observe how much effort the manager puts into the job. Similarly, the manager will often end up having better information than the owner about the opportunities available to the firm.

Anticipating the development of such informational asymmetries, the contracting parties seek to design a contract that mitigates the difficulties they cause. These problems are endemic to situations in which one individual hires another to take some action for him as his "agent." For this reason, this contract design problem has come to be known as the *principal-agent problem.*

The literature has traditionally distinguished between two types of informational problems that can arise in these settings: those resulting from *hidden actions* and those resulting from *hidden information.* The hidden action case, also known as *moral hazard,* is illustrated by the owner's inability to observe how hard his manager is working; the manager's coming to possess superior information about the firm's opportunities, on the other hand, is an example of hidden information.[1]

Although many economic situations (and some of the literature) contain elements of both types of problems, it is useful to begin by studying each in isolation. In Section 14.B, we introduce and study a model of hidden actions. Section 14.C analyzes

1. The literature's use of the term *moral hazard* is not entirely uniform. The term originates in the insurance literature, which first focused attention on two types of informational imperfections: the "moral hazard" that arises when an insurance company cannot observe whether the insured exerts effort to prevent a loss and the "adverse selection" (see Section 13.B) that occurs when the insured knows more than the company at the time he purchases a policy about his likelihood of an accident. Some authors use moral hazard to refer to either of the hidden action or hidden information variants of the principal-agent problem [see, for example, Hart and Holmstrom (1987)]. Here, however, we use the term in the original sense.

a hidden information model. Then, in Section 14.D, we provide a brief discussion of hybrid models that contain both of these features. We shall see that the presence of postcontractual asymmetric information often leads to welfare losses for the contracting parties relative to what would be achievable in the absence of these informational imperfections.

It is important to emphasize the broad range of economic relationships that fit into the general framework of the principal-agent problem. The owner–manager relationship is only one example; others include insurance companies and insured individuals (the insurance company cannot observe how much care is exercised by the insured), manufacturers and their distributors (the manufacturer may not be able to observe the market conditions faced by the distributor), a firm and its workforce (the firm may have more information than its workers about the true state of demand for its products and therefore about the value of the workers' product), and banks and borrowers (the bank may have difficulty observing whether the borrower uses the loaned funds for the purpose for which the loan was granted). As would be expected given this diversity of examples, the principal-agent framework has found application in a broad range of applied fields in economics. Our discussion will focus on the owner–manager problem.

The analysis in this chapter, particularly that in Section 14.C, is closely related to that in two other chapters. First, the techniques developed in Section 14.C can be applied to the analysis of screening problems in which, in contrast with the case studied in Section 13.D, only one uninformed party screens informed individuals. We discuss the analysis of this *monopolistic screening problem* in small type at the end of Section 14.C. Second, the principal-agent problem is actually a special case of "mechanism design," the topic of Chapter 23. Thus, the material here constitutes a first pass at this more general issue. Mastery of the fundamentals of the principal-agent problem, particularly the material in Section 14.C, will be helpful when you study Chapter 23.

A good source for further reading on topics of this chapter is Hart and Holmstrom (1987).

14.B Hidden Actions (Moral Hazard)

Imagine the owner of a firm (the *principal*) wishes to hire a manager (the *agent*) for a one-time project. The project's profits are affected, at least in part, by the manager's actions. If these actions were observable, the contracting problem between the owner and the manager would be relatively straightforward; the contract would simply specify the exact actions to be taken by the manager and the compensation (wage payment) that the owner is to provide in return.[2] When the manager's actions are not observable, however, the contract can no longer specify them in an effective manner, because there is simply no way to verify whether the manager has fulfilled his obligations. In this circumstance, the owner must design the manager's compensation scheme in a way that *indirectly* gives him the incentive to take the correct

2. Note that this requires not only that the manager's actions be observable to the owner but also that they be observable to any court that might be called upon to enforce the contract.

actions (those that would be contracted for if his actions were observable). In this section, we study this contract design problem.

To be more specific, let π denote the project's (observable) profits, and let e denote the manager's action choice. The set of possible actions is denoted by E. We interpret e as measuring managerial effort. In the simplest case that is widely studied in the literature, e is a one-dimensional measure of how "hard" the manager works, and so $E \subset \mathbb{R}$. More generally, however, managerial effort can have many dimensions—how hard the manager works to reduce costs, how much time he spends soliciting customers, and so on—and so e could be a vector with each of its elements measuring managerial effort in a distinct activity. In this case, $E \subset \mathbb{R}^M$ for some M.[3] In our discussion, we shall refer to e as the manager's *effort choice* or *effort level.*

For the nonobservability of managerial effort to have any consequence, the manager's effort must not be perfectly deducible from observation of π. Hence, to make things interesting (and realistic), we assume that although the project's profits are affected by e, they are not fully determined by it. In particular, we assume that the firm's profit can take values in $[\underline{\pi}, \bar{\pi}]$ and that it is stochastically related to e in a manner described by the conditional density function $f(\pi \mid e)$, with $f(\pi \mid e) > 0$ for all $e \in E$ and all $\pi \in [\underline{\pi}, \bar{\pi}]$. Thus, any potential realization of π can arise following any given effort choice by the manager.

In the discussion that follows, we restrict our attention to the case in which the manager has only two possible effort choices, e_H and e_L (see Appendix A for a discussion of the case in which the manager has many possible actions), and we make assumptions implying that e_H is a "high-effort" choice that leads to a higher profit level for the firm than e_L but entails greater difficulty for the manager. This fact will mean that there is a conflict between the interests of the owner and those of the manager.

More specifically, we assume that the distribution of π conditional on e_H first-order stochastically dominates the distribution conditional on e_L; that is, the distribution functions $F(\pi \mid e_L)$ and $F(\pi \mid e_H)$ satisfy $F(\pi \mid e_H) \leq F(\pi \mid e_L)$ at all $\pi \in [\underline{\pi}, \bar{\pi}]$, with strict inequality on some open set $\Pi \subset [\underline{\pi}, \bar{\pi}]$ (see Section 6.D). This implies that the level of expected profits when the manager chooses e_H is larger than that from e_L: $\int \pi f(\pi \mid e_H)\, d\pi > \int \pi f(\pi \mid e_L)\, d\pi$.

The manager is an expected utility maximizer with a Bernoulli utility function $u(w, e)$ over his wage w and effort level e. This function satisfies $u_w(w, e) > 0$ and $u_{ww}(w, e) \leq 0$ at all (w, e) (subscripts here denote partial derivatives) and $u(w, e_H) < u(w, e_L)$ at all w; that is, the manager prefers more income to less, is weakly risk averse over income lotteries, and dislikes a high level of effort.[4] In what follows, we focus on a special case of this utility function that has attracted much of the

3. In fact, more general interpretations are possible. For example, e could include non-effort-related managerial decisions such as what kind of inputs are purchased or the strategies that are adopted for appealing to buyers. We stick to the effort interpretation largely because it helps with intuition.

4. Note that in the multidimensional-effort case, it need not be that e_H has higher effort in every dimension; the only important thing for our analysis is that it leads to higher profits and entails a larger managerial disutility than does e_L.

attention in the literature: $u(w, e) = v(w) - g(e)$.[5] For this case, our assumptions on $u(w, e)$ imply that $v'(w) > 0$, $v''(w) \leq 0$, and $g(e_H) > g(e_L)$.

The owner receives the project's profits less any wage payments made to the manager. We assume that the owner is risk neutral and therefore that his objective is to maximize his expected return. The idea behind this simplifying assumption is that the owner may hold a well-diversified portfolio that allows him to diversify away the risk from this project. (Exercise 14.B.2 asks you to consider the case of a risk-averse owner.)

The Optimal Contract when Effort is Observable

It is useful to begin our analysis by looking at the optimal contracting problem when effort is observable.

Suppose that the owner chooses a contract to offer the manager that the manager can then either accept or reject. A contract here specifies the manager's effort $e \in \{e_L, e_H\}$ and his wage payment as a function of observed profits $w(\pi)$. We assume that a competitive market for managers dictates that the owner must provide the manager with an expected utility level of at least $\bar{u}$ if he is to accept the owner's contract offer ($\bar{u}$ is the manager's *reservation utility level*). If the manager rejects the owner's contract offer, the owner receives a payoff of zero.

We assume throughout that the owner finds it worthwhile to make the manager an offer that he will accept. The optimal contract for the owner then solves the following problem (for notational simplicity, we suppress the lower and upper limits of integration $\underline{\pi}$ and $\bar{\pi}$):

$$\max_{e \in \{e_L, e_H\}, w(\pi)} \int (\pi - w(\pi)) f(\pi \mid e)\, d\pi \qquad \text{(14.B.1)}$$

$$\text{s.t.} \int v(w(\pi)) f(\pi \mid e)\, d\pi - g(e) \geq \bar{u}.$$

It is convenient to think of this problem in two stages. First, for each choice of e that might be specified in the contract, what is the best compensation scheme $w(\pi)$ to offer the manager? Second, what is the best choice of e?

Given that the contract specifies effort level e, choosing $w(\pi)$ to maximize $\int (\pi - w(\pi)) f(\pi \mid e)\, d\pi = (\int \pi f(\pi \mid e)\, d\pi) - (\int w(\pi) f(\pi \mid e)\, d\pi)$ is equivalent to minimizing the expected value of the owner's compensation costs, $\int w(\pi) f(\pi \mid e)\, d\pi$, so (14.B.1) tells us that the optimal compensation scheme in this case solves

$$\min_{w(\pi)} \int w(\pi) f(\pi \mid e)\, d\pi \qquad \text{(14.B.2)}$$

$$\text{s.t.} \int v(w(\pi)) f(\pi \mid e)\, d\pi - g(e) \geq \bar{u}.$$

The constraint in (14.B.2) always binds at a solution to this problem; otherwise, the owner could lower the manager's wages while still getting him to accept the contract. Letting γ denote the multiplier on this constraint, at a solution to problem (14.B.2) the manager's wage $w(\pi)$ at each level of $\pi \in [\underline{\pi}, \bar{\pi}]$ must satisfy the first-order

5. Exercise 14.B.1 considers one implication of relaxing this assumption.

condition[6]

$$-f(\pi \mid e) + \gamma v'(w(\pi)) f(\pi \mid e) = 0,$$

or

$$\frac{1}{v'(w(\pi))} = \gamma. \tag{14.B.3}$$

If the manager is strictly risk averse [so that $v'(w)$ is strictly decreasing in w], the implication of condition (14.B.3) is that the optimal compensation scheme $w(\pi)$ is a constant; that is, the owner should provide the manager with a fixed wage payment. This finding is just a risk-sharing result: Given that the contract explicitly dictates the manager's effort choice and that there is no problem with providing incentives, the risk-neutral owner should fully insure the risk-averse manager against any risk in his income stream (in a manner similar to that in Example 6.C.1). Hence, given the contract's specification of e, the owner offers a fixed wage payment w_e^* such that the manager receives exactly his reservation utility level:

$$v(w_e^*) - g(e) = \bar{u}. \tag{14.B.4}$$

Note that since $g(e_H) > g(e_L)$, the manager's wage will be higher if the contract calls for effort e_H than if it calls for e_L.

On the other hand, when the manager is risk neutral, say with $v(w) = w$, condition (14.B.3) is necessarily satisfied for *any* compensation function. In this case, because there is no need for insurance, a fixed wage scheme is merely one of many possible optimal compensation schemes. Any compensation function $w(\pi)$ that gives the manager an expected wage payment equal to $\bar{u} + g(e)$ [the level derived from condition (14.B.4) when $v(w) = w$] is also optimal.

Now consider the optimal choice of e. The owner optimally specifies the effort level $e \in \{e_L, e_H\}$ that maximizes his expected profits less wage payments,

$$\int \pi f(\pi \mid e)\, d\pi - v^{-1}(\bar{u} + g(e)). \tag{14.B.5}$$

The first term in (14.B.5) represents the gross profit when the manager puts forth effort e; the second term represents the wages that must be paid to compensate the manager for this effort [derived from condition (14.B.4)]. Whether e_H or e_L is optimal depends on the incremental increase in expected profits from e_H over e_L compared with the monetary cost of the incremental disutility it causes the manager.

This is summarized in Proposition 14.B.1.

Proposition 14.B.1: In the principal-agent model with observable managerial effort, an optimal contract specifies that the manager choose the effort e^* that maximizes $[\int \pi f(\pi \mid e)\, d\pi - v^{-1}(\bar{u} + g(e))]$ and pays the manager a fixed wage $w^* = v^{-1}(\bar{u} + g(e^*))$. This is the uniquely optimal contract if $v''(w) < 0$ at all w.

6. The first-order condition for $w(\pi)$ is derived by taking the derivative with respect to the manager's wage at each level of π separately. To see this point, consider a discrete version of the model in which there is a finite number of possible profit levels $(\pi_1, \ldots, \pi_N)$ and associated wage levels $(w_1, \ldots, w_N)$. The first-order condition (14.B.3) is analogous to the condition one gets in the discrete model by examining the first-order conditions for each w_n, $n = 1, \ldots, N$ (note that we allow the wage payment to be negative). To be rigorous, we should add that when we have a continuum of possible levels of π, an optimal compensation scheme need only satisfy condition (14.B.3) at a set of profit levels that is of full measure.

The Optimal Contract when Effort is Not Observable

The optimal contract described in Proposition 14.B.1 accomplishes two goals: it specifies an efficient effort choice by the manager, and it fully insures him against income risk. When effort is not observable, however, these two goals often come into conflict because the only way to get the manager to work hard is to relate his pay to the realization of profits, which is random. When these goals come into conflict, the nonobservability of effort leads to inefficiencies.

To highlight this point, we first study the case in which the manager is risk neutral. We show that in this case, where the risk-bearing concern is absent, the owner can still achieve the same outcome as when effort is observable. We then study the optimal contract when the manager is risk averse. In this case, whenever the first-best (full observability) contract would involve the high-effort level, efficient risk bearing and efficient incentive provision come into conflict, and the presence of nonobservable actions leads to a welfare loss.

A risk-neutral manager

Suppose that $v(w) = w$. Applying Proposition 14.B.1, the optimal effort level e^* when effort is observable solves

$$\max_{e \in \{e_l, e_h\}} \int \pi f(\pi \mid e)\, d\pi - g(e) - \bar{u}. \tag{14.B.6}$$

The owner's profit in this case is the value of expression (14.B.6), and the manager receives an expected utility of exactly $\bar{u}$.

Now consider the owner's payoff when the manager's effort is not observable. In Proposition 14.B.2, we establish that the owner can still achieve his full-information payoff.

Proposition 14.B.2: In the principal-agent model with unobservable managerial effort and a risk-neutral manager, an optimal contract generates the same effort choice and expected utilities for the manager and the owner as when effort is observable.

Proof: We show explicitly that there is a contract the owner can offer that gives him the same payoff that he receives under full information. This contract must therefore be an optimal contract for the owner because the owner can never do better when effort is not observable than when it is (when effort is observable, the owner is always free to offer the optimal nonobservability contract and simply leave the choice of an effort level up to the manager).

Suppose that the owner offers a compensation schedule of the form $w(\pi) = \pi - \alpha$, where α is some constant. This compensation schedule can be interpreted as "selling the project to the manager" because it gives the manager the full return π except for the fixed payment α (the "sales price"). If the manager accepts this contract, he chooses e to maximize his expected utility,

$$\int w(\pi) f(\pi \mid e)\, d\pi - g(e) = \int \pi f(\pi \mid e)\, d\pi - \alpha - g(e). \tag{14.B.7}$$

Comparing (14.B.7) with (14.B.6), we see that e^* maximizes (14.B.7). Thus, this contract induces the first-best (full observability) effort level e^*.

The manager is willing to accept this contract as long as it gives him an expected utility of at least $\bar{u}$, that is, as long as

$$\int \pi f(\pi \mid e^*)\, d\pi - \alpha - g(e^*) \geq \bar{u}. \tag{14.B.8}$$

Let α^* be the level of α at which (14.B.8) holds with equality. Note that the owner's payoff if the compensation scheme is $w(\pi) = \pi - \alpha^*$ is exactly α^* (the manager gets all of π except for the fixed payment α^*). Rearranging (14.B.8), we see that $\alpha^* = \int \pi f(\pi \mid e^*)\, d\pi - g(e^*) - \bar{u}$. Hence, with compensation scheme $w(\pi) = \pi - \alpha^*$, both the owner and the manager get exactly the same payoff as when effort is observable. ■

The basic idea behind Proposition 14.B.2 is straightforward. If the manager is risk neutral, the problem of risk sharing disappears. Efficient incentives can be provided without incurring any risk-bearing losses by having the manager receive the full marginal returns from his effort.

A risk-averse manager

When the manager is strictly risk averse over income lotteries, matters become more complicated. Now incentives for high effort can be provided only at the cost of having the manager face risk. To characterize the optimal contract in these circumstances, we again consider the contract design problem in two steps: first, we characterize the optimal incentive scheme for each effort level that the owner might want the manager to select; second, we consider which effort level the owner should induce.

The optimal incentive scheme for implementing a specific effort level e minimizes the owner's expected wage payment subject to two constraints. As before, the manager must receive an expected utility of at least $\bar{u}$ if he is to accept the contract. When the manager's effort is unobservable, however, the owner also faces a second constraint: The manager must actually *desire* to choose effort e when facing the incentive scheme. Formally, the optimal incentive scheme for implementing e must therefore solve

$$\begin{aligned} \underset{w(\pi)}{\text{Min}} \quad & \int w(\pi) f(\pi \mid e)\, d\pi \qquad (14.B.9) \\ \text{s.t.} \quad & \text{(i)} \int v(w(\pi)) f(\pi \mid e)\, d\pi - g(e) \geq \bar{u} \\ & \text{(ii)} \quad e \text{ solves } \underset{\tilde{e}}{\text{Max}} \int v(w(\pi)) f(\pi \mid \tilde{e})\, d\pi - g(\tilde{e}). \end{aligned}$$

Constraint (ii) is known as the *incentive constraint*: it insures that under compensation scheme $w(\pi)$ the manager's optimal effort choice is e.

How does the owner optimally implement each of the two possible levels of e? We consider each in turn.

Implementing e_L: Suppose, first, that the owner wishes to implement effort level e_L. In this case, the owner optimally offers the manager the fixed wage payment $w_e^* = v^{-1}(\bar{u} + g(e_L))$, the same payment he would offer if contractually specifying effort e_L when effort is observable. To see this, note that with this compensation

scheme the manager selects e_L: His wage payment is unaffected by his effort, and so he will choose the effort level that involves the lowest disutility, namely e_L. Doing so, he earns exactly $\bar{u}$. Hence, this contract implements e_L at exactly the same cost as when effort is observable. But, as we noted in the proof of Proposition 14.B.2, the owner can never do better when effort is unobservable than when effort is observable [formally, in problem (14.B.9), the owner faces the additional constraint (ii) relative to problem (14.B.2)]; therefore, this must be a solution to problem (14.B.9).

Implementing e_H: The more interesting case arises when the owner decides to induce effort level e_H. In this case, constraint (ii) of (14.B.9) can be written as

$$(\text{ii}_H) \quad \int v(w(\pi))\, f(\pi \mid e_H)\, d\pi - g(e_H) \geq \int v(w(\pi))\, f(\pi \mid e_L)\, d\pi - g(e_L).$$

Letting $\gamma \geq 0$ and $\mu \geq 0$ denote the multipliers on constraints (i) and (ii_H), respectively, $w(\pi)$ must satisfy the following Kuhn–Tucker first-order condition at every $\pi \in [\underline{\pi}, \bar{\pi}]$:[7]

$$-f(\pi \mid e_H) + \gamma v'(w(\pi))\, f(\pi \mid e_H) + \mu[f(\pi \mid e_H) - f(\pi \mid e_L)]v'(w(\pi)) = 0$$

or

$$\frac{1}{v'(w(\pi))} = \gamma + \mu\left[1 - \frac{f(\pi \mid e_L)}{f(\pi \mid e_H)}\right]. \tag{14.B.10}$$

We first establish that in any solution to problem (14.B.9), where $e = e_H$, both γ and μ are strictly positive.

Lemma 14.B.1: In any solution to problem (14.B.9) with $e = e_H$, both $\gamma > 0$ and $\mu > 0$.

Proof: Suppose that $\gamma = 0$. Because $F(\pi \mid e_H)$ first-order stochastically dominates $F(\pi \mid e_L)$, there must exist an open set of profit levels $\tilde{\Pi} \subset [\underline{\pi}, \bar{\pi}]$ such that $[f(\pi \mid e_L)/f(\pi \mid e_H)] > 1$ at all $\pi \in \tilde{\Pi}$. But if $\gamma = 0$, condition (14.B.10) then implies that $v'(w(\pi)) \leq 0$ at any such π (recall that $\mu \geq 0$), which is impossible. Hence, $\gamma > 0$.

On the other hand, if $\mu = 0$ in the solution to problem (14.B.9) then, by condition (14.B.10), the optimal compensation schedule gives a fixed wage payment for every profit realization. But we know that this would lead the manager to choose e_L rather than e_H, violating constraint (ii_H) of problem (14.B.9). Hence, $\mu > 0$. ■

7. Although problem (14.B.9) may not appear to be a convex programming problem, a simple transformation of the problem shows that (14.B.10) is both a necessary and a sufficient condition for a solution. To see this, reformulate (14.B.9) as a problem of choosing the manager's level of utility for each profit outcome π, say $\bar{v}(\pi)$. Letting $\phi(\cdot) = v^{-1}(\cdot)$, the objective function becomes $\int \phi(\bar{v}(\pi))f(\pi \mid e_H)\, d\pi$, which is convex in $\bar{v}(\pi)$, and the constraints are then all linear in $\bar{v}(\pi)$. Thus, (Kuhn–Tucker) first-order conditions are both necessary and sufficient for a maximum of this reformulated problem (see Section M.K of the Mathematical Appendix). The first-order condition for this problem is

$$-\phi'(\bar{v}(\pi))f(\pi \mid e_H) + \gamma f(\pi \mid e_H) + \mu[f(\pi \mid e_H) - f(\pi \mid e_L)] = 0 \quad \text{for all } \pi \in [\underline{\pi}, \bar{\pi}].$$

Defining $w(\pi)$ by $v(w(\pi)) = \bar{v}(\pi)$, and noting that $\phi'(v(w(\pi))) = 1/v'(w(\pi))$, this gives (14.B.10).

Lemma 14.B.1 tells us that both constraints in problem (14.B.9) bind when $e = e_H$.[8] Moreover, given Lemma 14.B.1, condition (14.B.10) can be used to derive some useful insights into the shape of the optimal compensation schedule. Consider, for example, the fixed wage payment $\hat{w}$ such that $(1/v'(\hat{w})) = \gamma$. According to condition (14.B.10),

$$w(\pi) > \hat{w} \qquad \text{if} \qquad \frac{f(\pi \mid e_L)}{f(\pi \mid e_H)} < 1$$

and

$$w(\pi) < \hat{w} \qquad \text{if} \qquad \frac{f(\pi \mid e_L)}{f(\pi \mid e_H)} > 1.$$

This relationship is fairly intuitive. The optimal compensation scheme pays more than $\hat{w}$ for outcomes that are statistically relatively more likely to occur under e_H than under e_L in the sense of having a likelihood ratio $[f(\pi \mid e_L)/f(\pi \mid e_H)]$ less than 1. Similarly, it offers less compensation for outcomes that are relatively more likely when e_L is chosen. We should stress, however, that while this condition evokes a statistical interpretation, there is no actual statistical inference going on here; the owner *knows* what level of effort will be chosen given the compensation schedule he offers. Rather, the compensation package has this form because of its *incentive effects*. That is, by structuring compensation in this way, it provides the manager with an incentive for choosing e_H instead of e_L.

This point leads to what may at first seem a somewhat surprising implication: in an optimal incentive scheme, compensation is not necessarily monotonically increasing in profits. As is clear from examination of condition (14.B.10), for the optimal compensation scheme to be monotonically increasing, it must be that the likelihood ratio $[f(\pi \mid e_L)/f(\pi \mid e_H)]$ is decreasing in π; that is, as π increases, the likelihood of getting profit level π if effort is e_H relative to the likelihood if effort is e_L must increase. This property, known as the *monotone likelihood ratio property* [see Milgrom (1981)], is *not* implied by first-order stochastic dominance. Figures 14.B.1(a) and (b), for example, depict a case in which the distribution of π conditional on e_H stochastically dominates the distribution of π conditional on e_L but the monotone likelihood ratio property does not hold. In this example, increases in effort serve to convert low profit realizations into intermediate ones but have no effect on the likelihood of very high profit realizations. Condition (14.B.10) tells us that in this case, we should have higher wages at intermediate levels of profit than at very high ones because it is the likelihood of intermediate profit levels that is sensitive to increases in effort. The optimal compensation function for this example is shown in Figure 14.B.1(c).

8. A more direct argument for constraint (i) being binding goes as follows: Suppose that $w(\pi)$ is a solution to (14.B.9) in which constraint (i) is not binding. Consider a change in the compensation function that lowers the wage paid at each level of π in such a way that the resulting decrease in utility is equal at all π, that is, to a new function $\hat{w}(\pi)$ with $[v(w(\pi)) - v(\hat{w}(\pi))] = \Delta v > 0$ at all $\pi \in [\underline{\pi}, \bar{\pi}]$. This change does not affect the satisfaction of the incentive constraint (ii_H) since if the manager was willing to pick e_H when faced with $w(\pi)$, he will do so when faced with $\hat{w}(\pi)$. Furthermore, because constraint (i) is not binding, the manager will still accept this new contract if Δv is small enough. Lastly, the owner's expected wage payments will be lower than under $w(\pi)$. This yields a contradiction.

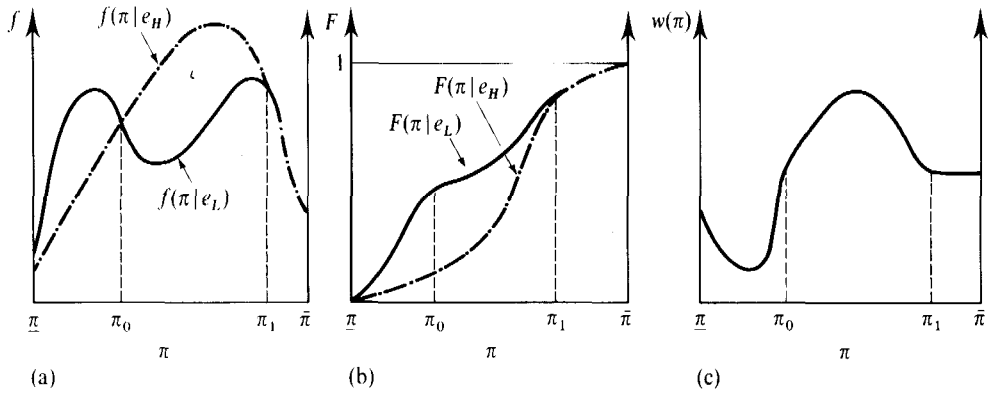


Figure 14.B.1
A violation of the monotone likelihood ratio property. (a) Densities. (b) Distribution functions. (c) Optimal wage scheme.

Condition (14.B.10) also implies that the optimal contract is not likely to take a simple (e.g., linear) form. The optimal shape of $w(\pi)$ is a function of the informational content of various profit levels (through the likelihood ratio), and this is unlikely to vary with π in a simple manner in most problems.

Finally, note that given the variability that is optimally introduced into the manager's compensation, the expected value of the manager's wage payment must be strictly greater than his (fixed) wage payment in the observable case, $w^*_{e_H} = v^{-1}(\bar{u} + g(e_H))$. Intuitively, because the manager must be assured an expected utility level of $\bar{u}$, the owner must compensate him through a higher average wage payment for any risk he bears. To see this point formally, note that since $E[v(w(\pi))|e_H] = \bar{u} + g(e_H)$ and $v''(\cdot) < 0$, Jensen's inequality (see Section M.C of the Mathematical Appendix) tells us that $v(E[w(\pi)|e_H]) > \bar{u} + g(e_H)$. But we know that $v(w^*_{e_H}) = \bar{u} + g(e_H)$, and so $E[w(\pi)|e_H] > w^*_{e_H}$. As a result, nonobservability increases the owner's expected compensation costs of implementing effort level e_H.

Given the preceding analysis, which effort level should the owner induce? As before, the owner compares the incremental change in expected profits from the two effort levels $[\int \pi f(\pi|e_H)\,d\pi - \int \pi f(\pi|e_L)\,d\pi]$ with the difference in expected wage payments in the contracts that optimally implement each of them, that is, with the difference in the value of problem (14.B.9) for $e = e_H$ compared with $e = e_L$.

From the preceding analysis, we know that the wage payment when implementing e_L is exactly the same as when effort is observable, whereas the expected wage payment when the owner implements e_H under nonobservability is strictly larger than his payment in the observable case. Thus, in this model, nonobservability raises the cost of implementing e_H and does not change the cost of implementing e_L. The implication of this fact is that nonobservability of effort can lead to an inefficiently low level of effort being implemented. When e_L would be the optimal effort level if effort were observable, then it still is when effort is nonobservable. In this case, nonobservability causes no losses. In contrast, when e_H would be optimal if effort were observable, then one of two things may happen: it may be optimal to implement e_H using an incentive scheme that faces the manager with risk; alternatively, the risk-bearing costs may be high enough that the owner decides that it is better to

simply implement e_L. In either case, nonobservability causes a welfare loss to the owner (the manager's expected utility is $\bar{u}$ in either case).[9]

These observations are summarized in Proposition 14.B.3.

Proposition 14.B.3: In the principal-agent model with unobservable manager effort, a risk-averse manager, and two possible effort choices, the optimal compensation scheme for implementing e_H satisfies condition (14.B.10), gives the manager expected utility $\bar{u}$, and involves a larger expected wage payment than is required when effort is observable. The optimal compensation scheme for implementing e_L involves the same fixed wage payment as if effort were observable. Whenever the optimal effort level with observable effort would be e_H, nonobservability causes a welfare loss.

The fact that nonobservability leads in this model only to *downward* distortions in the manager's effort level is a special feature of the two-effort-level specification. With many possible effort choices, nonobservability may still alter the level of managerial effort induced in an optimal contract from its level under full observability, but the direction of the bias can be upward as well as downward. (See Exercise 14.B.4 for an illustration.)

Imagine that another statistical signal of effort, say y, is available to the owner in addition to the realization of profits, and that the joint density of π and y given e is given by $f(\pi, y \mid e)$. In this case, the manager's compensation can, in principle, be made to depend on both π and y. When should compensation be made a function of this variable as well? That is, when does the optimal compensation function $w(\pi, y)$ actually depend on y?

To answer this question, suppose that the owner wishes to implement e_H. Following along the same lines as above, we can derive a condition analogous to condition (14.B.10):

$$\frac{1}{v'(w(\pi, y))} = \gamma + \mu\left[1 - \frac{f(\pi, y \mid e_L)}{f(\pi, y \mid e_H)}\right]. \tag{14.B.11}$$

Consider, first, the case in which y is simply a noisy random variable that is unrelated to e. Then we can write the density $f(\pi, y \mid e)$ as the product of two densities, $f_1(\pi \mid e)$ and $f_2(y)$: $f(\pi, y \mid e) = f_1(\pi \mid e) f_2(y)$. Substituting into (14.B.11), the $f_2(\cdot)$ terms cancel out, and so the optimal compensation package is independent of y.

The intuition behind this result is straightforward. Suppose that the owner is initially offering a contract that has wage payments dependent on y. Intuitively, this contract induces a randomness in the manager's wage that is unrelated to e and therefore makes the manager face risk without achieving any beneficial incentive effect. If the owner instead offers, for each realization of π, the certain payment $\bar{w}(\pi)$ such that

$$v(\bar{w}(\pi)) = E[v(w(\pi, y)) \mid \pi] = \int v(w(\pi, y)) f_2(y)\, dy,$$

9. Note, however, that although nonobservability leads to a welfare loss, the outcome here is a constrained Pareto optimum in the sense introduced in Section 13.B. To see this, note that the owner maximizes his profit subject to giving the manager an expected utility level no less than $\bar{u}$ and subject to constraints deriving from his inability to observe the manager's effort choice. As a result, no allocation that Pareto dominates this outcome can be achieved by a central authority who cannot observe the manager's effort choice. For market intervention by such an authority to generate a Pareto improvement, there must be externalities among the contracts signed by different pairs of individuals.

then the manager gets exactly the same expected utility under $\bar{w}(\pi)$ as under $w(\pi, y)$ for any level of effort he chooses. Thus, the manager's effort choice will be unchanged, and he will still accept the contract. However, because the manager faces less risk, the expected wage payments are lower and the owner is better off (this again follows from Jensen's inequality: for all π, $v(E[w(\pi, y)\,|\,\pi]) > E[v(w(\pi, y))\,|\,\pi]$, and so $\bar{w}(\pi) < E[w(\pi, y)\,|\,\pi]$).

This point can be pushed further. Note that we can always write

$$f(\pi, y\,|\,e) = f_1(\pi\,|\,e)f_2(y\,|\,\pi, e).$$

If $f_2(y\,|\,\pi, e)$ does not depend on e, then the $f_2(\cdot)$ terms in condition (14.B.11) again cancel out and the optimal compensation package does not depend on y. This condition on $f_2(y\,|\,\pi, e)$ is equivalent to the statistical concept that π is a *sufficient statistic* for y with respect to e. The converse is also true: As long as π is *not* a sufficient statistic for y, then wages *should* be made to depend on y, at least to some degree. See Holmstrom (1979) for further details.

A number of extensions of this basic analysis have been studied in the literature. For example, Holmstrom (1982), Nalebuff and Stiglitz (1983), and Green and Stokey (1983) examine cases in which many managers are being hired and consider the use of relative performance evaluation in such settings; Bernheim and Whinston (1986), on the other hand, extend the model in the other direction, examining settings in which a single agent is hired simultaneously by several principals; Dye (1986) considers cases in which effort may be observed through costly monitoring; Rogerson (1985a), Allen (1985), and Fudenberg, Holmstrom, and Milgrom (1990) examine situations in which the agency relationship is repeated over many periods, with a particular focus on the extent to which long-term contracts are more effective at resolving agency problems than is a sequence of short-term contracts of the type we analyzed in this section. (This list of extensions is hardly exhaustive.) Many of these analyses focus on the case in which effort is single-dimensional; Holmstrom and Milgrom (1991) discuss some interesting aspects of the more realistic case of multidimensional effort.

Holmstrom and Milgrom (1987) have pursued another interesting extension. Bothered by the simplicity of real-world compensation schemes relative to the optimal contracts derived in models like the one we have studied here, they investigate a model in which profits accrue incrementally over time and the manager is able to adjust his effort during the course of the project in response to early profit realizations. They identify conditions under which the owner can restrict himself without loss to the use of compensation schemes that are *linear* functions of the project's total profit. The optimality of linear compensation schemes arises because of the need to offer incentives that are "robust" in the sense that they continue to provide incentives regardless of how early profit realizations turn out. Their analysis illustrates a more general idea, namely, that complicating the nature of the incentive problem can actually lead to simpler forms for optimal contracts. For illustrations of this point, see Exercises 14.B.5 and 14.B.6.

The exercises at the end of the chapter explore some of these extensions.

14.C Hidden Information (and Monopolistic Screening)

In this section, we shift our focus to a setting in which the postcontractual informational asymmetry takes the form of hidden information.

Once again, an owner wishes to hire a manager to run a one-time project. Now, however, the manager's effort level, denoted by e, is fully observable. What is not observable after the contract is signed is the random realization of the manager's disutility from effort. For example, the manager may come to find himself well suited to the tasks required at the firm, in which case high effort has a relatively low disutility associated with it, or the opposite may be true. However, only the manager comes to know which case obtains.[10]

Before proceeding, we note that the techniques we develop here can also be applied to models of *monopolistic screening* where, in a setting characterized by *precontractual* informational asymmetries, a single uninformed individual offers a menu of contracts in order to distinguish, or *screen*, informed agents who have differing information at the time of contracting (see Section 13.D for an analysis of a competitive screening model). We discuss this connection further in small type at the end of this section.

To formulate our hidden information principal-agent model, we suppose that effort can be measured by a one-dimensional variable $e \in [0, \infty)$. Gross profits (excluding any wage payments to the manager) are a simple deterministic function of effort, $\pi(e)$, with $\pi(0) = 0$, $\pi'(e) > 0$, and $\pi''(e) < 0$ for all e.

The manager is an expected utility maximizer whose Bernoulli utility function over wages and effort, $u(w, e, \theta)$, depends on a state of nature θ that is realized after the contract is signed and that only the manager observes. We assume that $\theta \in \mathbb{R}$, and we focus on a special form of $u(w, e, \theta)$ that is widely used in the literature:[11]

$$u(w, e, \theta) = v(w - g(e, \theta)).$$

The function $g(e, \theta)$ measures the disutility of effort in monetary units. We assume that $g(0, \theta) = 0$ for all θ and, letting subscripts denote partial derivatives, that

$$g_e(e, \theta) \begin{cases} > 0 & \text{for } e > 0 \\ = 0 & \text{for } e = 0 \end{cases}$$

$$g_{ee}(e, \theta) > 0 \quad \text{for all } e$$

$$g_\theta(e, \theta) < 0 \quad \text{for all } e$$

$$g_{e\theta}(e, \theta) \begin{cases} < 0 & \text{for } e > 0 \\ = 0 & \text{for } e = 0. \end{cases}$$

Thus, the manager is averse to increases in effort, and this aversion is larger the greater the current level of effort. In addition, higher values of θ are more productive states in the sense that both the manager's total disutility from effort, $g(e, \theta)$, and his marginal disutility from effort at any current effort level, $g_e(e, \theta)$, are lower when θ

10. A seemingly more important source of hidden information between managers and owners is that the manager of a firm often comes to know more about the potential profitability of various actions than does the owner. In Section 14.D, we discuss one hybrid hidden action–hidden information model that captures this alternative sort of informational asymmetry; its formal analysis reduces to that of the model studied here.

11. Exercise 14.C.3 asks you to consider an alternative form for the manager's utility function.

is greater. We also assume that the manager is strictly risk averse, with $v''(\cdot) < 0$.[12] As in Section 14.B, the manager's reservation utility level, the level of expected utility he must receive if he is to accept the owner's contract offer, is denoted by $\bar{u}$. Note that our assumptions about $g(e, \theta)$ imply that the manager's indifference curves have the single-crossing property discussed in Section 13.C.

Finally, for expositional purposes, we focus on the simple case in which θ can take only one of two values, θ_H and θ_L, with $\theta_H > \theta_L$ and $\text{Prob}(\theta_H) = \lambda \in (0, 1)$. (Exercise 14.C.1 asks you to consider the case of an arbitrary finite number of states.)

A contract must try to accomplish two objectives here: first, as in Section 14.B, the risk-neutral owner should insure the manager against fluctuations in his income; second, although there is no problem here in insuring that the manager puts in effort (because the contract can explicitly state the effort level required), a contract that maximizes the surplus available in the relationship (and hence, the owner's payoff) must make the level of managerial effort responsive to the disutility incurred by the manager, that is, to the state θ. To fix ideas, we first illustrate how these goals are accomplished when θ is observable; we then turn to an analysis of the problems that arise when θ is observed only by the manager.

The State θ is Observable

If θ is observable, a contract can directly specify the effort level and remuneration of the manager contingent on each realization of θ (note that these variables fully determine the economic outcomes for the two parties). Thus, a complete information contract consists of two wage–effort pairs: $(w_H, e_H) \in \mathbb{R} \times \mathbb{R}_+$ for state θ_H and $(w_L, e_L) \in \mathbb{R} \times \mathbb{R}_+$ for state θ_L. The owner optimally chooses these pairs to solve the following problem:

$$\begin{aligned} \underset{\substack{w_L, e_L \geq 0 \\ w_H, e_H \geq 0}}{\text{Max}} \quad & \lambda[\pi(e_H) - w_H] + (1 - \lambda)[\pi(e_L) - w_L] \\ \text{s.t.} \quad & \lambda v(w_H - g(e_H, \theta_H)) + (1 - \lambda)v(w_L - g(e_L, \theta_L)) \geq \bar{u}. \end{aligned} \tag{14.C.1}$$

In any solution $[(w_L^*, e_L^*), (w_H^*, e_H^*)]$ to problem (14.C.1) the reservation utility constraint must bind; otherwise, the owner could lower the level of wages offered and still have the manager accept the contract. In addition, letting $\gamma \geq 0$ denote the multiplier on this constraint, the solution must satisfy the following first-order conditions:

$$-\lambda + \gamma\lambda v'(w_H^* - g(e_H^*, \theta_H)) = 0. \tag{14.C.2}$$

$$-(1 - \lambda) + \gamma(1 - \lambda)\, v'(w_L^* - g(e_L^*, \theta_L)) = 0. \tag{14.C.3}$$

$$\lambda\pi'(e_H^*) - \gamma\lambda v'(w_H^* - g(e_H^*, \theta_H))\, g_e(e_H^*, \theta_H) \begin{cases} \leq 0, \\ = 0 \quad \text{if } e_H^* > 0. \end{cases} \tag{14.C.4}$$

$$(1 - \lambda)\pi'(e_L^*) - \gamma(1 - \lambda)\, v'(w_L^* - g(e_L^*, \theta_L))\, g_e(e_L^*, \theta_L) \begin{cases} \leq 0, \\ = 0 \quad \text{if } e_L^* > 0. \end{cases} \tag{14.C.5}$$

12. As with the case of hidden actions studied in Section 14.B, nonobservability causes no welfare loss in the case of managerial risk neutrality. As there, a "sellout" contract that faces the manager with the full marginal returns from his actions can generate the first-best outcome. (See Exercise 14.C.2.)

These conditions indicate how the two objectives of insuring the manager and making effort sensitive to the state are handled. First, rearranging and combining conditions (14.C.2) and (14.C.3), we see that

$$v'(w_H^* - g(e_H^*, \theta_H)) = v'(w_L^* - g(e_L^*, \theta_L)), \tag{14.C.6}$$

so the manager's marginal utility of income is equalized across states. This is the usual condition for a risk-neutral party optimally insuring a risk-averse individual. Condition (14.C.6) implies that $w_H^* - g(e_H^*, \theta_H) = w_L^* - g(e_L^*, \theta_L)$, which in turn implies that $v(w_H^* - g(e_H^*, \theta_H)) = v(w_L^* - g(e_L^*, \theta_L))$; that is, the manager's utility is equalized across states. Given the reservation utility constraint in (14.C.1), the manager therefore has utility level $\bar{u}$ in each state.

Now consider the optimal effort levels in the two states. Since $g_e(0, \theta) = 0$ and $\pi'(0) > 0$, conditions (14.C.4) and (14.C.5) must hold with equality and $e_i^* > 0$ for $i = 1, 2$. Combining condition (14.C.2) with (14.C.4), and condition (14.C.3) with (14.C.5), we see that the optimal level of effort in state θ_i, e_i^*, satisfies

$$\pi'(e_i^*) = g_e(e_i^*, \theta_i) \qquad \text{for } i = L, H. \tag{14.C.7}$$

This condition says that the optimal level of effort in state θ_i equates the marginal benefit of effort in terms of increased profit with its marginal disutility cost.

The pair (w_i^*, e_i^*) is illustrated in Figure 14.C.1 (note that the wage is depicted on the vertical axis and the effort level on the horizontal axis). As shown, the manager is better off as we move to the northwest (higher wages and less effort), and the owner is better off as we move toward the southeast. Because the manager receives utility level $\bar{u}$ in state θ_i, the owner seeks to find the most profitable point on the manager's state θ_i indifference curve with utility level $\bar{u}$. This is a point of tangency between the manager's indifference curve and one of the owner's isoprofit curves. At this point, the marginal benefit to additional effort in terms of increased profit is exactly equal to the marginal cost borne by the manager.

The owner's profit level in state θ_i is $\Pi_i^* = \pi(e_i^*) - v^{-1}(\bar{u}) - g(e_i^*, \theta_i)$. As shown in Figure 14.C.1, this profit is exactly equal to the distance from the origin to the point at which the owner's isoprofit curve through point (w_i^*, e_i^*) hits the vertical

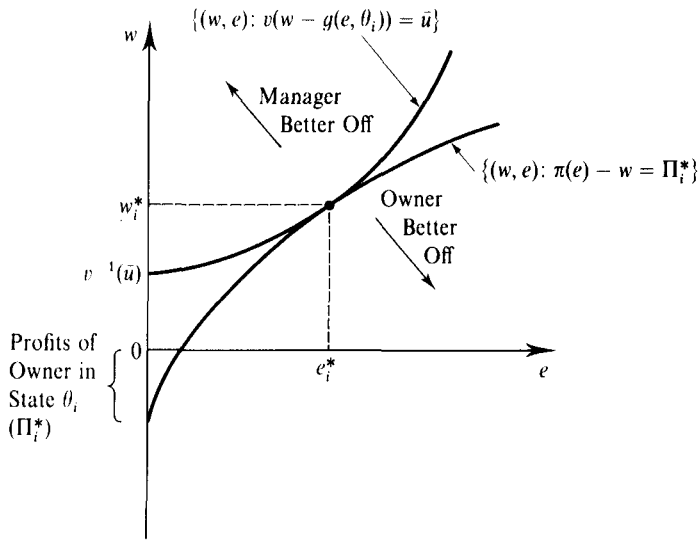


Figure 14.C.1
The optimal wage–effort pair for state θ_i when states are observable.

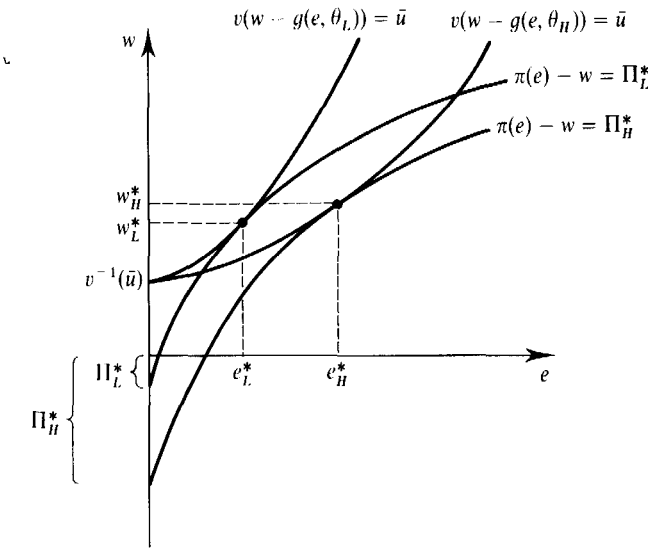


Figure 14.C.2
The optimal contract with full observability of θ.

axis [since $\pi(0) = 0$, if the wage payment at this point on the vertical axis is $\hat{w} < 0$, the owner's profit at (w_i^*, e_i^*) is exactly $-\hat{w}$].

From condition (14.C.7), we see that $g_{e\theta}(e, \theta) < 0$, $\pi''(e) < 0$, and $g_{ee}(e, \theta) > 0$ imply that $e_H^* > e_L^*$. Figure 14.C.2 depicts the optimal contract, $[(w_H^*, e_H^*), (w_L^*, e_L^*)]$.

These observations are summarized in Proposition 14.C.1.

Proposition 14.C.1: In the principal-agent model with an observable state variable θ, the optimal contract involves an effort level e_i^* in state θ_i such that $\pi'(e_i^*) = g_e(e_i^*, \theta_i)$ and fully insures the manager, setting his wage in each state θ_i at the level w_i^* such that $v(w_i^* - g(e_i^*, \theta_i)) = \bar{u}$.

Thus, with a strictly risk-averse manager, the first-best contract is characterized by two basic features: first, the owner fully insures the manager against risk; second, he requires the manager to work to the point at which the marginal benefit of effort exactly equals its marginal cost. Because the marginal cost of effort is lower in state θ_H than in state θ_L, the contract calls for more effort in state θ_H.

The State θ is Observed Only by the Manager

As in Section 14.B, the desire both to insure the risk-averse manager and to elicit the proper levels of effort come into conflict when informational asymmetries are present. Suppose, for example, that the owner offers a risk-averse manager the contract depicted in Figure 14.C.2 and relies on the manager to reveal the state voluntarily. If so, the owner will run into problems. As is evident in the figure, in state θ_H, the manager prefers point (w_L^*, e_L^*) to point (w_H^*, e_H^*). Consequently, in state θ_H he will *lie* to the owner, claiming that it is actually state θ_L. As is also evident in the figure, this misrepresentation lowers the owner's profit.

Given this problem, what is the optimal contract for the owner to offer? To answer this question, it is necessary to start by identifying the set of possible contracts that the owner can offer. One can imagine many different forms that a contract could conceivably take. For example, the owner might offer a compensation function $w(\pi)$ that pays the manager as a function of realized profit and that leaves the effort

choice in each state to the manager's discretion. Alternatively, the owner could offer a compensation schedule $w(\pi)$ but restrict the possible effort choices by the manager to some degree. Another possibility is that the owner could offer compensation as a function of the observable effort level chosen by the manager, possibly again with some restriction on the allowable choices. Finally, more complicated arrangements might be imagined. For example, the manager might be required to make an announcement about what the state is and then be free to choose his effort level while facing a compensation function $w(\pi \mid \hat{\theta})$ that depends on his announcement $\hat{\theta}$.

Although finding an optimal contract from among all these possibilities may seem a daunting task, an important result known as the *revelation principle* greatly simplifies the analysis of these types of contracting problems:[13]

Proposition 14.C.2: (*The Revelation Principle*) Denote the set of possible states by Θ. In searching for an optimal contract, the owner can without loss restrict himself to contracts of the following form:

(i) After the state θ is realized, the manager is required to announce which state has occurred.
(ii) The contract specifies an outcome $[w(\hat{\theta}), e(\hat{\theta})]$ for each possible announcement $\hat{\theta} \in \Theta$.
(iii) In every state $\theta \in \Theta$, the manager finds it optimal to report the state *truthfully*.

A contract that asks the manager to announce the state θ and associates outcomes with the various possible announcements is known as a *revelation mechanism*. The revelation principle tells us that the owner can restrict himself to using a revelation mechanism for which the manager always responds truthfully; revelation mechanisms with this truthfulness property are known as *incentive compatible* (or *truthful*) revelation mechanisms. The revelation principle holds in an extremely wide array of incentive problems. Although we defer its formal (and very general) proof to Chapter 23 (see Sections 23.C and 23.D), its basic idea is relatively straightforward.

For example, imagine that the owner is offering a contract with a compensation schedule $w(\pi)$ that leaves the choice of effort up to the manager. Let the resulting levels of effort in states θ_L and θ_H be e_L and e_H, respectively. We can now show that there is a truthful revelation mechanism that generates exactly the same outcome as this contract. In particular, suppose that the owner uses a revelation mechanism that assigns outcome $[w(\pi(e_L)), e_L]$ if the manager announces that the state is θ_L and outcome $[w(\pi(e_H)), e_H]$ if the manager announces that the state is θ_H. Consider the manager's incentives for truth telling when facing this revelation mechanism. Suppose, first, that the state is θ_L. Under the initial contract with compensation schedule $w(\pi)$, the manager could have achieved outcome $[w(\pi(e_H)), e_H]$ in state θ_L by choosing effort level e_H. Since he instead chose e_L, it must be that in state θ_L outcome $[w(\pi(e_L)), e_L]$ is at least as good for the manager as outcome $[w(\pi(e_H)), e_H]$. Thus, under the proposed revelation mechanism, the manager will find telling the truth to be an optimal response when the state is θ_L. A similar argument applies for state θ_H. We see therefore that this revelation mechanism results in truthful announcements

13. Two early discussions of the revelation principle are Myerson (1979) and Dasgupta, Hammond, and Maskin (1979).

by the manager and yields exactly the same outcome as the initial contract. In fact, a similar argument can be constructed for *any* initial contract (see Chapter 23), and so the owner can restrict his attention without loss to truthful revelation mechanisms.[14]

To simplify the characterization of the optimal contract, we restrict attention from this point on to a specific and extreme case of managerial risk aversion: *infinite* risk aversion. In particular, we take the expected utility of the manager to equal the manager's lowest utility level across the two states. Thus, for the manager to accept the owner's contract, it must be that the manager receives a utility of at least $\bar{u}$ in each state.[15] As above, efficient risk sharing requires that an infinitely risk-averse manager have a utility level equal to $\bar{u}$ in each state. If, for example, his utility is $\bar{u}$ in one state and $u' > \bar{u}$ in the other, then the owner's expected wage payment is larger than necessary for giving the manager an expected utility of $\bar{u}$.

Given this assumption about managerial risk preferences, the revelation principle allows us to write the owner's problem as follows:

$$\underset{w_H, e_H \geq 0, w_L, e_L \geq 0}{\text{Max}} \quad \lambda[\pi(e_H) - w_H] + (1-\lambda)[\pi(e_L) - w_L] \tag{14.C.8}$$

$$\text{s.t.} \quad \left.\begin{array}{ll} \text{(i)} & w_L - g(e_L, \theta_L) \geq v^{-1}(\bar{u}) \\ \text{(ii)} & w_H - g(e_H, \theta_H) \geq v^{-1}(\bar{u}) \end{array}\right\} \begin{array}{l} \textit{reservation utility} \\ \text{(or } \textit{individual rationality}\text{)} \\ \textit{constraint} \end{array}$$

$$\left.\begin{array}{ll} \text{(iii)} & w_H - g(e_H, \theta_H) \geq w_L - g(e_L, \theta_H) \\ \text{(iv)} & w_L - g(e_L, \theta_L) \geq w_H - g(e_H, \theta_L) \end{array}\right\} \begin{array}{l} \textit{incentive compatibility} \\ \text{(or } \textit{truth-telling} \\ \text{or } \textit{self-selection}\text{)} \\ \textit{constraints.} \end{array}$$

The pairs (w_H, e_H) and (w_L, e_L) that the contract specifies are now the wage and effort levels that result from different *announcements* of the state by the manager; that is, the outcome if the manager announces that the state is θ_i is (w_i, e_i). Constraints (i) and (ii) make up the *reservation utility* (or *individual rationality*) *constraint* for the infinitely risk-averse manager; if he is to accept the contract, he must be guaranteed a utility of at least $\bar{u}$ in each state. Hence, we must have $v(w_i - g(e_i, \theta_i)) \geq \bar{u}$ for $i = L, H$ or, equivalently, $w_i - g(e_i, \theta_i) \geq v^{-1}(\bar{u})$ for $i = L, H$. Constraints (iii) and (iv) are the *incentive compatibility* (or *truth-telling* or *self-selection*) *constraints* for the manager in states θ_H and θ_L, respectively. Consider, for example, constraint (iii). The

14. One restriction that we have imposed here for expositional purposes is to limit the outcomes specified following the manager's announcement to being nonstochastic (in fact, much of the literature does so as well). Randomization can sometimes be desirable in these settings because it can aid in satisfying the incentive compatibility constraints that we introduce in problem (14.C.8). See Maskin and Riley (1984a) for an example.

15. This can be thought of as the limiting case in which, starting from the concave utility function $v(x)$, we take the concave transformation $v_\rho(v) = -v(x)^\rho$ for $\rho < 0$ as the manager's Bernoulli utility function and let $\rho \to -\infty$. To see this, note that the manager's expected utility over the random outcome giving $(w_H - g(e_H, \theta_H))$ with probability λ and $(w_L - g(e_L, \theta_L))$ with probability $(1-\lambda)$ is then $EU = -[\lambda v_H^\rho + (1-\lambda)v_L^\rho]$, where $v_i = v(w_i - g(e_i, \theta_i))$ for $i = L, H$. This expected utility is correctly ordered by $(-EU)^{1/\rho} = [\lambda v_H^\rho + (1-\lambda)v_L^\rho]^{1/\rho}$. Now as $\rho \to -\infty$, $[\lambda v_H^\rho + (1-\lambda)v_L^\rho]^{1/\rho} \to \text{Min}\{v_H, v_L\}$ (see Exercise 3.C.6). Hence, a contract gives the manager an expected utility greater than his (certain) reservation utility if and only if $\text{Min}\{v(w_H - g(e_H, \theta_H)), v(w_L - g(e_L, \theta_L))\} \geq \bar{u}$.

manager's utility in state θ_H is $v(w_H - g(e_H, \theta_H))$ if he tells the truth, but it is $v(w_L - g(e_L, \theta_H))$ if he instead claims that it is state θ_L. Thus, he will tell the truth if $w_H - g(e_H, \theta_H) \geq w_L - g(e_L, \theta_H)$. Constraint (iv) follows similarly.

Note that the first-best (full observability) contract depicted in Figure 14.C.2 does not satisfy the constraints of problem (14.C.8) because it violates constraint (iii).

We analyze problem (14.C.8) through a sequence of lemmas. Our arguments for these results make extensive use of graphical analysis to build intuition. An analysis of this problem using Kuhn Tucker conditions is presented in Appendix B.

Lemma 14.C.1: We can ignore constraint (ii). That is, a contract is a solution to problem (14.C.8) if and only if it is the solution to the problem derived from (14.C.8) by dropping constraint (ii).

Proof: Whenever both constraints (i) and (iii) are satisfied, it must be that $w_H - g(e_H, \theta_H) \geq w_L - g(e_L, \theta_H) \geq w_L - g(e_L, \theta_L) \geq v^{-1}(\bar{u})$, and so constraint (ii) is also satisfied. This implies that the set of feasible contracts in the problem derived from (14.C.8) by dropping constraint (ii) is exactly the same as the set of feasible contracts in problem (14.C.8). ■

Lemma 14.C.1 is illustrated in Figure 14.C.3. By constraint (i), (w_L, e_L) must lie in the shaded region of the figure. But by constraint (iii), (w_H, e_H) must lie on or above the state θ_H indifference curve through point (w_L, e_L). As can be seen, this implies that the manager's state θ_H utility is at least $\bar{u}$, the utility he gets at point $(w, e) = (v^{-1}(\bar{u}), 0)$.

Therefore, from this point on we can ignore constraint (ii).

Lemma 14.C.2: An optimal contract in problem (14.C.8) must have $w_L - g(e_L, \theta_L) = v^{-1}(\bar{u})$.

Proof: Suppose not, that is, that there is an optimal solution $[(w_L, e_L), (w_H, e_H)]$ in which $w_L - g(e_L, \theta_L) > v^{-1}(\bar{u})$. Now, consider an alteration to the owner's contract

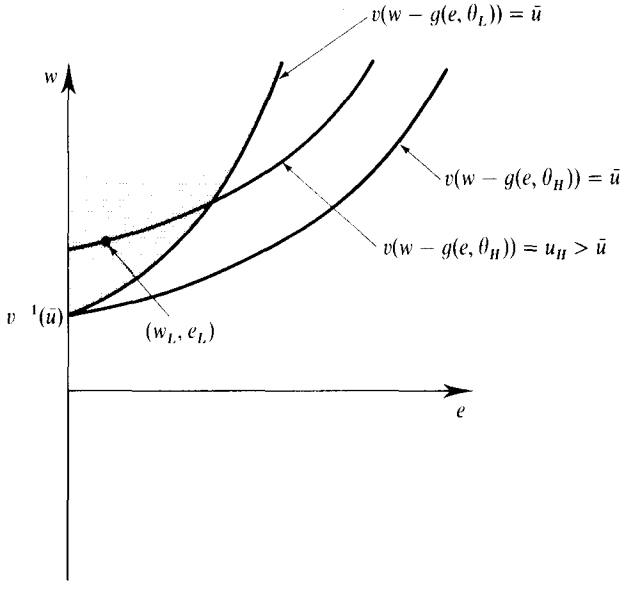
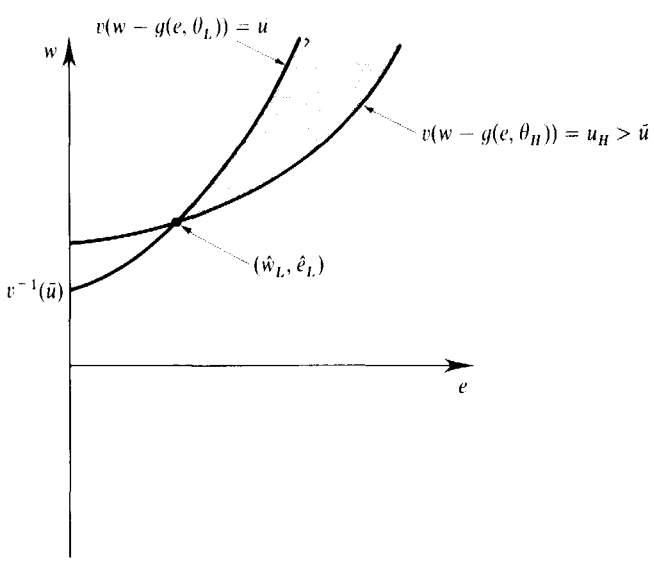


Figure 14.C.3
Constraint (ii) in problem (14.C.8) is satisfied by any contract satisfying constraints (i) and (iii).



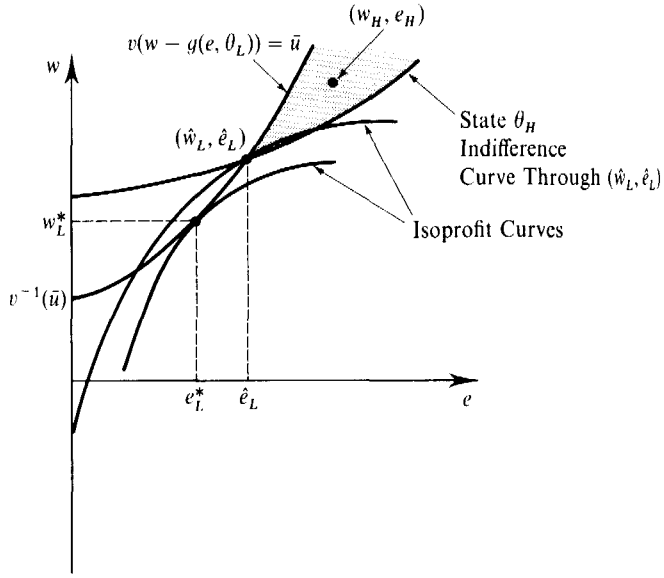


Figure 14.C.4 (left)
In a feasible contract offering $(\hat{w}_L, \hat{e}_L)$ for state θ_L, the pair (w_H, e_H) must lie in the shaded region.

Figure 14.C.5 (right)
An optimal contract has $e_L \leq e_L^*$.

in which the owner pays wages in the two states of $\hat{w}_L = w_L - \varepsilon$ and $\hat{w}_H = w_H - \varepsilon$, where $\varepsilon > 0$ (i.e., the owner lowers the wage payments in both states by ε). This new contract still satisfies constraint (i) as long as ε is chosen small enough. In addition, the incentive compatibility constraints are still satisfied because this change just subtracts a constant, ε, from each side of these constraints. But if this new contract satisfies all the constraints, the original contract could not have been optimal because the owner now has higher profits, which is a contradiction. ■

Lemma 14.C.3: In any optimal contract:

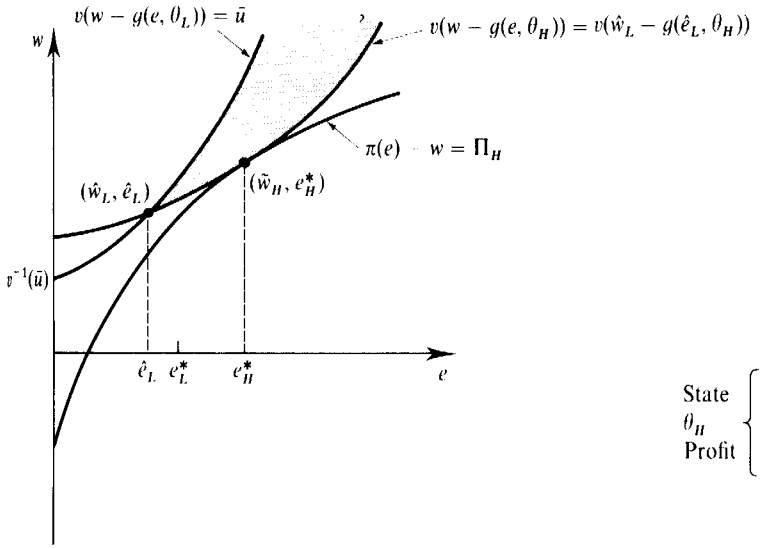
(i) $e_L \leq e_L^*$; that is, the manager's effort level in state θ_L is no more than the level that would arise if θ were observable.

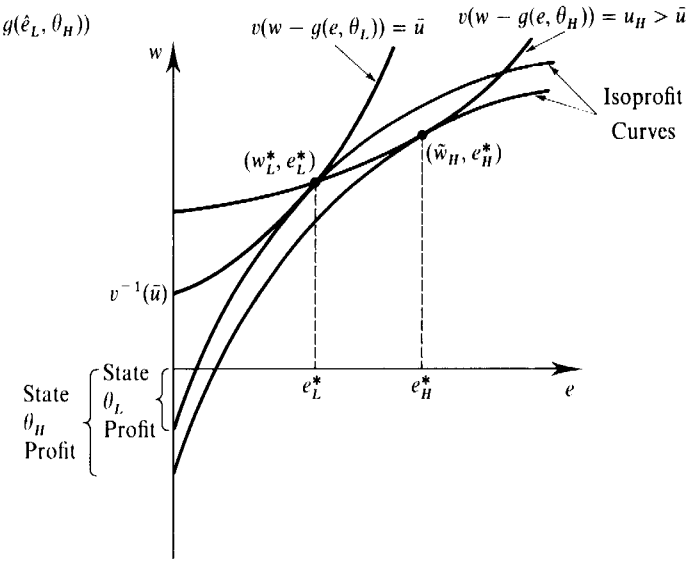
(ii) $e_H = e_H^*$; that is, the manager's effort level in state θ_H is exactly equal to the level that would arise if θ were observable.

Proof: Lemma 14.C.3 can best be seen graphically. By Lemma 14.C.2, (w_L, e_L) lies on the locus $\{(w, e): v(w - g(e, \theta_L)) = \bar{u}\}$ in any optimal contract. Figure 14.C.4 depicts one possible pair $(\hat{w}_L, \hat{e}_L)$. In addition, the truth-telling constraints imply that the outcome for state θ_H, (w_H, e_H), must lie in the shaded region of Figure 14.C.4. To see this, note that by constraint (iv), (w_H, e_H) must lie on or below the state θ_L indifference curve through $(\hat{w}_L, \hat{e}_L)$. In addition, by constraint (iii), (w_H, e_H) must lie on or above the state θ_H indifference curve through $(\hat{w}_L, \hat{e}_L)$.

To see part (i), suppose that we have a contract with $\hat{e}_L > e_L^*$. Figure 14.C.5 depicts such a contract offer: $(\hat{w}_L, \hat{e}_L)$ lies on the manager's state θ_L indifference curve with utility level $\bar{u}$, and (w_H, e_H) lies in the shaded region defined by the truth-telling constraints. The state θ_L indifference curve for the manager and the isoprofit curve for the owner which go through point $(\hat{w}_L, \hat{e}_L)$ have the relation depicted at point $(\hat{w}_L, \hat{e}_L)$ because $\hat{e}_L > e_L^*$.

As can be seen in the figure, the owner can raise his profit level in state θ_L by moving the state θ_L wage–effort pair down the manager's indifference curve from $(\hat{w}_L, \hat{e}_L)$ to its first-best point (w_L^*, e_L^*). This change continues to satisfy all the constraints in problem (14.C.8): The manager's utility in each state is unchanged,





and, as is evident in Figure 14.C.5, the truth-telling constraints are still satisfied. Thus, a contract with $\hat{e}_L > e_L^*$ cannot be optimal.

Now consider part (ii). Given any wage–effort pair $(\hat{w}_L, \hat{e}_L)$ with $\hat{e}_L \leq e_L^*$, such as that shown in Figure 14.C.6, the owner's problem is to find the location for (w_H, e_H) in the shaded region that maximizes his profit in state θ_H. The solution occurs at a point of tangency between the manager's state θ_H indifference curve through point $(\hat{w}_L, \hat{e}_H)$ and an isoprofit curve for the owner. This tangency occurs at point $(\tilde{w}_H, e_H^*)$ in the figure, and necessarily involves effort level e_H^* because all points of tangency between the manager's state θ_H indifference curves and the owner's isoprofit curves occur at effort level e_H^* [they are characterized by condition (14.C.7) for $i = H$]. Note that this point of tangency occurs strictly to the right of effort level $\hat{e}_L$ because $\hat{e}_L \leq e_L^* < e_H^*$. ■

Figure 14.C.6 (left)
An optimal contract has $e_H = e_H^*$.

Figure 14.C.7 (right)
The best contract with $e_L = e_L^*$.

A secondary point emerging from the proof of Lemma 14.C.3 is that only the truth-telling constraint for state θ_H is binding in the optimal contract. This property is common to many of the other applications in the literature.[16]

Lemma 14.C.4: In any optimal contract, $e_L < e_L^*$; that is, the effort level in state θ_L is necessarily *strictly* below the level that would arise in state θ_L if θ were observable.

Proof: Again, this point can be seen graphically. Suppose we start with $(w_L, e_L) = (w_L^*, e_L^*)$, as in Figure 14.C.7. By Lemma 14.C.3, this determines the state θ_H outcome, denoted by $(\tilde{w}_H, e_H^*)$ in the figure. Note that by the definition of (w_L^*, e_L^*), the isoprofit curve through this point is tangent to the manager's state θ_L indifference curve.

Recall that the absolute distance between the origin and the point where each state's isoprofit curve hits the vertical axis represents the profit the owner earns in that state. The owner's overall expected profit with this contract offer is therefore

16. In models with more than two types, this property takes the form that only the incentive constraints between adjacent types bind, and they do so only in one direction. (See Exercise 14.C.1.)

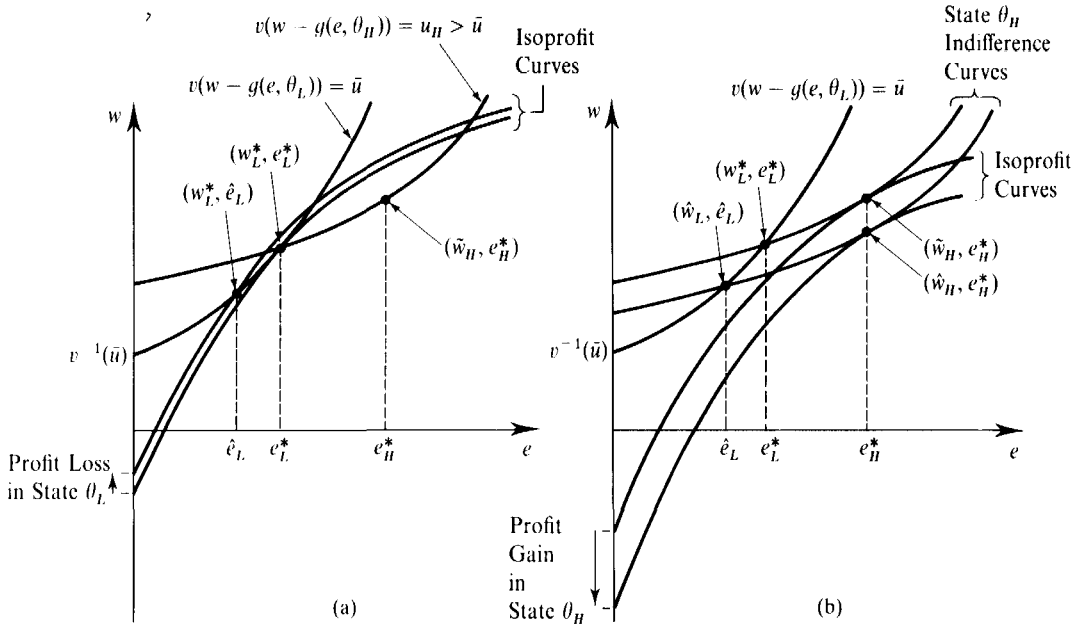


Figure 14.C.8 (a) The change in profits in state θ_L from lowering e_L slightly below e_L^*. (b) The change in profits in state θ_H from lowering e_L slightly below e_L^* and optimally adjusting w_H.

equal to the average of these two profit levels (with weights equal to the relative probabilities of the two states).

We now argue that a change in the state θ_L outcome that lowers this state's effort level to one slightly below e_L^* necessarily raises the owner's expected profit. To see this, start by moving the state θ_L outcome to a slightly lower point, $(\hat{w}_L, \hat{e}_L)$, on the manager's state θ_L indifference curve. This change is illustrated in Figure 14.C.8, along with the owner's isoprofit curve through this new point. As is evident in Figure 14.C.8(a), this change lowers the profit that the owner earns in state θ_L. However, it also relaxes the incentive constraint on the state θ_H outcome and, by doing so, it allows the owner to offer a lower wage in that state. Figure 14.C.8(b) shows the new state θ_H outcome, say $(\hat{w}_H, e_H^*)$, and the new (higher-profit) isoprofit curve through this point.

Overall, this change results in a lower profit for the owner in state θ_L and a higher profit for the owner in state θ_H. Note, however, that because we started at a point of tangency at (w_L^*, e_L^*), the profit loss in state θ_L is small relative to the gain in state θ_H. Indeed, if we were to look at the derivative of the owner's profit in state θ_L with respect to an *infinitesimal* change in that state's outcome, we would find that it is *zero*. In contrast, the derivative of profit in state θ_H with respect to this infinitesimal change would be strictly positive. The zero derivative in state θ_L is an envelope theorem result: because we started out at the first-best level of effort in state θ_L, a small change in (w_L, e_L) that keeps the manager's state θ_L utility at $\bar{u}$ has no first-order effect on the owner's profit in that state; but because it relaxes the state θ_H incentive constraint, for a small-enough change the owner's expected profit is increased. ■

How far should the owner go in lowering e? In answering this question, the owner must weigh the marginal loss in profit in state θ_L against the marginal gain in state

θ_H [note that once we move away from (w_L^*, e_L^*), the envelope result no longer holds and the marginal reduction in state θ_L's profit is strictly positive]. It should not be surprising that the extent to which the owner wants to make this trade-off depends on the relative probabilities of the two states. In particular, the greater the likelihood of state θ_H, the more the owner is willing to distort the state θ_L outcome to increase profit in state θ_H. In the extreme case in which the probability of state θ_L gets close to zero, the owner may set $e_L = 0$ and hire the manager to work only in state θ_H.[17]

The analysis in Appendix B confirms this intuition. There we show that the optimal level of e_L satisfies the following first-order condition:

$$[\pi'(e_L) - g_e(e_L, \theta_L)] + \frac{\lambda}{1-\lambda}[g_e(e_L, \theta_H) - g_e(e_L, \theta_L)] = 0. \qquad (14.C.9)$$

The first term of this expression is zero at $e_L = e_L^*$ and is strictly positive at $e_L < e_L^*$; the second term is always strictly negative. Thus, we must have $e_L < e_L^*$ to satisfy this condition, confirming our finding in Lemma 14.C.4. Differentiating this expression reveals that the optimal level of e_L falls as $\lambda/(1-\lambda)$ rises.

These findings are summarized in Proposition 14.C.3.

Proposition 14.C.3: In the hidden information principal-agent model with an infinitely risk-averse manager the optimal contract sets the level of effort in state θ_H at its first-best (full observability) level e_H^*. The effort level in state θ_L is distorted downward from its first-best level e_L^*. In addition, the manager is inefficiently insured, receiving a utility greater than $\bar{u}$ in state θ_H and a utility equal to $\bar{u}$ in state θ_L. The owner's expected payoff is strictly lower than the expected payoff he receives when θ is observable, while the infinitely risk-averse manager's expected utility is the same as when θ is observable (it equals $\bar{u}$).[18,19]

A basic, and very general, point that emerges from this analysis is that the optimal contract for the owner in this setting of hidden information necessarily *distorts* the effort choice of the manager in order to ameliorate the costs of asymmetric information, which here take the form of the higher expected wage payment that the owner makes because the manager has a utility in state θ_H in excess of $\bar{u}$.

Note that nothing would change if the profit level π were not publicly observable (and so could not be contracted on), since our analysis relied only on the fact that the effort level e was observable. Moreover, in the case in which π is not publicly observable, we can extend the model to allow the relationship between profits and effort to depend on the state; that is, the owner's profits in states θ_L and θ_H given effort level e might be given by the functions $\pi_L(e)$ and $\pi_H(e)$.[20] As long as

17. In fact, this can happen only if $g_e(0, \theta_L) > 0$.

18. Recall that an infinitely risk-averse manager's expected utility is equal to his lowest utility level across the two states.

19. Note, however, that while the outcome here is Pareto inefficient, it is a constrained Pareto optimum in the sense introduced in Section 13.B; the reasons parallel those given in footnote 9 of Section 14.B for the hidden action model (although here it is θ that the authority cannot observe rather than e).

20. The nonobservability of profits is important for this extension because if π could be contracted upon, the manager could be punished for misrepresenting the state by simply comparing the realized profit level with the profit level that should have been realized in the announced state for the specified level of effort.

$\pi'_H(e) \geq \pi'_L(e) > 0$ for all $e \geq 0$, the analysis of this model follows exactly along the lines of the analysis we have just conducted (see Exercise 14.C.5).

As in the case of hidden action models, a number of extensions of this basic hidden information model have been explored in the literature. Some of the most general treatments appear in the context of the "mechanism design" literature associated with social choice theory. A discussion of these models can be found in Chapter 23.

The Monopolistic Screening Model

In Section 13.D, we studied a model of *competitive screening* in which firms try to design their employment contracts in a manner that distinguishes among workers who, at the time of contracting, have different unobservable productivity levels (i.e., there is *precontractual* asymmetric information). The techniques that we have developed in our study of the principal-agent model with hidden information enable us to formulate and solve a model of *monopolistic screening* in which, in contrast with the analysis in Section 13.D, only a single firm offers employment contracts (actually, this might more properly be called a *monopsonistic* screening model because the single firm is on the demand side of the market).

To see this, suppose that, as in the model in Section 13.D, there are two possible types of workers who differ in their productivity. A worker of type θ has utility $u(w, t \mid \theta) = w - g(t, \theta)$ when he receives a wage of w and faces task level t. His reservation utility level is $\bar{u}$. The productivities of the two types of workers are θ_H and θ_L, with $\theta_H > \theta_L > 0$. The fraction of workers of type θ_H is $\lambda \in (0, 1)$. We assume that the firm's profits, which are not publicly observable, are given by the function $\pi_H(t)$ for a type θ_H worker and by $\pi_L(t)$ for a type θ_L worker, and that $\pi'_H(t) \geq \pi'_L(t) > 0$ for all $t \geq 0$ [e.g., as in Exercise 13.D.1, we could have $\pi_i(t) = \theta_i(1 - \mu t)$ for $\mu > 0$].[21]

The firm's problem is to offer a set of contracts that maximizes its profits given worker self-selection among, and behavior within, its offered contracts. Once again, the revelation principle can be invoked to greatly simplify the firm's problem. Here the firm can restrict its attention to offering a menu of wage task pairs $[(w_H, t_H), (w_L, t_L)]$ to solve

$$\max_{w_H, t_H \geq 0,\, w_L, t_L \geq 0} \quad \lambda[\pi_H(t_H) - w_H] + (1 - \lambda)[\pi_L(t_L) - w_L] \tag{14.C.10}$$

$$\begin{aligned} \text{s.t.} \quad & \text{(i)} \quad w_L - g(t_L, \theta_L) \geq \bar{u} \\ & \text{(ii)} \quad w_H - g(t_H, \theta_H) \geq \bar{u} \\ & \text{(iii)} \quad w_H - g(t_H, \theta_H) \geq w_L - g(t_L, \theta_H) \\ & \text{(iv)} \quad w_L - g(t_L, \theta_L) \geq w_H - g(t_H, \theta_L). \end{aligned}$$

This problem has exactly the same structure as (14.C.8) but with the principal's (here the firm's) profit being a function of the state. As noted above, the analysis of this problem follows exactly the same lines as our analysis of problem (14.C.8).

This class of models has seen wide application in the literature (although often with a continuum of types assumed). Maskin and Riley (1984b), for example, apply this model to the study of monopolistic price discrimination. In their model, a consumer of type θ has utility $v(x, \theta) - T$ when he consumes x units of a monopolist's good and makes a total payment of T to the monopolist, and can earn a reservation utility level of $v(0, \theta) = 0$ by not purchasing from the monopolist. The monopolist has a constant unit cost of production equal to $c > 0$

21. The model studied in Section 13.D with $\pi_i(t) = \theta_i$ corresponds to the limiting case where $\mu \to 0$.

and seeks to offer a menu of (x_i, T_i) pairs to maximize its profit. The monopolist's problem then takes the form in (14.C.10) where we take $t_i = x_i$, $w_i = -T_i$, $\bar{u} = 0$, $g(t_i, \theta_i) = -v(x_i, \theta_i)$, and $\pi_i(t_i) = -cx_i$.

Baron and Myerson's (1982) analysis of optimal regulation of a monopolist with unknown costs provides another example. There, a regulated firm faces market demand function $x(p)$ and has unobservable unit costs of θ. The regulator, who seeks to design a regulatory policy that maximizes consumer surplus, faces the monopolist with a choice among a set of pairs (p_i, T_i), where p_i is the allowed retail price and T_i is a transfer payment from the regulator to the firm. The regulated firm is able to shut down if it cannot earn profits of at least zero from any of the regulator's offerings. The regulator's problem then corresponds to (14.C.10) with $t_i = p_i$, $w_i = T_i$, $u = 0$, $g(t_i, \theta_i) = -(p_i - \theta_i)x(p_i)$, and $\pi_i(t_i) = \int_{p_i}^{\infty} x(s)\, ds$.[22]

Exercises 14.C.7 to 14.C.9 ask you to study some examples of monopolistic screening models.

14.D Hidden Actions and Hidden Information: Hybrid Models

Although the hidden action–hidden information dichotomization serves as a useful starting point for understanding principal-agent models, many real-world situations (and some of the literature as well) involve elements of both problems.

To consider an example of such a model, suppose that we augment the simple hidden information model considered in Section 14.C in the following manner: let the level of effort e now be unobservable, and let profits be a stochastic function of effort, described by conditional density function $f(\pi \mid e)$. In essence, what we now have is a hidden action model, but one in which the owner also does not know something about the disutility of the manager (which is captured in the state variable θ).

Formal analysis of this model is beyond the scope of this chapter, but the basic thrust of the revelation principle extends to the analysis of these types of hybrid problems. In particular, as Myerson (1982) shows, the owner can now restrict attention to contracts of the following form:

(i) After the state θ is realized, the manager announces which state has occurred.
(ii) The contract specifies, for each possible announcement $\hat{\theta} \in \Theta$, the effort level $e(\hat{\theta})$ that the manager should take and a compensation scheme $w(\pi \mid \hat{\theta})$.
(iii) In every state θ, the manager is willing to be both *truthful* in stage (i) and *obedient* following stage (ii) [i.e., he finds it optimal to choose effort level $e(\theta)$ in state θ].

This contract can be thought of as a revelation game, but one in which the outcome of the manager's announcement about the state is a hidden action-style contract, that is, a compensation scheme and a "recommended action." The requirement of "obedience" amounts to an incentive constraint that is like that in the hidden action

22. The regulator's objective function can be generalized to allow a weighted average of consumer and producer surplus, with greater weight on consumers. In this case, the function $\pi_i(\cdot)$ will depend on θ_i.

model considered in Section 14.B; the "truthfulness" constraints are generalizations of those considered in our hidden information model. See Myerson (1982) for details.

One special case of this hybrid model deserves particular mention because its analysis reduces to that of the pure hidden information model considered in Section 14.C. In particular, suppose that effort is unobservable but that the relationship between effort and profits is *deterministic*, given by the function $\pi(e)$. In that case, for any particular announcement $\hat{\theta}$, it is possible to induce any wage–effort pair that is desired, say $(\hat{w}, \hat{e})$, by use of a simple "forcing" compensation scheme: Just reward the manager with a wage payment of $\hat{w}$ if profits are $\pi(\hat{e})$, and give him a wage payment of $-\infty$ otherwise. Thus, the combination of the observability of π and the one-to-one relationship between π and e effectively allows the contract to specify e. The analysis of this model is therefore identical to that of the hidden information model considered in Section 14.C, where wage–effort pairs could be specified directly as functions of the manager's announcement.

To see this point in a slightly different way, note first that because of the ability to write forcing contracts, in this model an optimal contract can be thought of as specifying, for each announcement $\hat{\theta}$, a wage–profit pair $(w(\hat{\theta}), \pi(\hat{\theta}))$. Now, for any required profit level π, the effort level necessary to achieve a profit of π is $\tilde{e}$ such that $\pi(\tilde{e}) = \pi$. Let the function $\tilde{e}(\pi)$ describe this effort level. We can now think of the manager as having a disutility function defined directly over the profit level which is given by $\tilde{g}(\pi, \theta) = g(\tilde{e}(\pi), \theta)$. But this model looks just like a model with *observable* effort where the effort variable is π, the disutility function over this effort is $\tilde{g}(\pi, \theta)$, and the profit function is $\tilde{\pi}(\pi) = \pi$. Thus, the analysis of this model is identical to that in a pure hidden information model.

A similar point applies to a closely related hybrid model in which, instead of the manager's disutility of effort, it is the relation between profit and effort that depends on the state. In particular, suppose that the disutility of effort is given by the function $g(e)$ and profits are given by the function $\pi(e, \theta)$, where $\pi_e(\cdot) > 0$, $\pi_{ee}(\cdot) < 0$, $\pi_\theta(\cdot) > 0$, and $\pi_{e\theta}(\cdot) > 0$. Effort is not observable, but profits are. The idea is that the manager knows more than the owner does about the true profit opportunities facing the firm (e.g., the marginal productivity of effort). Again, we can think of a contract as specifying, for each announcement by the manager, a wage–profit pair (implicitly using forcing contracts). In this context, the effort needed to achieve any given level of profit π in state θ is given by some function $\hat{e}(\pi, \theta)$, and the disutility associated with this effort is then $\hat{g}(\pi, \theta) = g(\hat{e}(\pi, \theta))$. But this model is also equivalent to our basic hidden information model with observable effort: just let the effort variable be π, the disutility of this effort be $\hat{g}(\pi, \theta)$, and the profit function be $\hat{\pi}(\pi) = \pi$. Again, our results from Section 14.C apply.

APPENDIX A: MULTIPLE EFFORT LEVELS IN THE HIDDEN ACTION MODEL

In this appendix, we discuss additional issues that arise when the effort choice in the hidden action (moral hazard) model discussed in Section 14.B is more complex than the simple two-effort-choice specification $e \in \{e_L, e_H\}$ analyzed there. Here, we return

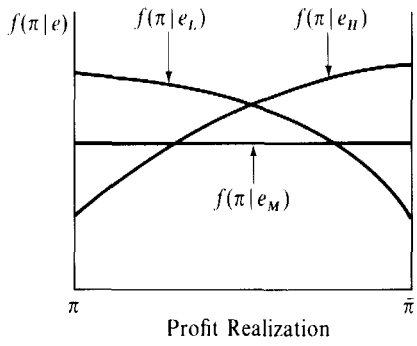


Figure 14.AA.1
Density functions for $E = \{e_L, e_M, e_H\}$: effort choice e_M may not be implementable.

to the more general specification initially introduced in Section 14.B in which E is the feasible set of effort choices.

As in Section 14.B, we can break up the principal's (the owner's) problem into several parts:

(a) What are the effort levels e that it is possible to induce?
(b) What is the optimal contract for inducing each specific effort level $e \in E$?
(c) Which effort level $e \in E$ is optimal?

In a multiple-action setting, each of these three parts becomes somewhat more complicated. For example, with just two actions, part (a) was trivial: e_L could be induced with a fixed wage contract, and e_H could always be induced by giving incentives that were sufficiently high at outcomes that were more likely to arise when e_H is chosen. With more than two actions, however, this may not be so. For example, consider the three-action case in which $E = \{e_L, e_M, e_H\}$ and the conditional density functions are those depicted in Figure 14.AA.1. As is suggested by the figure, it may be impossible to design incentives such that e_M is chosen because for any $w(\pi)$ the agent may prefer either e_L or e_H to e_M. (Exercise 14.B.4 provides an example along these lines.)

Part (b) also becomes more involved. The optimal contract for implementing effort choice e solves

$$\underset{w(\pi)}{\text{Min}} \int w(\pi) f(\pi|e)\, d\pi \tag{14.AA.1}$$

$$\text{s.t. (i)} \int v(w(\pi)) f(\pi|e)\, d\pi - g(e) \geq \bar{u}$$

$$\text{(ii) } e \text{ solves } \underset{\tilde{e} \in E}{\text{Max}} \int v(w(\pi)) f(\pi|\tilde{e})\, d\pi - g(\tilde{e}).$$

If we have K possible actions in set E, the incentive constraints in problem (14.AA.1) [constraints (ii)] consist of $(K-1)$ constraints that must be satisfied. In this case, with a change of variables in which we maximize over the level of utility that the manager gets conditional on π, say $\bar{v}(\pi)$, we have a problem with K linear constraints and a convex objective function [see Grossman and Hart (1983) and footnote 7 for more on this].

However, if E is a continuous set of possible actions, say $E = [0, \bar{e}] \subset \mathbb{R}$, then we have an *infinity* of incentive constraints. One trick sometimes used in this case to

simplify problem (14.AA.1) is to replace constraint (ii) with a *first-order condition* (this is sometimes called the *first-order approach*). For example, if e is a one-dimensional measure of effort, then the manager's first-order condition is

$$\int v(w(\pi))\, f_e(\pi \mid e)\, d\pi - g'(e) = 0, \tag{14.AA.2}$$

where $f_e(\pi \mid e) = \partial f(\pi \mid e)/\partial e$. If we replace constraint (ii) with (14.AA.2) and solve the resulting problem, we can derive a condition for $w(\pi)$ that parallels condition (14.B.10):

$$\frac{1}{v'(w(\pi))} = \lambda + \mu\left[\frac{f_e(\pi \mid e)}{f(\pi \mid e)}\right]. \tag{14.AA.3}$$

The condition that ratio $[f_e(\pi \mid e)/f(\pi \mid e)]$ be increasing in π is the differential version of the monotone likelihood ratio property (see Exercise 14.AA.1).

In general, however, a solution to the problem resulting from this substitution is not necessarily a solution to the actual problem (14.AA.1). The reason is that the agent may satisfy first-order condition (14.AA.2) even when effort level e is not his optimal effort choice. First, effort level e could be a *minimum* rather than a maximum; therefore, we at least want the agent to also be satisfying a local second-order condition. But even this will not be sufficient. In general, we need to be sure that the agent's objective function is concave in e. Note that this is not a simple matter because the concavity of his objective function in e will depend both on the shape of $f(\pi \mid e)$ *and* on the shape of the incentive contract $w(\pi)$ that is offered. The known conditions which insure that this condition is met are very restrictive. See Grossman and Hart (1983) and Rogerson (1985b) for details. Exercise 14.AA.2 provides a very simple example.

Finally, to answer part (c), we need to compute the optimal contract from part (b) for each action that part (a) reveals is implementable and then compare their relative profits for the principal. With more than two effort choices, two features of the two-effort-choice case fail to generalize. First, nonobservability can lead to an upward distortion in effort. (Exercise 14.B.4 provides an example.) Second, at the optimal contract under nonobservability we can get *both* an inefficient effort choice and inefficiencies resulting from managerial risk bearing.

APPENDIX B: A FORMAL SOLUTION OF THE PRINCIPAL-AGENT PROBLEM WITH HIDDEN INFORMATION

Recall problem (14.C.8):

$$\max_{w_H, e_H \geq 0,\, w_L, e_L \geq 0} \quad \lambda[\pi(e_H) - w_H] + (1 - \lambda)[\pi(e_L) - w_L]$$

$$\text{s.t.} \quad \text{(i)} \; w_L - g(e_L, \theta_L) \geq v^{-1}(\bar{u})$$

$$\text{(ii)} \; w_H - g(e_H, \theta_H) \geq v^{-1}(\bar{u})$$

$$\text{(iii)} \; w_H - g(e_H, \theta_H) \geq w_L - g(e_L, \theta_H)$$

$$\text{(iv)} \; w_L - g(e_L, \theta_L) \geq w_H - g(e_H, \theta_L).$$

Using Lemma 14.C.1 we can restate problem (14.C.8) as

$$\max_{w_H, e_H \geq 0,\, w_L, e_L \geq 0} \lambda[\pi(e_H) - w_H] + (1 - \lambda)[\pi(e_L) - w_L] \tag{14.BB.1}$$

$$\text{s.t. (i) } w_L - g(e_L, \theta_L) \geq v^{-1}(\bar{u})$$

$$\text{(iii) } w_H - g(e_H, \theta_H) \geq w_L - g(e_L, \theta_H)$$

$$\text{(iv) } w_L - g(e_L, \theta_L) \geq w_H - g(e_H, \theta_L).$$

Letting $(\gamma, \phi_H, \phi_L) \geq 0$ be the multipliers on constraints (i), (iii), and (iv), respectively, the Kuhn–Tucker conditions for this problem can be written (see Section M.K of the Mathematical Appendix)

$$-\lambda + \phi_H - \phi_L = 0. \tag{14.BB.2}$$

$$-(1 - \lambda) + \gamma - \phi_H + \phi_L = 0. \tag{14.BB.3}$$

$$\lambda\pi'(e_H) - \phi_H g_e(e_H, \theta_H) + \phi_L g_e(e_H, \theta_L) \begin{cases} \leq 0 \\ = 0 \quad \text{if } e_H > 0 \end{cases} \tag{14.BB.4}$$

$$(1 - \lambda)\pi'(e_L) - (\gamma + \phi_L) g_e(e_L, \theta_L) + \phi_H g_e(e_L, \theta_H) \begin{cases} \leq 0 \\ = 0 \quad \text{if } e_L > 0 \end{cases} \tag{14.BB.5}$$

along with the complementary slackness conditions for constraints (i), (iii), and (iv) [conditions (M.K.7)].

Let us break up the analysis of these conditions into several steps.

Step 1: Condition (14.BB.2) implies that $\phi_H > 0$. Thus, constraint (iii) must bind (hold with equality) at an optimal solution.

Step 2: Adding conditions (14.BB.2) and (14.BB.3) implies that $\gamma = 1$. Hence, constraint (i) must bind at an optimal solution.

Step 3: Both e_L and e_H are strictly positive. To see this, note that condition (14.BB.4) cannot hold at $e_H = 0$ because $\pi'(0) > 0$ and $g_e(0, \theta_i) = 0$ for $i = L, H$. Similarly for condition (14.BB.5) and e_L.

Step 4: Steps 1 to 3 imply that $\phi_L = 0$. Suppose not: i.e., that $\phi_L > 0$. Then constraint (iv) must be binding. We shall now derive a contradiction. First, substitute for ϕ_H in conditions (14.BB.4) and (14.BB.5) using the fact that $\phi_H = \phi_L + \lambda$ from condition (14.BB.2). Then, using the fact that $(e_L, e_H) \gg 0$, we can write conditions (14.BB.4) and (14.BB.5) as

$$\lambda[\pi'(e_H) - g_e(e_H, \theta_H)] + \phi_L[g_e(e_H, \theta_L) - g_e(e_H, \theta_H)] = 0$$

and

$$(1 - \lambda)[\pi'(e_L) - g_e(e_L, \theta_H)] + (1 + \phi_L)[g_e(e_L, \theta_H) - g_e(e_L, \theta_L)] = 0.$$

But $\phi_L > 0$ then implies that

$$\pi'(e_L) - g_e(e_L, \theta_H) > 0 > \pi'(e_H) - g_e(e_H, \theta_H),$$

which implies $e_H > e_L$ since $\pi(e) - g(e, \theta_H)$ is concave in e. But if $e_H > e_L$ and constraint (iii) binds (which it does from Step 1), then constraint (iv) must be slack

because we then have

$$\begin{aligned}(w_H - w_L) &= g(e_H, \theta_H) - g(e_L, \theta_H) \\ &= \int_{e_L}^{e_H} g_e(e, \theta_H)\, de \\ &< \int_{e_L}^{e_H} g_e(e, \theta_L)\, de \\ &= g(e_H, \theta_L) - g(e_L, \theta_L).\end{aligned}$$

This is our desired contradiction.

Step 5: Since $\phi_L = 0$, we know from (14.BB.2) that $\phi_H = \lambda$. Substituting these two values into conditions (14.BB.4) and (14.BB.5) we have

$$\pi'(e_H) - g_e(e_H, \theta_H) = 0 \tag{14.BB.6}$$

and

$$[\pi'(e_L) - g_e(e_L, \theta_L)] + \frac{\lambda}{1 - \lambda}[g_e(e_L, \theta_H) - g_e(e_L, \theta_L)] = 0. \tag{14.BB.7}$$

Conditions (14.BB.6) and (14.BB.7) characterize the optimal values of e_H and e_L, respectively. The optimal values for w_L and w_H are then determined from constraints (i) and (iii), which we have seen hold with equality at the solution.

An alternative approach to solving problem (14.BB.1) that avoids this somewhat cumbersome argument involves the following "trick": Solve problem (14.BB.1) ignoring constraint (iv). Then show that the solution derived in this way also satisfies constraint (iv). If so, this must be a solution to the (more constrained) problem (14.BB.1). (Exercise 14.BB.1 asks you to try this approach.)

REFERENCES

Allen, F. (1985). Repeated principal-agent relationships with lending and borrowing. *Economic Letters* **17**: 27–31.

Baron, D., and R. Myerson. (1982). Regulating a monopolist with unknown costs. *Econometrica* **50**: 911–30.

Bernheim, B. D., and M. D. Whinston. (1986). Common agency. *Econometrica* **54**: 923–42.

Dasgupta, P., P. Hammond, and E. Maskin. (1979). The implementation of social choice rules: Some results on incentive compatibility. *Review of Economic Studies* **46**: 185–216.

Dye, R. (1986). Optimal monitoring policies in agencies. *Rand Journal of Economics* **17**: 339–50.

Fudenberg, D., B. Holmstrom, and P. Milgrom. (1990). Short-term contracts and long-term agency relationships. *Journal of Economic Theory* **52**: 194–206.

Green, J., and N. Stokey. (1983). A comparison of tournaments and contests. *Journal of Political Economy* **91**: 349–64.

Grossman, S. J., and O. D. Hart. (1983). An analysis of the principal-agent problem. *Econometrica* **51**: 7–45.

Hart, O. D., and B. Holmstrom. (1987). The theory of contracts. In *Advances in Economic Theory, Fifth World Congress*, edited by T. Bewley. New York: Cambridge University Press.

Holmstrom, B. (1979). Moral hazard and observability. *Bell Journal of Economics* **10**: 74–91.

Holmstrom, B. (1982). Moral hazard in teams. *Bell Journal of Economics* **13**: 324–40.

Holmstrom, B., and P. Milgrom. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica* **55**: 303–28.

Holmstrom, B., and P. Milgrom. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organizations* **7**: 24–52.

Maskin, E., and J. Riley. (1984a). Optimal auctions with risk averse buyers. *Econometrica* **52**: 1473–1518.

Maskin, E., and J. Riley. (1984b). Monopoly with incomplete information. *Rand Journal of Economics* **15**: 171–96.

Milgrom, P. (1981). Good news and bad news: Representation theorems and applications. *Bell Journal of Economics* **12**: 380–91.

Myerson, R. (1979). Incentive compatibility and the bargaining problem. *Econometrica* **47**: 61–74.

Myerson, R. (1982). Optimal coordination mechanisms in generalized principal-agent problems. *Journal of Mathematical Economics* **10**: 67–81.

Nalebuff, B., and J. E. Stiglitz. (1983). Prizes and incentives: Towards a general theory of compensation and competition. *Bell Journal of Economics* **13**: 21–43.

Rogerson, W. (1985a). Repeated moral hazard. *Econometrica* **53**: 69–76.

Rogerson, W. (1985b). The first-order approach to principal-agent problems. *Econometrica* **53**: 1357–68.

EXERCISES

14.B.1[B] Consider the two-effort-level hidden action model discussed in Section 14.B with the general utility function $u(w, e)$ for the agent. Must the reservation utility constraint be binding in an optimal contract?

14.B.2[B] Derive the first-order condition characterizing the optimal compensation scheme for the two-effort-level hidden action model studied in Section 14.B when the principal is strictly risk averse.

14.B.3[B] Consider a hidden action model in which the owner is risk neutral while the manager has preferences defined over the mean and the variance of his income w and his effort level e as follows: Expected utility $= E[w] - \phi \operatorname{Var}(w) - g(e)$, where $g'(0) = 0$, $(g'(e), g''(e), g'''(e)) \gg 0$ for $e > 0$, and $\lim_{e \to \infty} g'(e) = \infty$. Possible effort choices are $e \in \mathbb{R}_+$. Conditional on effort level e, the realization of profit is normally distributed with mean e and variance σ^2.

(a) Restrict attention to linear compensation schemes $w(\pi) = \alpha + \beta\pi$. Show that the manager's expected utility given $w(\pi)$, e, and σ^2 is given by $\alpha + \beta e - \phi\beta^2\sigma^2 - g(e)$.

(b) Derive the optimal contract when e is observable.

(c) Derive the optimal linear compensation scheme when e is not observable. What effects do changes in β and σ^2 have?

14.B.4[B] Consider the following hidden action model with three possible actions $E = \{e_1, e_2, e_3\}$. There are two possible profit outcomes: $\pi_H = 10$ and $\pi_L = 0$. The probabilities of π_H conditional on the three effort levels are $f(\pi_H | e_1) = \frac{2}{3}$, $f(\pi_H | e_2) = \frac{1}{2}$, and $f(\pi_H | e_3) = \frac{1}{3}$. The agent's effort cost function has $g(e_1) = \frac{5}{3}$, $g(e_2) = \frac{8}{5}$, $g(e_3) = \frac{4}{3}$. Finally, $v(w) = \sqrt{w}$, and the manager's reservation utility is $\bar{u} = 0$.

(a) What is the optimal contract when effort is observable?

(b) Show that if effort is not observable, then e_2 is not implementable. For what levels of $g(e_2)$ would e_2 be implementable? [*Hint*: Focus on the utility levels the manager will get for the two outcomes, v_1 and v_2, rather than on the wage payments themselves.]

(c) What is the optimal contract when effort is not observable?

(d) Suppose, instead, that $g(e_1) = \sqrt{8}$, and let $f(\pi_H | e_1) = x \in (0, 1)$. What is the optimal contract if effort is observable as x approaches 1? What is the optimal contract as x approaches 1 if it is not observable? As x approaches 1, is the level of effort implemented higher or lower when effort is not observable than when it is observable?