Radner, R. 1981. Monitoring cooperative agreements in a repeated principal-agent relationship. *Econometrica* 49: 1127–1148.

Radner, R. 1985. Repeated principal agent games with discounting. *Econometrica* 53: 1173–1198.

Radner, R. 1986. Repeated partnership games with imperfect monitoring and no discounting. *Review of Economic Studies* 53: 43–58.

Radner, R., R. Myerson, and E. Maskin. 1986. An example of a repeated partnership game with discounting and with uniformly inefficient equilibria. *Review of Economic Studies* 53: 59–70.

Rosenthal, R. 1979. Sequences of games with varying opponents. *Econometrica* 47: 1353–1366.

Rosenthal, R., and H. Landau. 1979. A game-theoretic analysis of bargaining with reputations. *Journal of Mathematical Psychology* 20: 235–255.

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

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

Rubinstein, A. 1979b. An optimal conviction policy for offenses that may have been committed by accident. In *Applied Game Theory*, ed. S. Brams, A. Schotter, and G. Schwodiauer. Physica-Verlag.

Rubinstein, A., and M. Yaari. 1983. Repeated insurance contracts and moral hazard. *Journal of Economic Theory* 30: 74–97.

Scherer, F. M. 1980. *Industrial Market Structure and Economic Performance*, second edition. Houghton Mifflin.

Shapiro, C. 1982. Consumer information, product quality, and seller reputation. *Bell Journal of Economics* 13: 20–35.

Simon, H. 1951. A formal theory of the employment relationship. *Econometrica* 19: 293–305.

Smith, L. 1989. Folk theorems in overlapping-generations games. Mimeo.

Smith, L. 1990. Folk theorems: Two-dimensionality is (almost) enough. Mimeo.

Sorin, S. 1986. On repeated games with complete information. *Mathematics of Operations Research* 11: 147–160.

Sorin, S. 1988. Supergames. Mimeo.

Stigler, G. 1964. A theory of oligopoly. *Journal of Political Economy* 72: 44–61.

van Damme, E. 1989. Renegotiation-proof equilibria in repeated prisoner's dilemma. *Journal of Economic Theory* 47: 206–207.

III STATIC GAMES OF INCOMPLETE INFORMATION

6 Bayesian Games and Bayesian Equilibrium

6.1 Incomplete Information[†]

When some players do not know the payoffs of the others, the game is said to have *incomplete information.* Many games of interest have incomplete information to at least some extent; the case of perfect knowledge of payoffs is a simplifying assumption that may be a good approximation in some cases.

As a particularly simple example of a game in which incomplete information matters, consider an industry with two firms: an incumbent (player 1) and a potential entrant (player 2). Player 1 decides whether to build a new plant, and simultaneously player 2 decides whether to enter. Imagine that player 2 is uncertain whether player 1's cost of building is 3 or 0, while player 1 knows her own cost. The payoffs are depicted in figure 6.1. Player 2's payoff depends on whether player 1 builds, but is not directly influenced by player 1's cost. Entering is profitable for player 2 if and only if player 1 does not build. Note also that player 1 has a dominant strategy: "build" if her cost is low and "don't build" if her cost is high.

Let p_1 denote the prior probability player 2 assigns to player 1's cost being high. Because player 1 builds if and only if her cost is low, player 2 enters whenever $p_1 > \frac{1}{2}$ and stays out if $p_1 < \frac{1}{2}$. Thus, we can solve the game in figure 6.1 by the iterated deletion of strictly dominated strategies. Section 6.6 gives a careful analysis of iterated dominance arguments in games of incomplete information.

The analysis of the game becomes more complex when the low cost is only 1.5 instead of 0, as in figure 6.2. In this new game, "don't build" is still a dominant strategy for player 1 when her cost is high. However, when her cost is low, player 1's optimal strategy depends on her prediction of y, the probability that player 2 enters: Building is better than not building if

$$1.5y + 3.5(1 - y) > 2y + 3(1 - y),$$

or

$$y < \tfrac{1}{2}.$$

Thus, player 1 must try to predict player 2's behavior to choose her own action, and player 2 cannot infer player 1's action from his knowledge of player 1's payoffs alone.

Harsanyi (1967–68) proposed that the way to model and understand this situation is to introduce a prior move by nature that determines player 1's "type" (here, her cost). In the transformed game, player 2's *incomplete* information about player 1's cost becomes *imperfect* information about nature's moves, so the transformed game can be analyzed with standard techniques.

	Enter	Don't
Build	0,−1	2,0
Don't Build	2,1	3,0

Payoffs if 1's building cost is high

	Enter	Don't
Build	3,−1	5,0
Don't Build	2,1	3,0

Payoffs if 1's building cost is low

Figure 6.1

	Enter	Don't
Build	0,−1	2,0
Don't Build	2,1	3,0

Payoffs if 1's building cost is high

	Enter	Don't
Build	1.5,−1	3.5,0
Don't Build	2,1	3,0

Payoffs if 1's building cost is low

Figure 6.2

The transformation of incomplete information into imperfect information is illustrated in figure 6.3, which depicts Harsanyi's rendering of the game of figure 6.2. N denotes "nature," who chooses player 1's type. (In the figure, numbers in brackets are probabilities of nature's moves.) The figure incorporates the standard assumption that all players have the same prior beliefs about the probability distribution on nature's moves. (Although this is a standard assumption, it may be more plausible when nature's moves represent public events, such as the weather, than when nature's moves model the determination of the players' payoffs and other private characteristics.) Once this common-prior assumption is imposed, we have a standard game, to which Nash equilibrium can be applied. Harsanyi's *Bayesian equilibrium* (or Bayesian Nash equilibrium) is precisely the Nash equilibrium of the imperfect-information representation of the game.

For instance, in the game of figure 6.2 (or figure 6.3), let x denote player 1's probability of building when her cost is low (player 1 never builds when her cost is high), and let y denote player 2's probability of entry. The optimal strategy for player 2 is $y = 1$ (enter) if $x < 1/[2(1 - p_1)]$, $y = 0$ if $x > 1/[2(1 - p_1)]$, and $y \in [0, 1]$ if $x = 1/[2(1 - p_1)]$. Similarly, the best response for the low-cost player 1 is $x = 1$ (build) if $y < \frac{1}{2}$, $x = 0$ if $y > \frac{1}{2}$, and $x \in [0, 1]$ if $y = \frac{1}{2}$. The search for a Bayesian equilibrium boils down to finding a pair (x, y) such that x is optimal for player 1 with low cost against player 2 and y is optimal for player 2 against player 1 given beliefs p_1 and

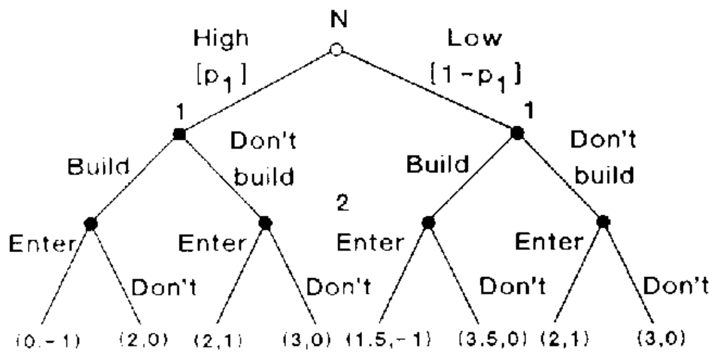


Figure 6.3

player 1's strategy. For instance, $(x = 0, y = 1)$ (player 1 does not build, player 2 enters) is an equilibrium for any p_1, and $(x = 1, y = 0)$ (player 1 builds if her cost is low, and player 2 does not enter) is an equilibrium if and only if $p_1 \leq \frac{1}{2}$.[1]

The remainder of the chapter is organized as follows. Section 6.2 gives a second example of Bayesian equilibrium in a game of incomplete information. Section 6.3 discusses the notion of type, and section 6.4 gives a formal definition of Bayesian equilibrium. Section 6.5 returns to illustrations but emphasizes the details of the characterization of Bayesian equilibria rather than motivation. The details of the analysis are somewhat involved, and many of the examples could be skipped on a first reading. Section 6.6 discusses the iterated deletion of dominated strategies in games of incomplete information. Here the issue arises of whether different "types" of a single player should be viewed as separate individuals, with potentially different beliefs about the strategies of their opponents, or as a single individual with fixed beliefs. Section 6.7 develops an incomplete-information justification of mixed strategies in games of complete information. Section 6.8 presents more technical material on games in which players have a continuum of types.

6.2 Example 6.1: Providing a Public Good under Incomplete Information[†]

The supply of a public good gives rise to the celebrated *free-rider problem*. Each player benefits when the public good is provided, but each would prefer the other players to incur the cost of supplying it. There are numerous variants of the public-good paradigm; we consider one studied experimentally by Palfrey and Rosenthal (1989). There are two players, $i = 1, 2$. Players decide simultaneously whether to contribute to the public good, and contributing is a 0-1 decision. Each player derives a benefit of 1 if at

1. In this case, there is also a mixed-strategy equilibrium: $(x = 1/[2(1-p_1)], y = \frac{1}{2})$.

	Contribute	Don't
Contribute	$1-c_1, 1-c_2$	$1-c_1, 1$
Don't	$1, 1-c_2$	$0,0$

Figure 6.4

least one of them provides the public good and 0 if none does; player i's cost of contributing is c_i. The payoffs are depicted in figure 6.4.[2]

The benefits of the public good—1 each—are common knowledge, but each player's cost is known only to that player. However, both players believe it is common knowledge that the c_i are drawn independently from the same continuous and strictly increasing cumulative distribution function, $P(\cdot)$, on $[\underline{c}, \overline{c}]$, where $\underline{c} < 1 < \overline{c}$ (so $P(\underline{c}) = 0$ and $P(\overline{c}) = 1$). The cost c_i is player i's "type."

A pure strategy in this game is a function $s_i(c_i)$ from $[\underline{c}, \overline{c}]$ into $\{0, 1\}$, where 1 means "contribute" and 0 means "don't contribute." Player i's payoff is

$$u_i(s_i, s_j, c_i) = \max(s_1, s_2) - c_i s_i.$$

(Note that player i's payoff does not depend $c_j, j \neq i$.)

A Bayesian equilibrium is a pair of strategies $(s_1^*(\cdot), s_2^*(\cdot))$ such that, for each player i and every possible value of c_i, strategy $s_i^*(c_i)$ maximizes $E_{c_j} u_i(s_i, s_j^*(c_j), c_i)$. Let $z_j \equiv \text{Prob}(s_j^*(c_j) = 1)$ be the equilibrium probability that player j contributes. To maximize his expected payoff, player i will contribute if his cost c_i is less than $1 \cdot (1 - z_j)$, which is his benefit from the public good times the probability that player j does not contribute. Thus, $s_i^*(c_i) = 1$ if $c_i < 1 - z_j$, and conversely, $s_i^*(c_i) = 0$ if $c_i > 1 - z_j$.[3] This shows that the types of player i who contribute lie in an interval $[\underline{c}, c_i^*]$: Player i contributes only if his cost is sufficiently low. (We adopt the convention that $[\underline{c}, c_i^*]$ is empty if $c_i^* < \underline{c}$.) Similarly, player j contributes if and only if $c_j \in [\underline{c}, c_j^*]$ for some c_j^*. Such "monotonicity" properties are frequent in economic applications; they will be useful in characterizing Bayesian equilibria in section 6.5, and they will be developed in more detail in chapter 7.

Since $z_j = \text{Prob}(\underline{c} \leq c_j \leq c_j^*) = P(c_j^*)$, the equilibrium cutoff levels c_i^* must satisfy $c_i^* = 1 - P(c_j^*)$. Thus, c_1^* and c_2^* must both satisfy the equation

2. The important feature of this game as a model of the provision of public goods is that if both players choose to contribute they both pay the full cost, as opposed to sharing the cost equally. One in interpretation of this model is that the two players belong to a committee. If either player attends the committee's meeting, the outcome is the one both prefer; if neither attends, the outcome is bad for both. The time to attend the meeting costs c_i utils.

3. Type $c_i = 1 - z_j$ is indifferent between contributing and not contributing, but since $P(\cdot)$ is continuous the probability of this (or any) particular type is 0.

$c^* = 1 - P(1 - P(c^*))$. If there is a unique c^* that solves this equation, then necessarily $c_i^* = c^* = 1 - P(c^*)$. For instance, if P is uniform on $[0, 2]$ $(P(c) \equiv c/2)$, then c^* is unique and is equal to $\frac{2}{3}$. (As a check on the analysis, note that if a player does not contribute his expected payoff is $P(c^*) = \frac{1}{3}$, and if a player with cost c^* contributes his payoff is $1 - c^* = \frac{1}{3}$.) A player does not contribute if his cost belongs to $(\frac{2}{3}, 1)$ even though his cost of providing the good is less than his benefit, and even though there is probability $1 - P(c^*) = \frac{2}{3}$ that the good will not be supplied by the other player.

If, instead of $\underline{c} = 0$, we suppose that $\underline{c} \geq 1 - P(1)$, the game has two asymmetric Nash equilibria. In these equilibria, one player never contributes and the other player contributes for all $c \leq 1$. For instance, the equilibrium where player 1 never contributes is $c_1^* = 1 - P(1) < \underline{c}$, and $c_2^* = 1$. The player who never contributes prefers not to, as his minimum cost of $\underline{c}$ exceeds the gain of $1 \cdot (1 - P(1))$ from increased supply of the good; the player who contributes for all $c \leq 1$ is playing optimally in view of the fact that if he does not contribute there is probability 0 of obtaining the public good.

6.3 The Notions of Type and Strategy[+]

In the examples of sections 6.1 and 6.2, a player's "type"—his private information—was simply his cost. More generally, the "type" of a player embodies any private information (more precisely, any information that is not common knowledge to all players) that is relevant to the player's decision making. This may include, in addition to the player's payoff function, his beliefs about other players' payoff functions, his beliefs about what other players believe his beliefs are, and so on.

We have already seen examples where the players' types are identified with their payoff functions. For an example where the type includes more than this, consider disarmament talks between two negotiators. Player 2's objective function is public information; player 1 is uncertain whether player 2 knows player 1's objectives. To model this, suppose that player 1 has two possible types—a "tough" type, who prefers no agreement to making substantial concessions, and a "weak" type, who prefers any agreement to none at all—and that the probability that player 1 is tough is p_1. Furthermore, suppose that player 2 has two types—"informed," who observes player 1's type, and "uninformed," who does not observe player 1's type. The probability that player 2 is informed is p_2, and player 1 does not observe player 2's type.

It is easy to construct more complicated versions of this game where, say, player 1's prior beliefs about player 2 can be either p_2 or p_2', and player 2 does not know which. In practice, though, these sorts of complications make the models difficult to work with, and in most applications a player's beliefs about his opponent are assumed to be completely determined by his own payoff function.

More generally, Harsanyi assumed that the players' types $\{\theta_i\}_{i=1}^{I}$ are drawn from some objective distribution $p(\theta_1, \ldots, \theta_I)$, where θ_i belongs to some space Θ_i. For simplicity, let us assume that Θ_i has a finite number $\#\Theta_i$ of elements. θ_i is observed by player i only. $p(\theta_{-i}|\theta_i)$ denotes player i's conditional probability about his opponent's types $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_I)$ given his type θ_i. We assume that the marginal $p_i(\theta_i)$ on each type $\theta_i \in \Theta_i$ is strictly positive.

To complete the description of a Bayesian game, we must specify a pure-strategy space S_i (with elements s_i, and mixed strategies $\sigma_i \in \Sigma_i$) and a payoff function $u_i(s_1, \ldots, s_I, \theta_1, \ldots, \theta_I)$ for each player i.[4] As in the previous chapters, the usual interpretation is that all the exogenous data of the game—the strategy spaces, payoff functions, possible types, and prior distributions—are "common knowledge" in an informal sense (i.e., every player knows them, knows that everybody knows them, and so on). In other words, any initial private information that a player may have is included in the description of his type.[5]

As usual, these strategy spaces are abstract objects which may be contingent plans in some extensive-form game, but for the time being it may be easiest to think of the strategy spaces S_i as representing choices of (uncontingent) actions. Paralleling our development of concepts in parts I and II, we will begin by discussing the solution concepts of Nash equilibrium and iterated strict dominance, which are often strong enough for reasonable predictions in static games but which are typically too weak for strong predictions in dynamic games. Chapters 8 and 11 develop "equilibrium refinements" for dynamic games of incomplete information.

Since each player's choice of strategy can depend on his type, we let $\sigma_i(\theta_i)$ denote the (possibly mixed) strategy player i chooses when his type is θ_i. If player i knew the strategies $\{\sigma_j(\cdot)\}_{j\neq i}$ of the other players as a function of their type, player i could use his beliefs $p(\theta_{-i}|\theta_i)$ to compute the expected utility to each choice and thus find his optimal response $\sigma_i(\theta_i)$. (Aumann (1964) pointed out that there are technical (measurability) problems with this way of modeling strategies when there is a continuum of types. We will say more about this at the end of this chapter when we discuss the work of Milgrom and Weber (1986).)

4. As in earlier chapters, we allow the payoff functions to be expectations over moves by nature (random variables) not known by any player when the players pick their strategies.
5. For more on this, see Mertens and Zamir 1985 and chapter 3 of Mertens, Sorin, and Zamir 1990.

6.4 Bayesian Equilibrium[†]

Definition 6.1 A *Bayesian equilibrium* in a game of incomplete information with a finite number of types θ_i for each player i, prior distribution p, and pure-strategy spaces S_i is a Nash equilibrium of the "expanded game" in which each player i's space of pure strategies is the set $S_i^{\Theta_i}$ of maps from Θ_i to S_i.[6]

Given a strategy profile $s(\cdot)$, and an $s_i'(\cdot) \in S_i^{\Theta_i}$, let $(s_i'(\cdot), s_{-i}(\cdot))$ denote the profile where player i plays $s_i'(\cdot)$ and the other players follow $s(\cdot)$, and let

$$(s_i'(\theta_i), s_{-i}(\theta_{-i})) = (s_1(\theta_1), \ldots, s_{i-1}(\theta_{i-1}), s_i'(\theta_i), s_{i+1}(\theta_{i+1}), \ldots, s_I(\theta_I))$$

denote the value of this profile at $\theta = (\theta_i, \theta_{-i})$. Then, profile $s(\cdot)$ is a (pure-strategy) Bayesian equilibrium if, for each player i,

$$s_i(\cdot) \in \arg\max_{s_i'(\cdot) \in S_i^{\Theta_i}} \sum_{\theta_i} \sum_{\theta_{-i}} p(\theta_i, \theta_{-i}) u_i(s_i'(\theta_i), s_{-i}(\theta_{-i}), (\theta_i, \theta_{-i})).$$

Because each type has positive probability, this *ex ante* formulation is equivalent to player i maximizing his expected utility conditional on θ_i for each θ_i:

$$s_i(\theta_i) \in \arg\max_{s_i' \in S_i} \sum_{\theta_{-i}} p(\theta_{-i} \mid \theta_i) u_i(s_i', s_{-i}(\theta_{-i}), (\theta_i, \theta_{-i})).$$

The existence of a Bayesian equilibrium is an immediate consequence of the Nash existence theorem. (Since Bayesian equilibrium, like Nash equilibrium, is essentially a consistency check, players' beliefs about others' beliefs do not enter the definition—all that matters is each player's own beliefs about the distribution of types and his opponents' type-contingent strategies. Beliefs about beliefs, and so on, become relevant when one is considering the likelihood that play resembles a Bayesian equilibrium, and when one is considering equilibrium refinements.)

6.5 Further Examples of Bayesian Equilibria[††]

This section sketches the analyses of several Bayesian games. Although the first example is straightforward, the details of the other examples become somewhat involved, and many readers may wish to skip them. However, we refer to several of them in section 6.7.

Example 6.2: Cournot Competition with Incomplete Information
Consider a duopoly playing Cournot (quantity) competition. Let firm i's profit be quadratic: $u_i = q_i(\theta_i - q_i - q_j)$, where θ_i is the difference between

6. The "expanded game" here closely parallels the expanded game used in describing correlated equilibrium in section 2.2.

the intercept of the linear demand curve and firm i's constant unit cost ($i = 1, 2$) and where q_i is the quantity chosen by firm i ($s_i = q_i$). It is common knowledge that, for firm 1, $\theta_1 = 1$ ("firm 2 has complete information about firm 1," or "firm 1 has only one potential type"). Firm 2, however, has private information about its unit cost. Firm 1 believes that $\theta_2 = \frac{3}{4}$ with probability $\frac{1}{2}$ and $\theta_2 = \frac{5}{4}$ with probability $\frac{1}{2}$, and this belief is common knowledge. Thus, firm 2 has two potential types, which we will call the "low-cost type" ($\theta_2 = \frac{5}{4}$) and the "high-cost type" ($\theta_2 = \frac{3}{4}$). The two firms choose their outputs simultaneously.

Let us look for a pure-strategy equilibrium of this game. We denote firm 1's output by q_1, firm 2's output when $\theta_2 = \frac{5}{4}$ by q_2^{L}, and firm 2's output when $\theta_2 = \frac{3}{4}$ by q_2^{H}. Firm 2's equilibrium choice $q_2(\theta_2)$ must satisfy

$$q_2(\theta_2) \in \arg\max_{q_2} \{q_2(\theta_2 - q_1 - q_2)\} \Rightarrow q_2(\theta_2) = (\theta_2 - q_1)/2.$$

Firm 1 does not know which type of firm 2 it faces, so its payoff is the expected value over firm 2's types:

$$q_1 \in \arg\max_{q_1} \{\tfrac{1}{2}q_1(1 - q_1 - q_2^{\mathrm{H}}) + \tfrac{1}{2}q_1(1 - q_1 - q_2^{\mathrm{L}})\}$$

$$\Rightarrow q_1 = \frac{2 - q_2^{\mathrm{H}} - q_2^{\mathrm{L}}}{4}.$$

Plugging in for $q_2(\theta_2)$, we obtain ($q_1 = 1/3, q_2^{\mathrm{L}} = 11/24, q_2^{\mathrm{H}} = 5/24$) as a Bayesian equilibrium. (In fact, this is the unique equilibrium.)

Example 6.3: War of Attrition

Consider an incomplete-information version of the war of attrition discussed in chapter 4. Player i chooses a number s_i in $[0, +\infty)$. Both players choose simultaneously. The payoffs are

$$u_i = \begin{cases} -s_i & \text{if } s_j \geq s_i \\ \theta_i - s_j & \text{if } s_j < s_i. \end{cases}$$

Player i's type, θ_i, is private information, and takes values in $[0, +\infty)$ with cumulative distribution P and density p. Types are independent between the players. θ_i is the prize received by the winner (i.e., the player whose s_i is highest). The game resembles a second-bid auction in that the winner pays the second bid. However, it differs from the second-bid auction in that the loser also pays the second bid.

Let us look for a (pure-strategy) Bayesian equilibrium $(s_1(\cdot), s_2(\cdot))$ of this game. For each θ_i, $s_i(\theta_i)$ must satisfy

$$s_i(\theta_i) \in \arg\max_{s_i} \left\{ -s_i \operatorname{Prob}(s_j(\theta_j) \geq s_i) + \int_{\{\theta_j : s_j(\theta_j) < s_i\}} (\theta_i - s_j(\theta_j)) p_j(\theta_j)\, d\theta_j \right\}. \tag{6.1}$$

We will look for profiles in which each player's strategy is a strictly increasing and continuous function of his type. In fact, it can be shown that every equilibrium profile satisfies these properties. To see that equilibrium strategies must be nondecreasing, notice that equilibrium requires that type θ_i' prefer $s_i' = s_i(\theta_i')$ to $s_i'' = s_i(\theta_i'')$ and that type θ_i'' prefer s_i'' to s_i'. Thus,

$$\theta_i' \operatorname{Prob}(s_j(\theta_j) < s_i') - s_i' \operatorname{Prob}(s_j(\theta_j) \geq s_i') - \int_{\{\theta_j | s_j(\theta_j) < s_i'\}} s_j(\theta_j) p_j(\theta_j)\, d\theta_j$$

$$> \theta_i' \operatorname{Prob}(s_j(\theta_j) < s_i'') - s_i'' \operatorname{Prob}(s_j(\theta_j) \geq s_i'') - \int_{\{\theta_j | s_j(\theta_j) < s_i''\}} s_j(\theta_j) p_j(\theta_j)\, d\theta_j,$$

and

$$\theta_i'' \operatorname{Prob}(s_j(\theta_j) < s_i'') - s_i'' \operatorname{Prob}(s_j(\theta_j) \geq s_i'') - \int_{\{\theta_j | s_j(\theta_j) < s_i''\}} s_j(\theta_j) p_j(\theta_j)\, d\theta_j$$

$$> \theta_i'' \operatorname{Prob}(s_j(\theta_j) < s_i') - s_i' \operatorname{Prob}(s_j(\theta_j) \geq s_i') - \int_{\{\theta_j | s_j(\theta_j) < s_i'\}} s_j(\theta_j) p_j(\theta_j)\, d\theta_j.$$

Subtracting the right-hand side of the second inequality from the left-hand side of the first, and subtracting the left-hand side of the second inequality from the right-hand side of the first, yields

$$(\theta_i'' - \theta_i')[\operatorname{Prob}(s_j(\theta_j) \geq s_i') - \operatorname{Prob}(s_j(\theta_j) \geq s_i'')] \geq 0,$$

so $s_i'' > s_i'$ if $\theta_i'' \geq \theta_i'$. (This is the monotonicity property mentioned in example 6.1.)

The argument that strategies must be *strictly* increasing and continuous is more involved, and we will only give the intuition. First, if strategies were not strictly increasing, there would be an "atom" at some $s > 0$, i.e., an s such that $\operatorname{Prob}(s_j(\theta_j) = s) > 0$. In this case, player i would assign probability 0 to the interval $[s - \varepsilon, s)$ for ε small, as she does better playing just above s (this argument is a bit loose, but can be made rigorous). Thus, the types of player j that play s would be better off playing $s - \varepsilon$, because this would not reduce the probability of winning and would lead to reduced cost, so there cannot be an atom at s after all. A similar intuition underlies the argument that strategies must be continuous. If they were discontinuous, then there would be an $s' \geq 0$ and an $s'' > s'$ such that $\operatorname{Prob}(s_j(\theta_j) \in [s', s'']) = 0$ while $s_j(\hat{\theta}_j) = s'' + \varepsilon$ for some small $\varepsilon \leq 0$ for some $\hat{\theta}_j$. In this case, player i strictly prefers $s_i = s'$ to any $s_i \in (s', s'')$, as the probability of winning is the same and the expected cost is reduced. But then quitting "at or just beyond" s'' is not optimal for player j with type $\hat{\theta}_j$.

Let us look for a strictly increasing, continuous function s_i with inverse Φ_i that is, $\Phi_i(s_i)$ is the type that plays s_i. Transforming the variable of integration from θ_j to s_j in equation 6.1 (using the formula for the transformation of the densities[7]) gives

$$s_i(\theta_i) \in \arg\max_{s_i} \left\{ -s_i(1 - P_j(\Phi_j(s_i))) + \int_0^{s_i} (\theta_i - s_j) p_j(\Phi_j(s_j)) \Phi_j'(s_j)\, ds_j \right\}. \tag{6.2}$$

The corresponding first-order conditions are that type θ_i cannot increase its payoff by playing $s_i + ds_i$ instead of s_i where $s_i \equiv s_i(\theta_i)$. This change costs ds_i if player j plays above $s_i + ds_i$, which has probability $1 - P_j(\Phi_j(s_i + ds_i))$; thus, the expected incremental cost is $(1 - P_j(\Phi_j(s_i)))\, ds_i$ to the first order in ds_i. The change yields a gain of $\theta_i = \Phi_i(s_i)$ if player j plays in the interval $[s_i, s_i + ds_i)$, which occurs if θ_j is the interval $[\Phi_j(s_i), \Phi_j(s_i + ds_i))$; this has probability $p_j(\Phi_j(s_i))\Phi_j'(s_i)\, ds_i$. Equating the costs and benefits, we obtain the first-order conditions[8]

$$\Phi_i(s_i) p_j(\Phi_j(s_i)) \Phi_j'(s_i) = 1 - P_j(\Phi_j(s_i)). \tag{6.3}$$

7. If random variable x has density $p(x)$ and $f: X \to Y$ is one-to-one, then $y = f(x)$ has density g given by

$$g(y) = \frac{p(f^{-1}(y))}{f'(f^{-1}(y))} = p(f^{-1}(y))(f^{-1})'(y).$$

8. Let us show that the global second-order conditions are satisfied if the first-order conditions are. Let $U_i(s_i, \theta_i)$ denote the maximand in equation 6.2. Note that

$$\frac{\partial^2 U_i}{\partial s_i \partial \theta_i} = p_j(\Phi_j(s_i))\Phi_j'(s_i) > 0.$$

Suppose that there exist a type θ_i and a strategy s_i' such that

$$U_i(s_i', \theta_i) > U_i(s_i, \theta_i),$$

where $s_i = s_i(\theta_i)$. This implies that

$$\int_{s_i}^{s_i'} \frac{\partial U_i}{\partial s}(s, \theta_i)\, ds > 0.$$

Or, using the first-order condition $(\partial U_i/\partial s)(s, \Phi_i(s)) = 0$ for all s,

$$\int_{s_i}^{s_i'} \left(\frac{\partial U_i}{\partial s}(s, \theta_i) - \frac{\partial U_i}{\partial s}(s, \Phi_i(s)) \right) ds > 0$$

or

$$\int_{s_i}^{s_i'} \int_{\Phi_i(s)}^{\theta_i} \frac{\partial^2 U_i}{\partial s \partial \theta}(s, \theta)\, d\theta\, ds > 0.$$

If, for instance, $s_i' > s_i$, then $\Phi_i(s) > \theta_i$ for all $s \in (s_i, s_i']$, and the last inequality cannot hold. And similarly for $s_i' < s_i$. So s_i is globally optimal for type θ_i.

Next, we suppose that $P_1 = P_2 = P$, and look for a symmetric equilibrium. Substituting $\theta = \Phi(s)$ in equation 6.3, and using the fact that $\Phi' = 1/s'$,[9] we have

$$s'(\theta) = \frac{\theta\, p(\theta)}{1 - P(\theta)}, \tag{6.4a}$$

or

$$s(\theta) = \int_0^\theta \left(\frac{x\, p(x)}{1 - P(x)} \right) dx, \tag{6.4b}$$

where the constant of integration is determined by $s(0) = 0$: Types with 0 value for the good are unwilling to fight for it.

We leave it to the reader to check that, for a symmetric exponential distribution $P(\theta) = 1 - \exp(-\theta)$, there exists a symmetric equilibrium: $\Phi(s) = \sqrt{2s}$, which corresponds to $s(\theta) = \theta^2/2$. (Riley (1980) shows that there also exists a continuum of asymmetric equilibria: $\Phi_1(s_1) = K\sqrt{s_1}$ and $\Phi_2(s_2) = (2/K)\sqrt{s_2}$ for $K > 0$.)

Aside We can give this war of attrition the following industrial-organization interpretation: Suppose that there are two firms in the market. Each firm loses 1 per unit of time when they compete. They make a monopoly profit when their opponents have left the market, the present discounted value of which is θ_i. (More realistically, we could allow the duopoly and monopoly profit to be correlated, but this would not change the results very much.) Then, s_i is the length of time firm i intends to stay in the market, if firm j has not exited before.[10,11]

Example 6.4: Double Auction

In a double auction, potential sellers and buyers of a single good move simultaneously, with the sellers submitting asking prices and the buyers submitting bids. An auctioneer then chooses a price p that clears the market: All the sellers who ask less than p sell, all the buyers who bid more

9. This is the inverse-function theorem.

10. See chapter 4 for an introduction to the symmetric-information war of attrition. The incomplete-information war of attrition was introduced in the theoretical biology literature by Bishop, Cannings, and Maynard Smith (1978), and extended by Riley (1980), Kreps and Wilson (1982), Nalebuff (1982), Nalebuff and Riley (1983), and Bliss and Nalebuff (1984). For a characterization of the set of equilibria and a uniqueness result with non stationary flow payoffs and/or large uncertainty over types, see Fudenberg and Tirole 1986.

11. Some readers may wonder whether the concept of Nash equilibrium is sufficiently strong for this dynamic interpretation of the game and whether a stronger equilibrium concept might reduce this multiplicity. In our study of the stationary complete-information war of attrition in chapter 4, we saw that all the Nash equilibria are subgame perfect. Similarly, the multiple equilibria just described satisfy the concept of perfect Bayesian equilibrium we introduce in chapter 8. (They trivially satisfy the concept of subgame perfection introduced in chapter 3, as the only proper subgame is the game itself.)

than p buy, and the total number of units supplied at price p equals the number demanded. (Any buyers or sellers who named exactly p are indifferent, and their allocations are chosen so that the quantity demanded equals the quantity supplied.)

Chatterjee and Samuelson (1983) consider the simplest example of a double auction, in which a single seller and a single buyer may trade 0 or 1 unit of a good. The seller (player 1) has cost c, and the buyer (player 2) has valuation v, where v and c belong to the interval $[0, 1]$. The seller and the buyer simultaneously choose bids b_1, $b_2 \in [0, 1]$. If $b_1 \leq b_2$, the two parties trade at price $t = (b_1 + b_2)/2$.[12] If $b_1 > b_2$, the parties do not trade the good and do not transfer money. The seller's utility is thus $u_1 = (b_1 + b_2)/2 - c$ if $b_1 \leq b_2$, and 0 if $b_1 > b_2$; the buyer's utility is $u_2 = v - (b_1 + b_2)/2$ if $b_1 \leq b_2$, and 0 if $b_1 > b_2$.

Under symmetric information (that is, with v and c common knowledge when the two parties bid), this is known as the *Nash* (1953) *demand game*. If we assume $v > c$ to make things interesting, the symmetric-information game has a continuum of pure-strategy, efficient equilibria in which the two parties bid the same amount: $b_1 = b_2 = t \in [c, v]$. In such equilibria the two traders realize positive surplus. If either tries to be more greedy (the seller asks for more than t or the buyer bids less than t), trade does not occur. There are also inefficient equilibria, in which the parties make nonserious offers: The seller asks for more than v and the buyer bids less than c.[13]

Now consider asymmetric information, where the seller's cost is distributed according to distribution P_1 in $[0, 1]$ and the buyer's valuation has distribution P_2 on the same interval. These distributions are common knowledge. Chatterjee and Samuelson look for a pure-strategy equilibrium $(s_1(\cdot), s_2(\cdot))$ where s_1 and s_2 map $[0, 1]$ into $[0, 1]$. Let $F_1(\cdot)$ and $F_2(\cdot)$ denote the equilibrium cumulative distributions of the seller's and the buyer's bids, respectively. That is, $F_1(b)$ is the probability that the seller has a cost that induces him to bid less than b:

$$F_1(b) = \text{Prob}(s_1(c) \leq b).$$

And similarly for the buyer.

For types who trade with positive probability, equilibrium bids are necessarily increasing in type. Consider for instance two costs, c' and c'', for the seller, and let $b_1' \equiv s_1(c')$ and $b_1'' \equiv s_1(c'')$. Then optimization by the seller requires that

$$\int_{b_1'}^{1} \left(\frac{b_1' + b_2}{2} - c'\right) dF_2(b_2) \geq \int_{b_1''}^{1} \left(\frac{b_1'' + b_2}{2} - c'\right) dF_2(b_2)$$

12. Chatterjee and Samuelson thus assume that the parties split the gains from the trade. More generally, the price in a two-player double auction is set at $kb_1 + (1 - k)b_2$, where $k \in [0, 1]$.
13. See exercise 1.3 for Nash's suggestion on how to select among equilibria.

and

$$\int_{b_1''}^{1}\left(\frac{b_1''+b_2}{2}-c''\right)dF_2(b_2)\geq\int_{b_1'}^{1}\left(\frac{b_1'+b_2}{2}-c''\right)dF_2(b_2).$$

Combining these inequalities gives

$$(c''-c')[F_2(b_1'')-F_2(b_1')]\geq 0,$$

so that $b_1''\geq b_1'$ if $c''>c'$.[14] And similarly for the buyer.

Chatterjee and Samuelson require further that each player's strategy as a function of his type be strictly increasing and continuously differentiable. The maximization problem of the type-c seller is then

$$\underset{b_1}{\text{Max}}\int_{b_1}^{1}\left(\frac{b_1+b_2}{2}-c\right)dF_2(b_2),$$

which implies that either

(i) $\frac{1}{2}[1-F_2(s_1(c))]-(s_1(c)-c)f_2(s_1(c))=0$

or

(ii) $\frac{1}{2}[1-F_2(s_1(c))]-(s_1(c)-c)f_2(s_1(c))>0$ and $s_1(c)=1$

or

(iii) $\frac{1}{2}[1-F_2(s_1(c))]-(s_1(c)-c)f_2(s_1(c))<0$ and $s_1(c)=0$.

Since $F_2(1)=1$ and $F_2(0)=0$, the boundary constraints $s_1\in[0,1]$ do not bind and the relevant condition is i. Note that for cost c above the highest buyer bid $\bar{s}_2$ the seller's optimal bid is any $s_1>\bar{s}_2$, and all such bids satisfy the seller's first-order condition, since for these bids both $f_2(s_1(c))$ and $1-F_2(s_1(c))$ equal 0. (Similar remarks apply to the buyer's first-order condition below.) Note that this first-order condition yields the same formula as for a monopoly seller, except that when the seller raises his price by 1 the trading price increases by $\frac{1}{2}$ instead of 1. We have an analogous formula for the buyer:

$$\underset{b_2}{\text{Max}}\int_{0}^{b_2}\left(v-\frac{b_1+b_2}{2}\right)dF_1(b_1)\Rightarrow[v-s_2(v)]f_1(s_2(v))=\tfrac{1}{2}F_1(s_2(v)).$$

Suppose now, following Chatterjee and Samuelson, that P_1 and P_2 are uniform distributions on [0, 1], and look for linear strategies, so that

$$s_1(c)=\alpha_1+\beta_1 c$$

14. To make this conclusion rigorous, we observe that $F_2(b_1'')=F_2(b_1')<1$ is impossible: Type c' of the seller would be better off asking the higher price because this would not affect the probability of trade, and trade would take place at a higher price when it does.

and

$$s_2(v) = \alpha_2 + \beta_2 v.$$

Then

$$F_i(b) = P_i(s_i^{-1}(b)) = s_i^{-1}(b) = (b - \alpha_i)/\beta_i$$

and

$$f_i(b) = 1/\beta_i.$$

Plugging this into the first-order conditions yields

$$2[\alpha_1 + (\beta_1 - 1)c]/\beta_2 = [\beta_2 - (\alpha_1 + \beta_1 c) + \alpha_2]/\beta_2$$

and

$$2[(1 - \beta_2)v - \alpha_2]/\beta_1 = (\alpha_2 + \beta_2 v - \alpha_1)/\beta_1.$$

Since these equations must hold for all c and v, we can identify the constant terms and coefficients of c and v on the two sides of these, obtaining

$$2(\beta_1 - 1) = -\beta_1,$$

$$2(1 - \beta_2) = \beta_2,$$

$$2\alpha_1 = \beta_2 - \alpha_1 + \alpha_2,$$

$$-2\alpha_2 = \alpha_2 - \alpha_1.$$

Solving this system, we have

$$\beta_1 = \beta_2 = \tfrac{2}{3},$$

$$\alpha_1 = \tfrac{1}{4},$$

$$\alpha_2 = \tfrac{1}{12}.$$

With these strategies, player 1's bid of $\frac{1}{4} + \frac{2}{3}c$ is less than his cost if $c > \frac{3}{4}$. However, for costs in this range, $s_1(c)$ also exceeds $\frac{3}{4}$, which is player 2's maximum bid, so player 1's strategy never leads him to sell at a price below his cost. Similarly, player 2's bid exceeds his valuation when $v < \frac{1}{4}$, but again for such bids trade never takes place.

In equilibrium the parties trade if and only if $\alpha_2 + \beta_2 v \geq \alpha_1 + \beta_1 c$, or $v \geq c + \frac{1}{4}$. Comparing with the *ex post* efficient trading pattern (trade if and only if $v \geq c$), we conclude that there is too little trading in equilibrium.

As one would expect from the symmetric-information case, there are other equilibria in this double auction. In particular, both parties making nonserious offers ($b_1 = 1$ and $b_2 = 0$) is an equilibrium. There also exists a continuum of "single-price" equilibria at $b \in [0, 1]$. The seller asks b if $c \leq b$ and 1 if $c > b$, and the buyer offers b if $v \geq b$ and 0 if $v < b$. Because the price is "fixed" at b if trade takes place, no player has any incentive to deviate. More interestingly, Leininger, Linhart, and Radner (1989) show that there exists a one-parameter continuum of differentiable and symmetric (but nonlinear) equilibrium strategies. (There exists a two-parameter continuum of differentiable, asymmetric equilibrium strategies; see Satterthwaite and Williams 1989.) Leininger et al. also show existence of other, discontinuous equilibria.

Example 6.5: First-Price Auction with a Continuum of Types (technical)
In a first-price auction, the bidder who offers the highest price gets the good and pays his bid (in contrast with the second-price auction analyzed in subsection 1.1.3, in which the highest bidder pays the second highest bid); the other bidders do not pay anything. In this example, we study the equilibria of two-bidder, symmetric-uncertainty first-price auctions when the valuations belong to an interval; the next example considers the same game when each valuation belongs to a two-point set. The point of going through two examples of first-price auctions is to illustrate the different techniques used to solve the continuous and discrete cases. (The analysis of the first example is fairly complicated.) There are two bidders, $i = 1, 2$, and one unit of a good for sale. Player i's valuation is θ_i and belongs to $[\underline{\theta}, \bar{\theta}]$, where $\underline{\theta} \geq 0$. Each player knows his own valuation and has beliefs P with positive density p on $[\underline{\theta}, \bar{\theta}]$ about his rival's valuation. The valuations are independent. The seller imposes a reservation price $s_0 > \underline{\theta}$, meaning that bids below s_0 are rejected. Player i's bid is s_i. The utility of player i is $u_i = \theta_i - s_i$ if $s_i > s_j$ and $s_i \geq s_0$; it is $u_i = 0$ if $s_i < s_j$ or $s_i < s_0$. If both players bid the same amount, we assume that each gets the good with probability $\frac{1}{2}$: If $s_i = s_j \geq s_0$, $u_i = (\theta_i - s_i)/2$. Let $s_i(\cdot)$ denote the (pure) equilibrium strategy of player i. (We leave it to the reader to show that s_i is increasing in θ_i, by following the steps of the monotonicity proofs in examples 6.3 and 6.4.)

Bayesian equilibrium strategies can be characterized intuitively as follows.[15] First, note that a player with valuation less than s_0 does not bid (or, rather, bids less than s_0). Second, as in the war of attrition, show that the strategies have no atoms at bids greater than s_0. Next, argue that the strategies have no "gaps." Suppose that player i, whatever his type, does not bid in the interval $[s_i^-, s_i^+]$, where $s_i^- \geq s_0$, but there are types of player i who bid s_i^+ or arbitrarily close to s_i^+. Then player j, whatever his type, ought not to bid $s_j \in (s_i^-, s_i^+)$: Starting from any such s_j, if player j reduces his bid slightly, he does not affect his probability of winning, and he reduces the price he pays when he wins. But then a type of player i who bids s_i^+ or

15. The (rigorous) characterization is given in Maskin and Riley 1986a. The style of proof that the distribution of bids has no atoms and that the strategies are strictly increasing is also common in search theory (e.g., Butters 1977) and in wars of attrition (e.g., Fudenberg and Tirole 1986).

arbitrarily close to s_i^+ would be better off bidding just above s_i^-, as he would reduce his probability of winning by an infinitesimal amount (recall that player i has no atom at s_i^+) and would substantially reduce his payment when winning.

In this manner one can show that the strategies are continuous and strictly increasing beyond s_0. It is easy to see that $s_i(\bar{\theta}) = s_j(\bar{\theta}) \equiv \bar{s}$. (If $s_i(\bar{\theta}) > s_j(\bar{\theta})$, then type $\bar{\theta}$ of player i could lower his bid slightly and still win the auction with probability 1.) Let $\theta_i = \Phi_i(s)$ denote the inverse function of $s_i(\cdot)$ on $(s_0, \bar{s}]$. That is, player i bids s when his valuation is $\Phi_i(s)$. The function $\Phi_i(\cdot)$ is differentiable almost everywhere because it is monotonic.

Type θ_i maximizes $(\theta_i - s)P(\Phi_j(s))$ over s. This yields

$$P(\Phi_j(s)) = [\Phi_i(s) - s]p(\Phi_j(s))\Phi_j'(s). \tag{6.5}$$

Equation (6.5) and the symmetric equation obtained by switching i and j yield two first-order differential equations in the functions $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$. Let $G_j(\cdot)$ denote the cumulative distribution of bids, $G_j(s) = P(\Phi_j(s))$, with density $g_j(s) = p(\Phi_j(s))\Phi_j'(s)$. Equation 6.5 can then be rewritten as

$$G_j(s) = [\Phi_i(s) - s]g_j(s). \tag{6.6}$$

Note the analogy with monopoly pricing: A unit increase in price raises revenue by the expected probability of winning, $G_j(s)$, but the bidder's surplus $(\Phi_i(s) - s)$ is lost with probability $g_j(s)$.

We now investigate the boundary conditions for equation 6.5. Recall that $\Phi_i(\bar{s}) = \bar{\theta}$ for all i. Further, $\lim_{s \downarrow s_0} \Phi_i(s) = s_0$ for at least some i. (Suppose both players have atoms at s_0, i.e., that types $\theta_i \in [s_0, s_0 + a_i]$, $a_i > 0$, bid s_0 for $i = 1, 2$. Then type $s_0 + a_i$ of player i could bid slightly more than s_0 and increase its probability of winning by a nonnegligible amount.) So these two boundary conditions might seem to pin down a unique solution to equation 6.5.

Although the solution is indeed unique, the reasoning is a bit more complex than this, because Φ_j' in equation 6.5 is not Lipschitz continuous at s_0 if $\Phi_i(s_0) = s_0$.[16] Integrating equation 6.5 yields

$$\ln \frac{P(\Phi_2(s))}{P(\Phi_1(s))} = \int_s^{\bar{s}} \left(\frac{1}{\Phi_2(x) - x} - \frac{1}{\Phi_1(x) - x} \right) dx. \tag{6.7}$$

Equation 6.5 shows that if $\Phi_1(s) = \Phi_2(s)$ for some $s \in (s_0, \bar{s}]$, then the solution is symmetric: $\Phi_1(s) = \Phi_2(s)$ for all $s \in (s_0, \bar{s}]$ (and, by continuity, for $s = s_0$ as well). Can there exist an asymmetric solution? By the previous reasoning, $\Phi_1(s) \neq \Phi_2(s)$ for all s in $(s_0, \bar{s}]$. Suppose that, without loss of

16. That is, the slope of Φ_j goes to infinity. Standard results on the uniqueness of solutions to differential equations require Lipschitz continuity. The war of attrition of example 6.3 is not Lipschitz continuous at $s = 0$, which is why it is possible for the system represented in equation 6.3 to have multiple solutions.

generality, $\Phi_2(s) > \Phi_1(s)$ for all s in $(s_0, \bar{s}]$. Then equation 6.7 implies that $P(\Phi_2(s))/P(\Phi_1(s))$ is greater than 1 and increases from s to $\bar{s}$. Hence it cannot converge to 1 at $\bar{s}$, a contradiction.

We conclude that any equilibrium is symmetric, which implies that there is no atom at s_0. From equation 6.5, $\Phi_1 = \Phi_2 = \Phi$ satisfies

$$\ln(P(\Phi(s))) = -\int_s^{\bar{s}} \frac{dx}{\Phi(x) - x}. \tag{6.8}$$

To show that there exists a unique equilibrium it suffices to note that there exists a unique $\bar{s}$ such that, if $\Phi(\cdot)$ is given by equation 6.8, then $\Phi(s_0) = s_0$.[17]

We thus conclude that as long as $s_0 > \underline{\theta}$ there is a unique solution; it is symmetric and satisfies $P(\Phi(s)) = [\Phi(s) - s]p(\Phi(s))\Phi'(s)$ and $\Phi(s_0) = s_0$. The equilibrium strategy $s(\cdot)$ is the inverse of the function $\Phi(\cdot)$.

Example 6.6: First-Price Auction with Two Types

As a last example, we analyze equilibrium in the first-price auction (see example 6.5) when each of two bidders has one of two possible valuations, $\underline{\theta}$ and $\bar{\theta}$ (with $\underline{\theta} < \bar{\theta}$). The valuations are independent; let $\bar{p}$ and $\underline{p}$ denote the probability that θ_i equals $\bar{\theta}$ and $\underline{\theta}$, respectively (with $\underline{p} + \bar{p} = 1$). To make things interesting, assume that the seller's reservation price or minimum bid is lower than $\underline{\theta}$. The new technical twist when the support of the distribution of types is discrete rather than continuous is that players must play a mixed strategy in equilibrium.

We look for an equilibrium where, for each player, type $\underline{\theta}$ bids $\underline{\theta}$ and type $\bar{\theta}$ randomizes according to the continuous distribution $F(s)$ on $[\underline{s}, \bar{s}]$. (It can be shown that the equilibrium is unique.) Clearly, $\underline{s} = \underline{\theta}$: If $\underline{s} > \underline{\theta}$, then a player with valuation $\bar{\theta}$ would be better off bidding just above $\underline{\theta}$ rather than bidding $\underline{s}$ (or close to $\underline{s}$), as this would not reduce his probability of winning and would reduce his payment when he wins. In order for player i with type $\bar{\theta}$ to play a mixed strategy with support $[\underline{s}, \bar{s}]$, it must be the case that

$$\forall s \in [\underline{s}, \bar{s}],\ (\bar{\theta} - s)[\underline{p} + \bar{p}F(s)] = \text{constant}. \tag{6.9}$$

(Type $\bar{\theta}$'s expected payoff is not affected by bids he makes with probability 0. Thus, even though playing $\underline{s}$ with positive probability will result in a lower expected payoff because it risks tying with type $\underline{\theta}$, bid $\underline{s}$ can still belong to the support of type $\bar{\theta}$'s equilibrium strategy.[18]) Because $F(\underline{\theta}) = 0$, the constant is equal to $(\bar{\theta} - \underline{\theta})\underline{p}$. Thus, $F(\cdot)$ is defined by

17. The proof is similar to that proving that there cannot be asymmetric equilibria: Consider two highest bids, $\bar{s}^1 > \bar{s}^2$, and let Φ^1 and Φ^2 denote the corresponding solutions. Then $\Phi^2(\bar{s}^2) - \theta > \Phi^1(\bar{s}^2)$. For any $s \leq s^2$, $P(\Phi^2(s))/P(\Phi^1(s))$ is greater than 1 and is decreasing. Hence, $P(\Phi^2(s))/P(\Phi^1(s))$ cannot converge to 1 when s converges to s_0.

18. Recall that the support of a probability distribution is the smallest closed set that has probability 1.

$$(\theta - s)[\underline{p} + \bar{p}F(s)] = (\bar{\theta} - \theta)\underline{p}. \tag{6.10}$$

Letting $G(s) \equiv \underline{p} + \bar{p}F(s)$ denote the cumulative distribution of bids for $s > \theta$, we can rewrite equation 6.10 as

$$(\theta - s)G(s) = (\bar{\theta} - \underline{\theta})\,p. \tag{6.11}$$

Last, $F(\bar{s}) = 1$ implies that

$$(\theta - \bar{s}) = (\bar{\theta} - \underline{\theta})\underline{p}, \text{ or } \bar{s} = \bar{p}\theta + \underline{p}\underline{\theta}. \tag{6.12}$$

Since the seller's reservation price is below $\underline{\theta}$, trade always takes place, and the seller's expected profit is equal to the expected social surplus minus the expected utility of the bidders. Expected social surplus is equal to $\underline{p}^2\theta + (1 - \underline{p}^2)\bar{\theta}$. Each bidder's net utility is 0 when he has type $\underline{\theta}$ and $\underline{p}(\bar{\theta} - \underline{\theta})$ when he has type $\bar{\theta}$. (Because type $\bar{\theta}$ is indifferent among bids in $(\theta, \bar{s}]$, his utility can be computed by assuming he bids just above $\underline{\theta}$, in which case he wins with probability $\underline{p}$.)

It is interesting to note that both expected social surplus and the bidders' utility (and therefore the seller's expected profit) are the same as in the second-price auction studied in chapter 1. This fact, known as the *revenue-equivalence theorem*, would also hold in the continuous case of example 6.5. (We will see in chapter 7 that the first-price and second-price auctions do not maximize the seller's expected revenue in the two-type case; they do maximize revenue under some conditions in the continuum case.)

6.6 Deletion of Strictly Dominated Strategies††

6.6.1 Interim vs. *Ex Ante* Dominance

If player i, instead of knowing the type-contingent strategies of his opponents, must try to predict them, then player i must be concerned with how player $j \neq i$ *thinks* player i would play for each possible type player i might have. And player i must also try to estimate player j's beliefs about player i's type, in order to predict the distribution of strategies that player i expects to face.

This brings us to the question of how the players predict their opponents' strategies, which in turn raises the following question: Should different types θ_1 and θ_1' of player 1 be viewed simply as a way of describing different information sets of a *single* player 1, who makes a type-contingent decision at the *ex ante* stage (that is, before he learns his type)? This interpretation seems natural in the Harsanyi formulation, which introduces a move by nature that determines the "type" of a single player 1. Alternatively, should we think of θ_1 and θ_1' as denoting two different "individuals," one of whom is selected by nature to "appear" when the game is played? In the first interpretation, the single *ex ante* player 1 should be thought of as predicting his opponents' play at the *ex ante* stage, so all types of player 1 would make the same prediction about the play of the other players. Under the second interpretation, the "different individuals" corresponding to different θ_1's would each make their predictions at the "interim" stage (i.e., after learning their type), and the different types could make different predictions. (This second interpretation may become more plausible if we imagine that the "types" correspond to aspects of preferences that are genetically determined, for here the "*ex ante*" stage is difficult to interpret literally.)

It is interesting to see that iterated strict dominance is at least as strong in the *ex ante* interpretation as in the interim interpretation and that the *ex ante* interpretation yields strictly stronger predictions in some games. To illustrate this, let us return to the public-good game of example 6.1. Using the interim approach to dominance, we ask which strategies are strictly dominated for player i when his cost is c_i. Not contributing is not dominated for any positive cost level, as it is always better not to contribute if you expect that the opponent will contribute. However, if c_i is greater than the private benefit of the good, which is 1, then contributing is strictly dominated for player i.

If the lowest possible cost, $\underline{c}$, is greater than $1 - P(1)$, the deletion process stops after only one round: For all types in $[\underline{c}, 1]$, neither "contribute" nor "don't contribute" is dominated. In particular, interim dominance does not preclude the situation where, for some c' between $\underline{c}$ and 1, all types in an interval $[c, c']$ don't contribute and all types in $(c', 1]$ do contribute—the types in $[c, c']$ will not contribute if they expect that their opponent will contribute whenever his cost is less than 1, while the types in $(c', 1]$ should contribute if they expect that no type of their opponent will contribute.

This situation could not arise in a Bayesian equilibrium, because, as we saw, in any Bayesian equilibrium each player's strategy must be a cutoff rule of the form "contribute if and only if $c_i \leq c'$ for some c'." That is, in Bayesian equilibrium, if the type of player i with a given cost level contributes, all player i's types with lower costs must contribute as well.

The conclusion that players' strategies must take the form of cutoff rules also follows from applying strict dominance at the *ex ante* stage. To see this, note that any strategy $s_i(\cdot)$ for player i that has player i contribute with probability $z > 0$ and is *not* a cutoff rule is strictly dominated *ex ante* by the strategy that has player i contribute if and only if $c_i < c'$, where c' is defined by $P(c') = z$. With this cutoff rule, for any strategy $s_j(\cdot)$ of the opponent, player i receives the public good with the same probability as when using s_i, and player i's expected cost of contributing is strictly lower. The point is that if player i is a single individual optimizing against the play of player j, then any beliefs about j's strategy that make it attractive to contribute with cost c' also make it attractive to contribute for all lower costs.

More generally, the reason that more strategies are dominated *ex ante* than *ex post* is that, for a given type-contingent strategy $\hat{\sigma}_1(\cdot)$ of player 1, it is easier to find a $\sigma_1(\cdot)$ satisfying the *ex ante* dominance condition

$$\sum_{\theta_1} p_1(\theta_1) \sum_{\theta_{-1}} p(\theta_{-1}|\theta_1)u_1(\sigma_1(\theta_1), \sigma_{-1}(\theta_{-1}), \theta)$$
$$> \sum_{\theta_1} p_1(\theta_1) \sum_{\theta_{-1}} p(\theta_{-1}|\theta_1)u_1(\hat{\sigma}_1(\theta_1), \sigma_{-1}(\theta_{-1}), \theta)$$

for all $\sigma_{-1}(\cdot)$ than it is to find a s_1 and a θ_1 that satisfy the interim constraints

$$\sum_{\theta_{-1}} p(\theta_{-1}|\theta_1)u_1(s_1, \sigma_{-1}(\theta_{-1}), \theta) > \sum_{\theta_{-1}} p(\theta_{-1}|\theta_1)u_1(\hat{\sigma}_1(\theta_1), \sigma_{-1}(\theta_{-1}), \theta)$$

for all $\sigma_{-1}(\cdot)$. (One way of putting this is that the *ex ante* approach "pools the domination constraints" and allows the use of "wasted slack" on some constraints.) This difference does not arise when the Nash concept is used, as Nash equilibrium supposes that all players make the same predictions about the strategies that will be played, whereas dominance arguments allow two players to make different predictions about the play of a third.

6.6.2 Examples of Iterated Strict Dominance

Now we present two examples of incomplete-information games where iterated dominance does lead to a unique prediction.

The first is the public-good game of example 6.1 when $\underline{c} < 1 - P(1)$ and there exists a unique c^* such that $c^* = 1 - P(1 - P(c^*))$. Here even interim iterated dominance gives a unique prediction.

Recall that at the first round of iteration we concluded that no type with cost over 1 would contribute. (Contributing is strictly dominated for all $c_i \in (c^1, \bar{c}]$, where $c^1 \equiv 1$.) At the second round, not contributing is strictly dominated for all $c_i \in [\underline{c}, c^2)$, where $c^2 \equiv 1 - P(1) = 1 - P(c^1)$. In contrast, the optimal strategy for types $c_i \in [c^2, c^1]$ depends on what types $c_j \in [c^2, c^1]$ do; hence, no strategy for these types can be eliminated in the second round. In the third round, types close to 1 should not contribute, as the cost of contributing is close to the private value of the public good, and there is a probability of at least $P(c^2)$ that the other player contributes. Thus, if $c_i > c^3 \equiv 1 - P(c^2)$, contributing is a strictly dominated strategy for player i, and so on. Iterating the process of deletion of strictly dominated strategies yields, at stage $2k + 1$ ($k = 0, 1, \ldots$), that contributing is a strictly dominated strategy for types greater than $c^{2k+1} \equiv 1 - P(c^{2k})$. And at stage $2k$ ($k = 1, 2, \ldots$), not contributing is a strictly dominated strategy for types lower than $c^{2k} \equiv 1 - P(c^{2k-1})$. The sequences $\{c^{2k+1}\}_{k=0,1,\ldots}$ and $\{c^{2k}\}_{k=1,2,\ldots}$ are strictly decreasing and strictly increasing, respectively. Because they are bounded, they converge to two numbers c^+ and c^-. Because P is continuous, $c^+ = 1 - P(c^-)$ and $c^- = 1 - P(c^+)$. If there is a unique c^* such that $c^* = 1 - P(1 - P(c^*))$, which is the condition for a

	L	R
U	10,12	10,0
D	0,0	12,10

$\theta_1 = \theta_1'$

	L	R
U	12,0	0,10
D	10,12	10,0

$\theta_1 = \theta_1''$

a

	L	R
UU	11,6	5,5
UD	10,12	10,0
DU	6,0	6,10
DD	5,6	11,5

b

Figure 6.5

unique Nash equilibrium, then $c^+ = c^- = c^*$ and the game is *solvable by (interim) iterated deletion of strictly dominated strategies.*

In our second example, *ex ante* iterated dominance gives a unique prediction, but interim iterated dominance does not.

Consider the game illustrated in figure 6.5. Player 1 has two possible types, θ_1' and θ_1'', each of which has prior probability $\frac{1}{2}$. Figure 6.5a displays the payoff matrices corresponding to player 1's two types; figure 6.5b shows the strategic form for the imperfect-information game where player 1 chooses type-contingent strategies. Here the first component of player 1's strategy is his play when he is of type θ_1', and the second component is his play when he is of type θ_1''; payoffs are obtained by taking the expected value with respect to the prior distribution.

Using interim dominance, neither U nor D can be eliminated for either type of player 1—both types prefer U if player 2 plays L, and both prefer D if player 2 plays R. And interim iterated dominance stops at this point. However, when the two types of player 1 are equally likely, as is assumed in figure 6.5b, the type-contingent strategy DU *is* strictly dominated by UD in figure 6.5b. And once DU is deleted for player 1, L dominates R for player 2. At the next round, UU dominates UD and DD, and the unique outcome surviving *ex ante* iterated dominance is (UU, L). (If the prior probability of θ_1' is 0.9, DU is not dominated by UD.)

6.7 Using Bayesian Equilibria to Justify Mixed Equilibria†

6.7.1 Examples

In chapter 1 we saw that simultaneous-move games of *complete* information often admit mixed-strategy equilibria. Some researchers are unhappy with this notion because, they argue, "real-world decision makers do not flip coins." However, as Harsanyi (1973) has shown, mixed-strategy equilibria of complete-information games can usually be interpreted as the limits of pure-strategy equilibria of slightly perturbed games of incomplete information. Indeed, we have already noticed that in a Bayesian game, once the players' type-contingent strategies have been computed, each player behaves as if he were facing mixed strategies by his opponents. (The uncertainty arises through the distribution of types rather than through "coin flips.")

Example 6.7: "Grab the Dollar"

To illustrate the mechanics of this construction, let us consider a one-period variant of the "grab the dollar" game introduced in chapter 4. Each player has two possible actions: invest ("grab") and don't invest. In the complete-information version of the game, a firm gains 1 if it is the only one to invest, loses 1 if both invest, and breaks even if it does not invest. (We can view this game as an extremely crude representation of entry into a natural-monopoly market.) The only symmetric equilibrium is that each firm invests with probability $\frac{1}{2}$. This clearly is an equilibrium, as each firm makes 0 if it does not invest and $\frac{1}{2}(1) + \frac{1}{2}(-1) = 0$ if it does. Now consider the same game with the following type of incomplete information: Each firm has the same payoff structure, except that when it wins it gets $(1 + \theta_i)$, where θ_i is uniformly distributed on $[-\varepsilon, \varepsilon]$. Each firm knows its type, θ_i, but not that of the other firm. Now, it is easily seen that the symmetric pure strategies "$s_i(\theta_i < 0)$ = do not invest, $s_i(\theta_i \geq 0)$ = invest" form a Bayesian equilibrium. From the point of view of each firm, the other firm invests with probability $\frac{1}{2}$. Thus, the firm should invest if and only if $\frac{1}{2}(1 + \theta_i) + \frac{1}{2}(-1) \geq 0$; that is, $\theta_i \geq 0$. Last, note that, when ε converges to 0, the pure-strategy Bayesian equilibrium converges to the mixed-strategy Nash equilibrium of the complete-information game.

Example 6.8: War of Attrition††

As another example, consider the symmetric war of attrition. Suppose that, in example 6.3, it is common knowledge that the payoffs are

$$u_i(s_i, s_j) = \begin{cases} -s_i & \text{if } s_j \geq s_i \\ \hat{\theta} - s_j & \text{if } s_j < s_i. \end{cases}$$

This game has asymmetric equilibria (for instance, firm 1 always stays in, and firm 2 always exits in the natural-monopoly interpretation). But there is a single symmetric equilibrium, which is in mixed strategies. Each player uses the distribution function $F(s) = 1 - \exp(-s/\hat{\theta})$, with corresponding density $f(s) = (1/\hat{\theta})\exp(-s/\hat{\theta})$; the hazard rate for this density—i.e., the probability that a player stops between s and $s + ds$ conditional on not stopping before s is $ds/\hat{\theta}$. That this profile is an equilibrium results from the fact that the expected gain of staying in ds more is $\hat{\theta}\cdot(ds/\hat{\theta})$, which equals the cost of ds. At each instant, conditional on both players still fighting, each player's valuation from that date on (which does not include the sunk cost of having fought to that date) is equal to 0, so the player is indifferent between fighting and quitting.

Can this mixed-strategy equilibrium be purified? That is, does there exist a sequence of continuous distributions of types that weakly converge to a point mass at $\hat{\theta}$, and such that each type plays a pure strategy and the equilibrium distributions of actions converge to the one associated with the mixed-strategy equilibrium of the complete-information game?

Consider a sequence of symmetric densities $p^n(\cdot)$ on $[0, \infty)$, with cumulative distribution functions $P^n(\cdot)$ such that $P^n(0) = 0$ and such that, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} [P^n(\hat{\theta} + \varepsilon) - P^n(\hat{\theta} - \varepsilon)] = 1.$$

Let $s^n(\cdot)$ be the symmetric-equilibrium strategy corresponding to p^n, and let Φ^n be the inverse of s^n.

Integrating equation 6.3 (the first-order condition for maximization) shows that

$$P^n(\Phi^n(s)) = 1 - \exp\left(-\int_0^s db/\Phi^n(b)\right). \tag{6.13}$$

Since $P^n(\hat{\theta} - \varepsilon)$ converges to 0, and $P^n(\hat{\theta} - \varepsilon) = P^n(\Phi^n(s^n(\hat{\theta} - \varepsilon)))$, equation 6.13 implies that $s^n(\hat{\theta} - \varepsilon)$ converges to 0 for all $\varepsilon > 0$. Similarly, one can show that $s^n(\hat{\theta} + \varepsilon)$ converges to infinity. Hence, for any $s > 0$ and $\varepsilon \in (0, \hat{\theta})$,

$$s^n(\hat{\theta} - \varepsilon) < s < s^n(\hat{\theta} + \varepsilon)$$

for n sufficiently large. Rewrite equation 6.13 as

$$\begin{aligned} P^n(\Phi^n(s)) &= 1 - \exp\left(-\int_0^{s^n(\hat{\theta}-\varepsilon)} \frac{db}{\Phi^n(b)}\right)\exp\left(-\int_{s^n(\hat{\theta}-\varepsilon)}^{s} \frac{db}{\Phi^n(b)}\right) \\ &= 1 - [1 - P^n(\hat{\theta} - \varepsilon)]\exp\left(-\int_{s^n(\hat{\theta}-\varepsilon)}^{s} \frac{db}{\Phi^n(b)}\right). \end{aligned} \tag{6.14}$$

Since $P^n(\hat{\theta} - \varepsilon)$ and $s^n(\hat{\theta} - \varepsilon)$ converge to 0, for n sufficiently large, $\Phi^n(b) \in b \in [\hat{\theta} - \varepsilon, \hat{\theta} + \varepsilon]$ for all $b \in [s^n(\hat{\theta} - \varepsilon), s]$, so that any accumulation point of

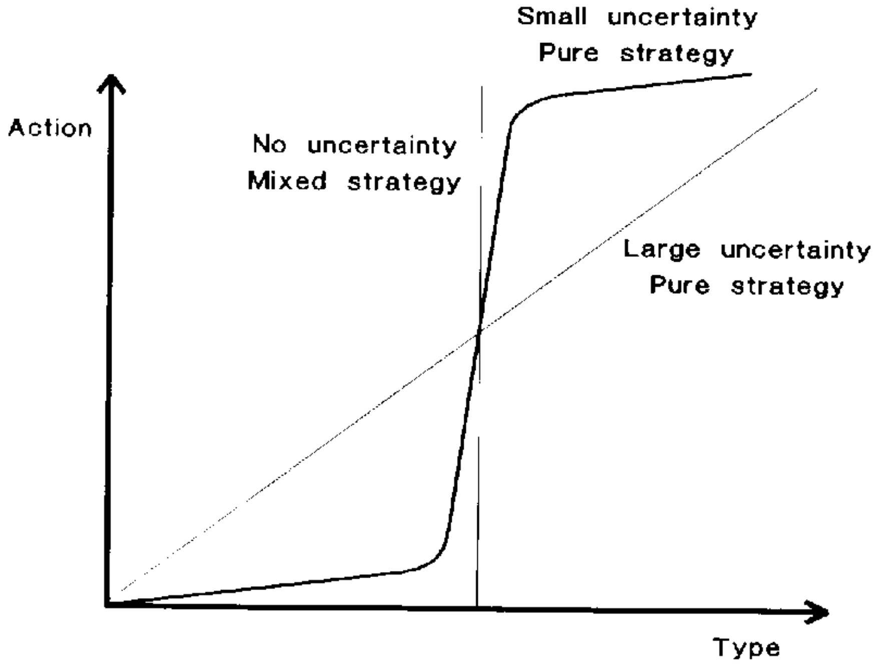


Figure 6.6

$P^n(\Phi^n(s))$ is between $1 - \exp[-s/(\hat{\theta} + \varepsilon)]$ and $1 - \exp[-s/(\hat{\theta} - \varepsilon)]$. Since this is true for all $\varepsilon > 0$, we conclude that

$$P^n(\Phi^n(s)) \to P(\Phi(s)) = 1 - \exp(-s/\hat{\theta}).$$

Thus, once again, a sequence of pure-strategy equilibria for an incomplete-information game "converges" to a mixed-strategy equilibrium of the complete-information version of the game. Our analysis focused on the convergence of the probability distributions over actions. Figure 6.6 illustrates the convergence in the space of strategies.

Example 6.9: First-Price Auction

As a last example, consider the first-price auction with a continuum of types and with two types (examples 6.5 and 6.6). Differentiating equation 6.11 (corresponding to the two-type case) for $s > \underline{\theta}$ yields

$$G(s) = (\theta - s)g(s). \tag{6.15}$$

To compare equation 6.15 with equation 6.6 (which corresponds to the continuum case),[19] consider a sequence of continuous distributions $P^n(\cdot)$ converging to spikes at $\underline{\theta}$ and θ ($\lim_{n\to\infty} P^n(\theta) = 0$ for $\theta < \underline{\theta}$, $= \underline{p}$ for $\theta \in [\underline{\theta}, \theta)$, $= 1$ for $\theta \geq \overline{\theta}$). If $\Phi^n(\cdot)$ denotes the equilibrium strategy for distribution $P^n(\cdot)$, then $\Phi^n(s)$ must converge to $\overline{\theta}$ for $s > \underline{\theta}$, and thus (loosely speaking) equation 6.6 converges to equation 6.15.

19. In the continuous example, we imposed a reservation price. Take the reservation price to be equal to θ in the discrete example to make the two games compatible.

6.7.2 Purification Theorem (technical)$^{++}$

Harsanyi (1973) shows that any mixed-strategy equilibrium can "almost always" be obtained as the limit of a pure-strategy equilibrium in a given sequence of slightly perturbed games. Consider a strategic-form game with finite strategy sets S_i and payoff functions u_i. Harsanyi perturbs the payoffs in the following way: Let θ_i^s denote a random variable with range a closed interval ($[-1, 1]$, say) and let $\varepsilon > 0$ denote a positive constant (which will later converge to 0). Player i's perturbed payoff function $\tilde{u}_i$ depends on player i's "type" $\theta_i \equiv \{\theta_i^s\}_{s \in S}$ and on the "scale of perturbation" ε:

$$\tilde{u}_i(s, \theta_i) = u_i(s) + \varepsilon\theta_i^s.$$

Harsanyi assumes that the players' types are statistically independent. Let $P_i(\cdot)$ denote the probability distribution for θ_i. It is assumed that P_i has a density function $p_i(\cdot)$ that is continuously differentiable for all θ_i. Harsanyi first shows that the best reply of any player i is an essentially unique pure strategy. That is, two best replies for player i, $\sigma_i(\cdot)$ and $\tilde{\sigma}_i(\cdot)$, must coincide for almost all θ_i, and furthermore they must be pure strategies for almost all θ_i. This is quite intuitive, since for given strategies of player i's opponents the coincidence of player i's payoffs for two pure strategies must be a rare event if θ_i is continuously distributed. As a consequence, in any equilibrium of a perturbed game, $\sigma_i(\theta_i)$ is a pure strategy for all i and for almost all $\theta \equiv (\theta_1, \ldots, \theta_I)$. Harsanyi shows that an equilibrium exists. He then proves the following result.

Theorem 6.1 (Harsanyi 1973) Fix a set of I players and strategy spaces S_i. For a set of payoffs $\{u_i(s)\}_{i \in \mathscr{I}, s \in S}$ of Lebesgue measure 1, for all independent, twice-differentiable distributions p_i on $\Theta_i = [-1, 1]^{\# S}$, any equilibrium of the payoffs u_i is the limit as $\varepsilon \to 0$ of a sequence of pure-strategy equilibria of the perturbed payoffs $\tilde{u}_i$. More precisely, the probability distributions over strategies induced by the pure-strategy equilibria of the perturbed game converge to the distribution of the equilibrium of the unperturbed game.

Note the order of quantifiers in the statement of the theorem: A single sequence of perturbed games can be used to "purify" *all* the mixed equilibria of the limit game.

Note also the restriction to a set of payoffs of full measure. There are two possible problems that occur for "pathological" payoffs. First, it may be that a given equilibrium can only be approximated by pure-strategy equilibria of a small subset of all perturbed games, and different perturbed games pick out different equilibria. Exercise 6.10 gives an example of this. Second, equilibria in weakly dominated strategies are not limits of equilibria of any perturbed game. In figure 6.7 (taken from Harsanyi 1973), the pure-strategy equilibrium (D, R) is not approachable by any equilibrium

	L	R
U	3,4	2,2
D	1,1	2,1

Figure 6.7

once the game is perturbed. For instance, suppose that the random variables θ_1^{UR} and θ_1^{DR} are symmetrically (e.g., uniformly) distributed on $[-1, 1]$. Then, for any probability that player 2 plays R, the probability that player 1 strictly prefers to play U is at least $\frac{1}{2}$. Hence, player 1's strategy in the perturbed game cannot converge to probability 1 on D. But games like that depicted in figure 6.7 are extraordinary. In the (D, R) equilibrium, players are indifferent between their equilibrium strategy and a dominating strategy a situation that is unlikely to occur if the entries in figure 6.7 are drawn "randomly."[20]

Our view is that games of complete information are an idealization, as players typically have at least a slight amount of incomplete information about the others' objectives. One consequence of that view, as Harsanyi's argument shows, is that the distinction between pure and mixed strategies may be artificial.

6.8 The Distributional Approach (technical)†††

Modeling mixed strategies as maps from types to mixtures over pure strategies has the drawback that it is not well defined in games with a continuum of types, as Aumann (1964) pointed out. Aumann proposed that a mixed strategy be a function σ_i from $[0, 1] \times \Theta_i$ into S_i. The interpretation is that type θ_i chooses among actions s_i on the basis of the outcome x_i of a lottery. Assuming without loss of generality that x_i is drawn from the uniform distribution on $[0, 1]$,[21] the probability that type θ_i of player i plays s_i is equal to the measure of the set of x_i such that $\sigma_i(x_i, \theta_i) = s_i$. There are, of course, an infinity of mixed strategies that describe a given behavior. For instance, the following mixed strategies are "equivalent":

$$\sigma_i(x_i, \theta_i) = s_i \text{ if } x_i \leq \tfrac{1}{3}, \quad \sigma_i(x_i, \theta_i) = s_i' \text{ if } x_i > \tfrac{1}{3},$$

and

20. Recall from chapter 3 that the set of all strategic-form payoffs arising from a given extensive form can have measure 0 in the space of all payoffs for that strategic form.

21. To see that we can assume a uniform distribution without loss of generality, consider a mixed strategy $\sigma_i(y_i, \theta_i)$ where y_i is distributed on $[0, 1]$ according to the increasing cumulative distribution function $F_i(y_i)(F_i(0) = 0,\ F_i(1) = 1)$. Define the new strategy $\tilde{\sigma}_i(x_i, \theta_i) \equiv \sigma_i(F_i(x_i), \theta_i)$. This mixed strategy is a function of the random variable x_i, which is uniformly distributed on $[0, 1]$ (as $\text{Prob}(x_i \leq x) = \text{Prob}(F_i^{-1}(x_i) \leq F_i^{-1}(x)) = F_i(F_i^{-1}(x)) = x$).

$$\tilde{\sigma}_i(x_i, \theta_i) = s_i \text{ if } x_i > \tfrac{2}{3}, \quad \tilde{\sigma}_i(x_i, \theta_i) = s_i' \text{ if } x_i \leq \tfrac{2}{3}.$$

In other words, the "Aumann fix" is not parsimonious.

In response, Milgrom and Weber (1986) introduced the concept of a "distributional strategy," which refers to the equivalence class of the mixed strategies that yield the same behavior. From the point of view of the other players, what matters is the joint distribution of player i's type and actions. This leads to the definition of a distributional strategy as a joint distribution on $\Theta_i \times S_i$ for which the marginal distribution on Θ_i is the one specified by the prior beliefs.

The equivalence between mixed strategies and distributional strategies is clear. A mixed strategy induces a joint distribution across types and actions. Conversely, a joint distribution can be generated by many mixed strategies.

The reader familiar with the notion of correlated equilibrium introduced in chapter 2 will note the analogy between definitions A and B of correlated equilibrium and the distinction between mixed and distributional strategies. In chapter 2, we noted that we could determine the set of correlated equilibria without considering all possible correlating devices, but instead could restrict attention to joint distributions over strategies. Similarly, we do not need to list all the possible relationships between the randomizing device and the strategy; instead we can focus on the joint distribution of the player's type and action.

Since pure-strategy equilibria need not exist in games of complete information, it is interesting to note that, under certain regularity conditions, pure-strategy equilibria do exist in games with an atomless distribution over types. (Mixed strategies are needed for existence in general incomplete-information games.)

The idea is that the effects of mixing can be duplicated by having each type play a pure strategy. If each player's payoff does not depend on the types of the others, then players care only about the distribution of their opponents' actions, and their payoffs are not affected by the replacement of an opponent's mixed strategy by a pure one that induces the same distribution.

To illustrate this, suppose that type θ_i is uniformly distributed on the interval $[0, 1]$ and that, given the strategies of player i's rivals, all types θ_i in $[0, \frac{1}{2}]$ are indifferent between actions s_i and s_i'. These types randomize in such a way that the probability that s_i (respectively, s_i') is chosen given that θ_i belongs to $[0, \frac{1}{2}]$ is α (respectively, $1 - \alpha$). Consider the following pure strategy: Types θ_i in $[0, \alpha/2]$ play s_i with probability 1, and types θ_i in $(\alpha/2, \frac{1}{2}]$ play s_i' with probability 1. Because types in $[0, \frac{1}{2}]$ are indifferent between the two actions, the pure strategy is an equilibrium behavior as long as the rivals do not change their behavior. The rivals' expected payoff function is not affected by the substitution if two conditions hold. The first condition is that θ_i should not enter player j's utility function or, more generally, should be separable from s_i in an appropriate way: Even though the marginal distributions of s_i and θ_i are not affected by the substitution, the distributional strategy is affected, and this matters if there are cross-effects between s_i and θ_i in u_j. The second condition is that the distributions of types should be independent among players. (If this is not the case, the distribution of s_i, conditional on θ_j, may be changed by the substitution.)

Along these lines, we can state a "purification theorem" due to Milgrom and Weber's (1986) extension of a similar result for the single-decision-maker setting of Dvoretzky, Wald, and Wolfowitz (1951). For this purpose (and for the rest of this section), we will assume that the structure of information takes the special form of a commonly observed variable θ_0 (common value) and some piece of private information $\tilde{\theta}_i$ for each player i (private values) such that, conditional on the realization of θ_0, the $\tilde{\theta}_i$ are independent. Let $\theta_i = (\theta_0, \tilde{\theta}_i)$; because θ_0 is commonly observed, $\tilde{\theta}_i$ will be called "player i's type" by abuse of terminology. We assume that $\theta_0 \in \Theta_0$ and $\tilde{\theta}_i \in \Theta_i$.

Definition 6.2 Preferences are *conditionally independent* if each player i's payoff can be written in the form $u_i = u_i(s, \theta_0, \tilde{\theta}_i)$, where $s \equiv (s_1, \ldots, s_I)$, and if, conditional on the realization of θ_0, the players' types $\tilde{\theta}_i$ are independent.

Theorem 6.2 (Milgrom and Weber 1986) Assume that preferences are "conditionally independent," that Θ_0 is finite, that the marginal distributions of types are atomless, that the game has continuous payoffs, and that each S_i is compact. Then every equilibrium point (we will later show that one exists) has a purification.

Remark 1 The assumption of conditionally independent preferences is obviously very strong. One may be able to purify mixed strategies even when it does not hold. When preferences are dependent, one must be able to replicate by a pure strategy not only the distribution on S_i, but the whole distributional strategy on $S_i \times \Theta_i$. That is, the reshuffling of weight we performed earlier must be "local" rather than "global." Unfortunately, we do not quite know what regularity conditions are required to this effect.[22] The issue is that the set of distributional strategies obtained from pure strategies is smaller than the set of distributional strategies obtained from mixed strategies. (Those two sets, however, are close to each other: The former is dense in the latter for the topology of weak convergence of

22. Aumann et al. (1982) allow dependence, but obtain only an approximate purification result. They show that with conditionally atomless distributions, any mixed strategy of a player can be ε-purified (i.e., replaced by a pure strategy that yields all players a payoff within ε of the payoff for the original mixed strategy), no matter what strategies the other players use, for any $\varepsilon > 0$.

probability measures. Hence, for any mixed-strategy equilibrium, there exists a set of nearly pure strategies that form an ε-equilibrium of the game.)

Remark 2 In subsection 6.7.2 we used the term "purification" in a different, although related, sense. We asked to what extent a mixed-strategy equilibrium of a game with complete information (or, more generally, with atoms of types) could be viewed as an approximation of pure-strategy equilibria of nearby games of incomplete information in which each player has a continuum of types.

With a continuum of types and/or a continuum of actions, some regularity conditions must be imposed in order to apply Glicksberg's existence theorem (see subsection 1.3.3). Let η and η_i $(i = 0, \ldots, I)$ denote, respectively, the probability measure over the set $\Theta = \Theta_0 \times \Theta_1 \times \cdots \times \Theta_I$ and the marginal distribution over Θ_i. The following existence result (a slightly stronger version of which can be found in Milgrom and Weber 1986) generalizes one obtained for independent types by Ambruster and Böge (1979).

Theorem 6.3 (Milgrom and Weber 1986) Assume that all S_i are compact; that (continuous information) the measure $\eta(\cdot)$ is absolutely continuous relative to the measure $\hat{\eta}(\cdot) = \eta_0(\cdot) \times \cdots \times \eta_I(\cdot)$;[23] and that (continuous payoffs) either all S_i are finite or, for all i, the function u_i is uniformly continuous on $\Theta \times S$. Then an equilibrium exists.

Exercises

Exercise 6.1** Consider the public-good game of section 6.2. Suppose that there are $I > 2$ players and that the public good is supplied (with benefit 1 for all players) only if at least $K \in \{1, \ldots, I\}$ players contribute. The players' costs of contributing, $\theta_1, \ldots, \theta_I$, are independently drawn from the distribution $P(\cdot)$ on $[\underline{\theta}, \bar{\theta}]$ where $\underline{\theta} < 1 < \bar{\theta}$.

(a) Generalize the Bayesian equilibrium of section 6.2 when $K = 1$.

(b) Suppose $K \geq 2$. Show that there always exists a trivial equilibrium in which nobody contributes. (Assume $\underline{\theta} > 0$.) Derive a more interesting Bayesian equilibrium.

(c) Showoffs: Apply the two concepts of iterated strict dominance to this game.

Exercise 6.2** Two firms simultaneously decide whether to enter a market. Firm i's entry cost is $\theta_i \in [0, +\infty)$. The two firms' entry costs are private

23. That is, a null-measure set for $\hat{\eta}$ is also a null-measure set for η. The Radon-Nikodym theorem (Royden 1968) implies that there exists a density f such that, for any subset S of θ, $\eta(S) = \int_S f(\theta)\, d\hat{\eta}(\theta)$. The continuous-information assumption holds for instance when the type spaces are finite or when the types are independently distributed.

information and are independently drawn from the distribution $P(\cdot)$ with strictly positive density $p(\cdot)$. Firm i's payoff is $\Pi^m - \theta_i$ if it is the only one to enter, $\Pi^d - \theta_i$ if both enter, and 0 if it does not enter. Π^m and Π^d are the monopoly and duopoly profits gross of entry costs and are common knowledge. $\Pi^m > \Pi^d > 0$.

(a) Point out the analogies and differences with the public-good game of section 6.2.

(b) Compute a Bayesian equilibrium. Show that it is unique.

(c) Apply *ex ante* and interim strict dominance.

(d) Instead of assuming simultaneous entry, suppose that the two firms are "around" at date 0. They incur cost $r\theta_i$ per unit of time of being in the market, where r is the rate of interest and θ_i is the value of firm i's assets in an alternative use. (Alternatively, $f_i \equiv r\theta_i$ is the fixed cost of production per unit of time.) $r\Pi^m$ and $r\Pi^d$ are the flow monopoly and duopoly payoffs gross of the opportunity or fixed cost. Follow the analysis of example 6.3 to derive the symmetric equilibrium of the war of attrition. (Watch out: There exists some time T after which firms don't drop out.) Show that there is no other equilibrium. For answers, see Fudenberg and Tirole 1986. Compare with the answer to question b.

Exercise 6.3* Consider the first-price auction of example 6.5. There are two bidders with valuations uniformly distributed on $[0, 1]$. The seller's reservation price (or minimum bid) is 0. Find an equilibrium in linear strategies, i.e., where $s_i(\theta_i) = a + c\theta_i$.

Exercise 6.4** This exercise analyzes the first- and second-price auctions with risk-averse bidders and two types per bidder. A bidder with valuation θ has utility $u(\theta - t)$ if he wins and pays transfer t, and utility $u(-t)$ if he loses and pays transfer t; u is increasing and concave. The bidders' valuations are independently drawn from the two-type distribution $\{\underline{\theta}$ with probability $\underline{p}$, $\bar{\theta}$ with probability $\bar{p}\}$.

(a) Show that in the second-price auction (in which the highest bidder wins but pays the second bid) each player bids his true valuation and so the seller's expected revenue is the same as with risk-neutral bidders.

(b) Now consider the first-price auction. Derive the analogue of equation 6.10 for the case of risk aversion. Show that type $\bar{\theta}$'s distribution of bids, $\tilde{F}$, first-order stochastically dominates the distribution F given by equation 6.10 (i.e., $\tilde{F}(s) \leq F(s)$ for all s). Use the revenue-equivalence theorem under risk neutrality (see example 6.6) to conclude that under risk aversion the seller prefers the first-price to the second-price auction. (The answer is in Maskin and Riley 1985.)

Exercise 6.5** Generalize the analysis of the first- and second-price auctions with two types and risk-neutral bidders to (a) asymmetric distributions (letting $\bar{p}_i \equiv \text{Prob}(\theta_i = \bar{\theta}), \bar{p}_1 \neq p_2$) and (b) correlated valuations.

Compare the seller's revenues in the two auctions (do parts a and b separately). (The answers are in Maskin and Riley 1985, 1986b.)

Exercise 6.6** In the examples in the chapter, equilibrium strategies are monotonic in type. Find and informally discuss examples in which such a monotonicity would not necessarily hold. (Hint: Consider the Chatterjee-Samuelson double auction with negatively correlated types. Find other examples. Discuss generally what goes wrong with the usual proof of monotonicity when types are correlated.)

Exercise 6.7* The (present discounted) value of a public good is 1 for all players $i = 1, \dots, I$. Time is continuous, and the rate of interest is r. Each player's cost c of supplying the public good is distributed according to the cumulative distribution function P on $[0, 1]$. Players' types are independent. The public good is supplied if at least one agent supplies it. The good is supplied at the first time at which at least one player chooses to contribute. Thus, the game is a kind of war of attrition. Look for a symmetric, pure-strategy equilibrium using the following outline:

(a) Argue formally or informally that the date at which a player with cost c supplies the public good, $s(c)$, is increasing in c.

(b) Show that $s(\cdot)$ satisfies

$$s'(c) = \frac{(I-1)c\,p(c)}{r(1-c)[1-P(c)]}.$$

Find a boundary condition. Infer that a player's waiting time to supply the good when there are $I-1$ other players is $I-1$ times his waiting time when there are two players. Show that each player's expected utility grows with I.

(Answers can be found in Bliss and Nalebuff 1984.)

Exercise 6.8** Kreps and Wilson (1982) consider the following war of attrition. There are two players, $i = 1, 2$. Time is continuous from 0 to 1. When one player concedes, the game ends. Each player can be either "strong" (with probability p for player 1 and q for player 2) or "weak" (with probabilities $1-p$ and $1-q$). A strong player enjoys fighting and therefore never concedes. A weak player 1 (respectively, a weak player 2) loses 1 per unit of time while fighting and makes $a>0$ (respectively, $b>0$) per unit of time when his rival has conceded. Thus, a weak player 1 has payoff $a(1-t)-t$ when it wins at t and payoff $-t$ when it concedes at t. There is no discounting.

(a) Show that from time 0^+ on, the posterior beliefs p_t and q_t of each player about the other must belong to the curve $q = p^{b/a}$.

(b) Show that one of the weak types exits with positive probability at date 0 exactly (that is, a player's cumulative probability distribution of exit times exhibits an atom at $t = 0$). How are the weak types' payoffs affected by a, b, p, and q?

	L	R
U	1,1	1,1
D	1,1	1,1

Figure 6.8

Exercise 6.9** Purify the mixed-strategy equilibrium in the inspection game of example 1.7.

Exercise 6.10** Consider the game illustrated in figure 6.8 (due to Harsanyi). Fix a continuous distribution over perturbations as in Harsanyi's construction (see section 6.7). Is any mixed-strategy equilibrium of the game in figure 6.8 a limit of pure-strategy equilibria of the perturbed games as ε tends to 0? Conclude that this game is "not generic."

Exercise 6.11*** Consider symmetric first-bid and second-bid auctions with correlated information and valuations. There are I bidders. Each bidder i has (unknown) valuation v_i and signal or information θ_i; let $z_i \equiv (\theta_i, v_i)$ and $z \equiv (z_1, \ldots, z_I)$. Player i knows only θ_i. The random variable z is distributed according to distribution $F(z)$ on a rectangular cell, with density $f(z)$. F is invariant under permutations of the bidders (symmetry). Correlation is described by the affiliation property: If $z \vee z'$ and $z \wedge z'$ are the component-wise maximum and minimum of z and z', then

$$f(z \vee z')f(z \wedge z') \geq f(z)f(z') \text{ for all } (z, z').$$

Let $\theta^1 \geq \theta^2 \geq \cdots \geq \theta^I$ denote a reordering in nonincreasing order of the signals. The conditional distribution of θ^2 given $\theta^1 = \gamma$ is denoted $\hat{F}(\cdot|\gamma)$ with density $\hat{f}(\cdot|\gamma)$. Affiliation implies that, for all μ, $\hat{f}(\mu|\gamma)/\hat{F}(\mu|\gamma)$ is nondecreasing in γ (monotone likelihood-ratio property). Let

$$v(\gamma, \mu) \equiv \mathrm{E}(v_i | \theta_i = \theta^1 = \gamma \text{ and } \theta^2 = \mu).$$

Look for symmetric, differentiable, and strictly increasing equilibrium bids $s(\theta_i)$.

Note that in the second-price auction $s(\theta) = v(\theta, \theta)$. Show that in the first-price auction

$$s(\theta) = v(\theta, \theta) - \int_{\underline{\theta}}^{\theta} K(\mu) \frac{dv}{d\mu}(\mu, \mu) \Big/ K(\theta),$$

where

$$K(\mu) = \exp\left(\int_{\underline{\theta}}^{\mu} \frac{\hat{f}(\gamma, \gamma)}{\hat{F}(\gamma, \gamma)} d\gamma \right)$$

and $\underline{\theta}$ is the lowest possible signal. (Hint: In the first-price auction, a bidder with type θ maximizes over his bid b:

$$\int_{\theta}^{s^{-1}(b)} [v(\theta, \mu) - b]\, d\hat{F}(\mu \mid \theta).$$

For the answer, look in Milgrom and Weber 1982 or Wilson 1990.)

References

Ambruster, W., and W. Böge. 1979. Bayesian game theory. In *Game Theory and Related Topics*, ed. O. Moeschlin and D. Pallascke. North-Holland.

Aumann, R. 1964. Mixed vs. behavior strategies in infinite extensive games. *Annals of Mathematics Studies* 52: 627–630.

Aumann, R., Y. Katznelson, R. Radner, R. Rosenthal, and B. Weiss. 1982. Approximate Purification of Mixed Strategies. *Mathematics of Operations Research* 8: 327–341.

Bishop, D. T., C. Cannings, and J. Maynard Smith. 1978. The war of attrition with random rewards. *Journal of Theoretical Biology* 3: 377–388.

Bliss, C., and B. Nalebuff. 1984. Dragon-slaying and ballroom dancing: The private supply of the public good. *Journal of Public Economics* 25: 1–12.

Butters, G. 1977. Equilibrium distribution of prices and advertising. *Review of Economic Studies* 44: 465–492.

Chatterjee, K., and W. Samuelson. 1983. Bargaining under incomplete information. *Operations Research* 31: 835–851.

Dvoretzky, A., A. Wald, and J. Wolfowitz. 1951. Elimination of randomization in certain statistical decision procedures and zero-sum two-person games. *Annals of Mathematics and Statistics* 22: 1–21.

Fudenberg, D., and J. Tirole. 1986. A theory of exit in duopoly. *Econometrica* 54: 943–960.

Harsanyi, J. 1967–68. Games with incomplete information played by Bayesian players. *Management Science* 14: 159–182, 320–334, 486–502.

Harsanyi, J. 1973. Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory* 2: 1–23.

Kreps, D., and R. Wilson. 1982. Reputation and imperfect information. *Journal of Economic Theory* 27: 253–279.

Leininger, W., P. Linhart, and R. Radner. 1989. Equilibria of the sealed-bid mechanism for bargaining with incomplete information. *Journal of Economic Theory* 48: 63–106.

Maskin, E., and J. Riley. 1985. Auction theory and private values. *American Economic Review Papers & Proceedings* 75: 150–155.

Maskin, E., and J. Riley. 1986a. Existence and uniqueness of equilibrium in sealed high bid auctions. Discussion paper 407, University of California, Los Angeles.

Maskin, E., and J. Riley. 1986b. Asymmetric auctions. Mimeo, UCLA and Harvard University.

Mertens, J. F., S. Sorin, and S. Zamir. 1990. Repeated games. Manuscript.

Mertens, J. F., and S. Zamir. 1985. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory* 10: 619–632.

Milgrom, P., and R. Weber. 1982. A Theory of Auctions and Competitive Bidding. *Econometrica* 50: 1089–1122.

Milgrom, P., and R. Weber. 1986. Distributional strategies for games with incomplete information. *Mathematics of Operations Research* 10: 619–631.

Nalebuff, B. 1982. Brinksmanship. Mimeo, Harvard University.

Nalebuff, B., and J. Riley. 1983. Asymmetric equilibria in the war of attrition. Mimeo, University of California, Los Angeles.

Nash, J. 1953. Two-person cooperative games. *Econometrica* 21: 128–140.

Palfrey, T., and H. Rosenthal. 1989. Underestimated probabilities that others free ride: An experimental test. Mimeo, California Institute of Technology and Carnegie-Mellon University.

Riley, J. 1980. Strong evolutionary equilibrium and the war of attrition. *Journal of Theoretical Biology* 82: 383–400.

Royden, H. 1968. *Real Analysis*. Macmillan.

Satterthwaite, M., and S. Williams. 1989. Bilateral trade with the sealed bid k-double auction: existence and efficiency. *Journal of Economic Theory* 48: 107–133.

Wilson, R. 1990. Strategic analysis of auctions. Mimeo, Graduate School of Business, Stanford University.

7 Bayesian Games and Mechanism Design

This chapter presents a thorough treatment of a special class of games of incomplete information known as *games of (static) mechanism design.* Examples of these games include monopolistic price discrimination, optimal taxation, the design of auctions, and mechanisms for the provision of public goods. In all of these cases, there is a "principal" who would like to condition her actions on some information that is privately known by the other players, called "agents." The principal could simply ask the agents for their information, but they will not report it truthfully unless the principal gives them an incentive to do so, either by monetary payments or with some other instrument that she controls. Since providing these incentives is costly, the principal faces a tradeoff that often results in an inefficient allocation.

The distinguishing characteristic of the mechanism-design approach is that the principal is assumed to choose the mechanism that maximizes her expected utility, as opposed to using a particular mechanism for historical or institutional reasons. This distinction can be illustrated using the topic of auctions: In chapters 1 and 6 we solved for the equilibrium bidding strategies of buyers in two particular mechanisms, the first-price and second-price auctions. When we study auctions in this chapter, we ask which form of auction maximizes the seller's expected revenue. Because of the pervasiveness of the first-price auction, it is very interesting to see that it (and the second-price auction) turns out to be optimal in some situations. Similarly, when we consider models where the principal is the government, we suppose that the government chooses a mechanism that maximizes its utility, which we take to be the total surplus in the economy. Thus, the applications to (e.g.) tax policy may be interpreted as normative as opposed to descriptive models.

Many applications of mechanism design consider games with a single agent. (These single-agent models also apply to situations with a continuum of infinitesimal agents, each of whom interacts with the principal but not with the other agents.) In second-degree price discrimination by a monopolist, the monopolist has incomplete information about the consumer's (the agent's) willingness to pay for her good. The monopolist designs a price schedule that determines the price to be paid by the consumer as a function of the quantity purchased. In the regulation of a natural monopoly under asymmetric information, the government has incomplete information about the regulated firm's (the agent's) cost structure. It designs an incentive scheme that determines the transfer received by the regulated firm as a function of its cost or its price (or both). In the study of optimal taxation, the government would like to raise tax revenue from a consumer (an agent) to finance public goods. The optimal level of tax depends on the consumer's ability to earn money. If the government knew this ability, it could levy an ability-dependent lump-sum tax that would not distort the consumer's labor supply. In the presence of incomplete information about ability, the government can only base the income tax on realized income. The income-tax schedule can be seen as an incentive scheme eliciting information about the consumer's ability.

Mechanism design can also be applied to games with several agents. In the public-good problem, a government must decide whether to supply a public good, but it has incomplete information about how much the good is valued by consumers. The government can then design a scheme determining the provision of the public good as well as transfers to be paid by the consumers as functions of their announced willingnesses to pay for the public good. In the design of auctions, a seller organizes an auction among the buyers for the purchase of a good. The seller, not knowing how much the buyers are willing to pay for the good, set up a mechanism that determines who purchases the good and the sale price. Finally, in problems of bilateral exchange, a mediator designs a trading mechanism between a seller who has private information about the production cost and a buyer who has private information about his willingness to pay for the good.

Mechanism design is typically studied as a three-step game of incomplete information, where the agents' types—e.g., willingness to pay—are private information. In step 1, the principal designs a "mechanism," or "contract," or "incentive scheme." A mechanism is a game in which the agents send costless messages, and an "allocation" that depends on the realized messages. The message game can have simultaneous announcements or a more complex communication process. The allocation is a decision about the level of some observable variable, e.g., the quantity consumed or the amount of public good provided, and a vector of transfers from the principal to the agents (which can be positive or negative). In step 2, the agents simultaneously accept or reject the mechanism. An agent who rejects the mechanism gets some exogenously specified "reservation utility" (usually, but not necessarily, a type-independent number). In step 3, the agents who accept the mechanism play the game specified by the mechanism.

Because a game of mechanism design can have many stages, the distinction between Nash and subgame-perfect equilibria in multi-stage games with complete information (see chapter 3) may suggest that the concept of Bayesian equilibrium is too weak to be useful here. Fortunately, a simple but fundamental result called the "revelation principle" (developed in section 7.2) shows that, to obtain her highest expected payoff, the principal can restrict attention to mechanisms that are accepted by all agents at step 2 and in which at step 3 all agents simultaneously and truthfully reveal their types. This implies in particular that the principal can obtain her highest expected payoff through a static Bayesian game among the agents. This is why we treat mechanism design in part III rather than in part IV of the book. (However, we invoke a mild perfection requirement: We do not allow agents to threaten to reject the principal's mechanism—or to misreport their types—if it is in their interest in steps 2 and 3 to accept the mechanism and to announce truthfully.)

In some situations (mainly situations in which the principal is the government), the "individual-rationality" or "participation" constraints—that the agents must be willing to participate in the principal's mechanism—are not imposed. That is, step 2 of the mechanism-design game is omitted. For instance, a government with coercive powers can choose an income tax that applies to all consumers (unless the possibility of emigration makes the participation constraints binding). Similarly, in some public-good problems, the government may impose decisions that the agents cannot veto. In contrast, the literature has assumed that consumers can refrain from buying from a firm, that bidders are free not to participate in an auction, and that regulated firms (or at least their managers) can refuse to produce (or to work). Whether an individual-rationality constraint should be included in the model depends on the extent of the coercive power of the principal, or, equivalently, on the distribution of property rights.[1]

An important focus of the mechanism-design literature is how the combination of incomplete information and binding individual-rationality constraints can prevent efficient outcomes.[2] Coase (1960) argued that, in the absence of transaction costs and with symmetric information, bargaining among parties concerned by a decision leads to an efficient decision, i.e., to the realization of gains from trade. With some exceptions (see the "efficiency results" in subsection 7.4.3), this is not so under asymmetric information. A constant theme of the mechanism-design literature is that the private information of the agents leads to inefficiency when individual-rationality constraints are binding.

The chapter is organized as follows. Section 7.1 illustrates individual rationality, truthful revelation, and optimal mechanism design in two simple examples. Section 7.2 develops the general framework and derives the revelation principle. Section 7.3 considers the case of a single agent. Besides being of considerable practical interest, this case offers a useful introduction to the more general multi-agent situation. Most of the steps involved in characterizing "implementable" or "incentive-compatible" allocations and in deriving the optimal mechanism for the principal are borrowed from the single-agent framework. Section 7.4 tackles the multi-agent case and characterizes implementable allocations. Subsections 7.4.3–7.4.6 apply this characterization to obtain some efficiency and inefficiency results in public-

1. The literature has made reasonable assumptions about the actual distribution of property rights, but little attention has been paid to what determines it in most of the contexts described above.

2. If the principal is a government that does not have a balanced-budget constraint, so that it can give all agents large positive transfers, the individual-rationality constraints will not bind.

and private-goods contexts. Section 7.5 builds on section 7.4 by analyzing the principal's optimal mechanism in two different contexts: auctions, in which a seller tries to extract the maximum expected revenue from buyers, and bilateral exchange, in which a mediator designs a mechanism so as to maximize expected gains from trade between a seller and a buyer. Section 7.6 mentions some additional topics and concludes.

The field of mechanism design is important enough to merit a book of its own.[3] We have not tried to provide a complete review of the field, choosing instead to develop a few main themes. Nevertheless, the material in this chapter could easily take a month to cover. Readers with little interest in mechanism design might choose to skip the chapter entirely, relying on the examples in chapter 6 to illustrate the application of Bayesian equilibrium. Those who are interested in mechanism design but are pressed for time might read through section 7.3, which completes the analysis of mechanism design with a single agent.

7.1 Examples of Mechanism Design†

This section contains two examples of mechanism design. To facilitate the exposition, they both involve a seller selling a good—to a single buyer in subsection 7.1.1 and to one of two buyers in subsection 7.1.2. This section is meant to provide motivation for the chapter; it should be skipped by any reader who has already seen some examples.

7.1.1 Nonlinear Pricing

A monopolist produces a good at constant marginal cost c and sells an amount $q \geq 0$ of this good to a consumer. (As is easily checked, nothing would be affected if she sold the good to several consumers who were *ex ante* identical.) The consumer receives utility

$$u_1(q, T, \theta) \equiv \theta V(q) - T,$$

where $\theta V(q)$ is his gross surplus, $V(0) = 0$, $V' > 0$, $V'' < 0$, and T is the transfer from the consumer to the seller. $V(\cdot)$ is common knowledge, but θ is private information to the consumer. The seller knows only that $\theta = \underline{\theta}$ with probability $\underline{p}$ and $\theta = \bar{\theta}$ with probability $\bar{p}$, where $\bar{\theta} > \underline{\theta} > 0$ and $\underline{p} + \bar{p} = 1$. The game proceeds as follows: The seller offers a (possibly nonlinear) tariff $T(q)$ specifying how much the consumer pays if he chooses consumption q. The consumer then either accepts the mechanism, chooses a consumption q, and pays $T(q)$, or else rejects the mechanism. Note that, without loss of generality, we can constrain the seller to offer a tariff such that $T(0) = 0$ and assume that the consumer always accepts the mechanism.

3. See, e.g., Green and Laffont 1979 and Laffont 1979.

If the seller knew the true value of θ, she would offer a fixed q and charge $T = \theta V(q)$. Her profit would then be $\theta V(q) - cq$, and it would be maximized at q given by $\theta V'(q) = c$. Because the consumer may have one of two types, the seller will want to offer two different bundles if she does not know θ. Let $(\underline{q}, \underline{T})$ denote the bundle intended for the type-$\underline{\theta}$ consumer, and let $(\bar{q}, \bar{T})$ be the bundle intended for the type-$\bar{\theta}$ consumer.[4] The seller's expected profit is thus

$$Eu_0 = \underline{p}(\underline{T} - c\underline{q}) + \bar{p}(\bar{T} - c\bar{q}).$$

The seller faces two kinds of constraints. The first kind requires that the consumer be willing to purchase. (As we noted above, this is without loss of generality, because the seller can always offer the bundle $(q, T) = (0, 0)$ which corresponds to not purchasing in her "menu" of bundles.) Such a constraint is called an *individual-rationality* (IR) or *participation* constraint. The "reservation utility" is the level of net utility obtained by the consumer by not purchasing, which is equal to 0 here. Thus, we require that

$$(\mathrm{IR}_1) \quad \underline{\theta} V(\underline{q}) - \underline{T} \geq 0$$

and

$$(\mathrm{IR}_2) \quad \bar{\theta} V(\bar{q}) - \bar{T} \geq 0.$$

The second kind of constraint requires that the consumer consume the bundle intended for his type. These are known as *incentive-compatibility* (IC) constraints. Thus, we require that

$$(\mathrm{IC}_1) \quad \underline{\theta} V(\underline{q}) - \underline{T} \geq \underline{\theta} V(\bar{q}) - \bar{T}$$

and

$$(\mathrm{IC}_2) \quad \bar{\theta} V(\bar{q}) - \bar{T} \geq \bar{\theta} V(\underline{q}) - \underline{T}.$$

The seller's problem is to choose $\{(\underline{q}, \underline{T}), (\bar{q}, \bar{T})\}$ so as to maximize her expected profit subject to the two IR and the two IC constraints.

A first step in solving this problem is to show that only IR_1 and IC_2 are binding. First, note that if IR_1 and IC_2 are satisfied, then

$$\bar{\theta} V(\bar{q}) - \bar{T} \geq (\bar{\theta} - \underline{\theta}) V(\underline{q}) \geq 0,$$

which reflects the fact that type $\bar{\theta}$ receives more surplus from consumption than type $\underline{\theta}$. Hence, IR_2 is satisfied as well. Furthermore, IR_2 will not be binding unless $\underline{q} = 0$, i.e., unless the seller does not sell to the low-type consumer. In contrast, IR_1 must be binding, i.e., $\underline{T} = \underline{\theta} V(\underline{q})$: If the two IR constraints were not binding, the seller could increase $\underline{T}$ and $\bar{T}$ by the same

4. The results of section 7.2 imply that the seller will not wish to offer several bundles intended for the same consumer.

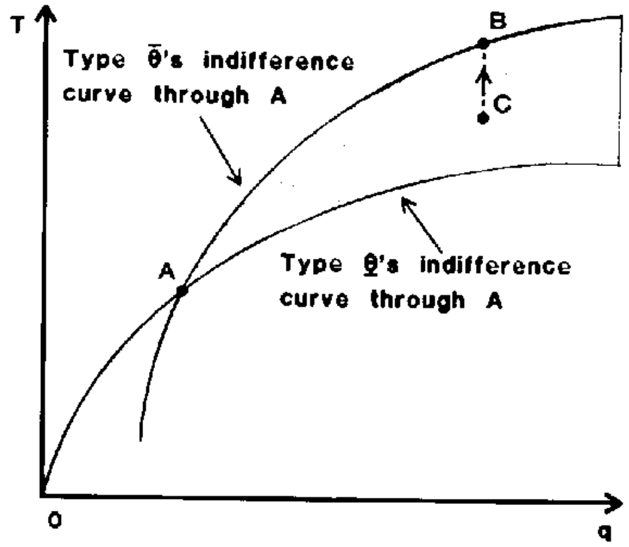


Figure 7.1

small positive amount, which would keep the incentive constraints satisfied, would not violate individual rationality, and would increase revenue.

Next, IC_2 must be binding; that is,

$$\bar{T} = \underline{T} + \bar{\theta}\, V(\bar{q}) - \bar{\theta}\, V(\underline{q}) = \bar{\theta}\, V(\bar{q}) - (\bar{\theta} - \underline{\theta}) V(\underline{q}).$$

If IC_2 were not binding, the seller could increase $\bar{T}$ slightly and keep all constraints satisfied. This is illustrated in figure 7.1. Let A denote the low-type consumer's allocation $(\underline{q}, \underline{T})$, and let B denote the high-type consumer's allocation $(\bar{q}, \bar{T})$. Draw the indifference curves of the low type and the high type through A. Note that, because the slope of the indifference curve of type θ is $\theta V'(q)$, the high-type consumer's indifference curve is always steeper than that of the low-type consumer at any allocation. The allocation B must belong to the shaded area in figure 7.1, because it must be (weakly) preferred by type $\bar{\theta}$ to A, and A must be (weakly) preferred by type $\underline{\theta}$ to B. (Note that this shows that $\bar{q} \geq \underline{q}$, i.e., a high-demand consumer must consume more than a low-demand one. We will analyze this "monotonicity property" at length in this chapter.) The figure also illustrates the fact that A for type $\underline{\theta}$ and C for type $\bar{\theta}$ cannot be optimal for the seller, who could increase her profit by increasing $\bar{T}$ and replacing C by B for type $\bar{\theta}$. Thus, IC_2 must bind.

Knowing that IR_1 and IC_2 are binding, we ignore IC_1 in the derivation of the seller's optimal nonlinear tariff, and solve the subconstrained program with IR_1 and IC_2 only. If the solution to this subconstrained program turns out to satisfy IC_1 as well (as the previous diagrammatic discussion indicates will be the case), then it is a solution to the overall program.

Maximizing Eu_0 subject to IR_1 and IC_2 binding is equivalent to maximizing

$$\underline{p}(\underline{T} - c\underline{q}) + \bar{p}(\bar{T} - c\bar{q}) = [(\underline{p}\underline{\theta} - \bar{p}(\bar{\theta} - \underline{\theta}))V(\underline{q}) - \underline{p}c\underline{q}]$$
$$+ \bar{p}(\bar{\theta}\, V(\bar{q}) - c\bar{q}).$$

The first-order conditions are (assuming $\bar{p}\bar{\theta} < \underline{\theta}$ and $V'(0) = 0$)

$$\underline{\theta}\, V'(\underline{q}) = c \Big/ \left(1 - \frac{\bar{p}(\bar{\theta} - \underline{\theta})}{\underline{p}\underline{\theta}}\right)$$

and

$$\bar{\theta}\, V'(\bar{q}) = c.$$

The quantity purchased by the high-demand consumer is socially optimal (the marginal utility of consumption of the good is equal to the marginal cost). If this were not so, the seller could increase or decrease $\bar{q}$ a little and increase or decrease $\bar{T}$ accordingly so as to keep the utility of the $\bar{\theta}$-type constant. The profit from this type would increase because efficiency would increase, and the new tariff would remain incentive compatible because IC_1 was not binding.

In contrast, the quantity purchased by the low-demand consumer is socially suboptimal (recall that $V'' < 0$). This can be easily understood: The seller lowers the consumption of the low-demand consumer to make it less attractive to the high-demand consumer to "cheat" and consume $\underline{q}$. This allows the seller to increase $\bar{T}$ or, equivalently, to reduce the rent of the high-demand consumer, $(\bar{\theta} - \underline{\theta})V(\underline{q})$. Thus, it is optimal for the seller to sacrifice some efficiency for the purpose of rent extraction. Note also that $\bar{q} > \underline{q}$; furthermore, if IR_1 and IC_2 are satisfied with equality, transfers $\underline{T}$ and $\bar{T}$ are determined by $\underline{q}$ and $\bar{q}$.

Last, we check that IC_1 is satisfied by the solution to the subconstrained program: that is,

$$\underline{\theta}\, V(\underline{q}) - \underline{T} = 0 \geq \underline{\theta}\, V(\bar{q}) - \bar{T}.$$

We find that

$$\underline{\theta}\, V(\bar{q}) - \bar{T} = -(\bar{\theta} - \underline{\theta})[V(\bar{q}) - V(\underline{q})] < 0.$$

Since we saw previously that IR_2 will always be satisfied if IR_1 is, we conclude that the solution to the subconstrained program is a solution to the overall program.

This inefficiency associated with incomplete information will be a constant theme of this chapter. The reader may note the resemblance to the analysis of (nondiscriminatory) monopoly pricing, which corresponds to the special case of our model where all consumers have 0-1 demand, i.e., $V(q) = 0$ for $q < 1$ and $V(q) = 1$ for $q \geq 1$. The two-type case we have considered corresponds to a market where demand d as a function of the transfer T is $d(T) = 1$ for $T \leq \underline{\theta}$, $d(T) = \bar{p}$ for $T \in (\underline{\theta}, \bar{\theta}]$, and $d(T) = 0$ for

$T > \theta$. (If we supposed a continuum of types, we could have a smooth demand function.) With this step-function demand curve, the seller's optimal tariff is either $T > \bar{\theta}$, with profit 0, or $T = \bar{\theta}$, with profit $\bar{p}(\bar{\theta} - c)$, or $T = \underline{\theta}$, with profit equal to $\underline{\theta} - c$. If $\bar{p}(\bar{\theta} - c) > \max(0, \underline{\theta} - c)$, then it is optimal to set $T = \bar{\theta}$, which "separates" the two types. Type $\bar{\theta}$ consumes one unit at price $\bar{\theta}$ and does not enjoy a rent from private information, and type $\underline{\theta}$ consumes 0. If $\underline{\theta} - c > \max(0, \bar{p}(\bar{\theta} - c))$, it is optimal to "bunch" the two types, i.e., have both of them consume one unit and let type $\bar{\theta}$ enjoy rent $\bar{\theta} - \underline{\theta}$. In section 7.3 and in the appendix to this chapter we will derive conditions under which separation or bunching is optimal in mechanism-design problems.[5]

Looking ahead to section 7.2, we can illustrate the revelation principle in this example. The seller indirectly elicits the consumer's information by having him consume a quantity that varies with his type. Alternatively, the seller could maximize her expected profit by asking the consumer to report his type directly. Let

$$\{(q^*(\theta), T^*(\theta))\}_{\theta \in \{\underline{\theta}, \bar{\theta}\}}$$

denote the solution obtained above. The seller can offer the following direct mechanism to the consumer: "Announce your type. If $\hat{\theta}$ is your announcement, you will consume $q^*(\hat{\theta})$ and pay $T^*(\hat{\theta})$." The incentive constraints IC_1 and IC_2 guarantee that it is optimal for the consumer to announce his type truthfully. Thus, the allocation is the same as under the indirect mechanism.

7.1.2 Auctions

A seller has one unit of a good for sale. There are two potential buyers ($i = 1, 2$) with unit demands, and they are *ex ante* identical. Their valuations, θ_1 and θ_2, take value $\underline{\theta}$ with probability $\underline{p}$ and $\bar{\theta}$ with probability $\bar{p}$, where $\underline{p} + \bar{p} = 1$ and θ_1 and θ_2 are independent. Each buyer knows his own valuation, but the seller and the other buyer do not.

One option for the seller is to use the first-price or second-price auctions considered in chapters 1 and 6. But do such auctions maximize the seller's profit? To provide an answer, we solve for the seller's optimal mechanism. As we will see, the familiar auction forms are in fact optimal in some situations.

Suppose that the seller sets up some "message game"—rules for sending and receiving messages—between the buyers, and specifies how the alloca-

5. Section 7.3 solves the more general case of a continuum of types. Approximating the two-type distribution by distributions with a continuum of types, the reader will check that the case of two types does not satisfy the monotone-hazard-rate condition (assumption A10 below), which plays a prominent role throughout the chapter. Bunching (not necessarily of the two "likely types") then occurs.

tion of the good and the transfers will depend on the messages chosen. Let s_1 and s_2 denote the realizations of the two buyers' strategies, σ_1 and σ_2, in this game. The mechanism specifies the probability $x_i(s_1, s_2)$ that the good is transferred to buyer i and the transfer $T_i(s_1, s_2)$ that is paid by buyer i to the seller. For instance, first-price and second-price auctions are mechanisms in which the messages s_i are bids, and bids are made simultaneously. (In both auctions, $x_i(s_1, s_2) = 1$ and $T_j(s_1, s_2) = 0$ if $s_i > s_j$. But when $s_i > s_j$, $T_i = s_i$ in a first-price auction, and $T_i = s_j$ in a second-price auction.)

Let $\{\sigma_1^*(\cdot), \sigma_2^*(\cdot)\}$ denote Bayesian equilibrium strategies in the game or mechanism. Because a buyer is free not to participate in the auction, buyer 1's individual-rationality constraint is that, for each θ_1 and for each s_1 belonging to the support of $\sigma_1^*(\theta_1)$,

$$\text{(IR)} \quad E_{\theta_2} E_{\sigma_2^*(\theta_2)}[\theta_1 x_1(s_1, s_2) - T_1(s_1, s_2)] \geq 0.$$

Similarly, the Bayesian equilibrium or incentive-compatibility condition is that, for each θ_1, each s_1 in the support of $\sigma_1^*(\theta_1)$, and each s_1',

$$\text{(IC)} \quad E_{\theta_2} E_{\sigma_2^*(\theta_2)}[\theta_1 x_1(s_1, s_2) - T_1(s_1, s_2)] \geq E_{\theta_2} E_{\sigma_2^*(\theta_2)}[\theta_1 x_1(s_1', s_2) - T_1(s_1', s_2)].$$

There are similar IR and IC constraints for buyer 2.

The optimal auction for the seller would be hard to define, let alone characterize, if one had to consider all possible message spaces. Fortunately, one can restrict one's attention to "direct-revelation games," in which the two buyers simultaneously make (possibly untruthful) announcements of their types $(\hat{\theta}_1, \hat{\theta}_2)$. To see this, define probabilities of consumption and payments by

$$\tilde{x}_i(\hat{\theta}_1, \hat{\theta}_2) \equiv E_{\{\sigma_1^*(\hat{\theta}_1), \sigma_2^*(\hat{\theta}_2)\}}[x_i(s_1, s_2)]$$

and

$$\tilde{T}_i(\hat{\theta}_1, \hat{\theta}_2) \equiv E_{\{\sigma_1^*(\hat{\theta}_1), \sigma_2^*(\hat{\theta}_2)\}}[T_i(s_1, s_2)].$$

IR and IC ensure that the buyers are willing to participate in the direct-revelation game and that a Bayesian equilibrium of this game is for both buyers to announce the truth $(\hat{\theta}_1 = \theta_1, \hat{\theta}_2 = \theta_2)$.

We will now solve for the optimal symmetric auction. (As we will see in section 7.5, the optimal auction is indeed symmetric.) Note that IR and IC involve only each buyer's expected probability of getting the good and expected payment to the seller, where the expectations are taken with respect to the other buyer's type. So let $\bar{X}, \underline{X}, \bar{T}$, and $\underline{T}$ denote the expected probabilities of getting the good and the expected payments when the buyer has type $\bar{\theta}$ and $\underline{\theta}$, respectively. The individual-rationality and incentive-compatibility constraints can be written as follows:

$$(\mathrm{IR}_1)\quad \underline{\theta}\underline{X} - \underline{T} \geq 0$$

$$(\mathrm{IR}_2)\quad \bar{\theta}\bar{X} - \bar{T} \geq 0$$

$$(\mathrm{IC}_1)\quad \underline{\theta}\underline{X} - \underline{T} \geq \underline{\theta}\bar{X} - \bar{T}$$

$$(\mathrm{IC}_2)\quad \bar{\theta}\bar{X} - \bar{T} \geq \bar{\theta}\underline{X} - \underline{T}.$$

The seller's expected profit per buyer, if his opportunity cost of selling the good is 0, is

$$\mathrm{E}u_0 = (\underline{p}\underline{T} + \bar{p}\bar{T}).$$

We build on the intuition developed in subsection 7.1.1 for the single-buyer case. That is, we guess that the only binding constraints are that the low-valuation type be willing to participate in the mechanism (IR_1) and that the high-valuation type not be tempted to claim a low valuation (IC_2). As in the single-buyer case, the reader can check that the other two constraints do not bind. IR_1 and IC_2 determine the expected payments: $\underline{T} = \underline{\theta}\underline{X}$ and $\bar{T} = \bar{\theta}(\bar{X} - \underline{X}) + \underline{\theta}\underline{X}$. Substituting into the seller's expected profit yields

$$\mathrm{E}u_0 = (\underline{\theta} - \bar{p}\bar{\theta})\underline{X} + \bar{p}\bar{\theta}\bar{X}.$$

Until now, we have not imposed constraints on the probabilities $\underline{X}$ and $\bar{X}$. If there were a single buyer, the constraints would obviously be $0 \leq \underline{X}$, $\bar{X} \leq 1$. With two buyers we must take account of the fact that, if one buyer gets the good, the other buyer does not. At the very least, it must be the case that *ex ante* probability of a player's getting the good (i.e., before knowing one's type) does not exceed $\frac{1}{2}$ (by symmetry):

$$\underline{p}\underline{X} + \bar{p}\bar{X} \leq \tfrac{1}{2}. \qquad (*)$$

As we will see shortly, this constraint does not fully describe the cross-buyer restrictions on probabilities.

First, suppose that $\underline{\theta} \leq \bar{p}\bar{\theta}$. Then $\mathrm{E}u_0$ is decreasing in $\underline{X}$ and increasing in $\bar{X}$. The seller thus wants to set $\underline{X} = 0$ and $\bar{X}$ "as large as possible." By symmetry, $\bar{X}$ cannot exceed $\underline{p} + \bar{p}/2$ because, when both buyers have valuation $\bar{\theta}$, each receives the good with probability $\frac{1}{2}$ (if any receives it). Hence, $\bar{X} = \underline{p} + \bar{p}/2$. The optimal mechanism is then to not sell if both buyers announce $\underline{\theta}$, to sell to the high type if only one buyer announces $\bar{\theta}$, and to sell with probability $\frac{1}{2}$ to each buyer if both buyers announce $\bar{\theta}$. Notice the strong analogy with the one-buyer case where, if $c = 0$ and $\bar{p}\bar{\theta} > \underline{\theta}$, the buyer buys if and only if $\theta = \bar{\theta}$, and enjoys no informational rent.

Second, suppose that $\underline{\theta} > \bar{p}\bar{\theta}$. Then $\mathrm{E}u_0$ is strictly increasing in both $\underline{X}$ and $\bar{X}$, and $(*)$ must be binding. Substituting $\bar{X}$ using $(*)$ in $\mathrm{E}u_0$ yields

$$\mathrm{E}u_0 = \frac{1}{2\underline{p}}(\underline{\theta} - \bar{p}\bar{\theta}) + \frac{\bar{p}}{\underline{p}}(\bar{\theta} - \underline{\theta})\bar{X}.$$

Hence, again, $\bar{X} = \underline{p} + \bar{p}/2$. And, from (*), $\underline{X} = \underline{p}/2$. If only one buyer announces the high valuation, he receives the good; if both buyers announce the high valuation or both announce the low one, each buyer receives the good with probability $\frac{1}{2}$. This completes the derivation of the optimal mechanism.

A famous result in auction theory (Vickrey 1961) is that, under some assumptions, the first- and second-price auctions yield the seller the optimal expected revenue. We will show in section 7.5 that this is the case if the buyers are symmetric, have independent valuations, and have a continuum of potential valuations (instead of two), and if a technical condition is satisfied.[6] An auction is optimal if it has a (symmetric) equilibrium that yields the same expected transfers, $\underline{T}$ and $\bar{T}$, and the same expected probabilities, $\underline{X}$ and $\bar{X}$, that were obtained above. The symmetric equilibrium of a first-price auction (see chapter 6) and the equilibrium of a second-price auction indeed yield the same $\underline{X}$ and $\bar{X}$ as above if $\underline{\theta} \geq \bar{p}\bar{\theta}$, as the good is sold to the highest-valuation buyer. If $\underline{\theta} < \bar{p}\bar{\theta}$, then the same $\underline{X}$ (i.e., 0) is obtained by adding a "reservation price"—$\bar{\theta}$, say—under which all bids are rejected. However, expected transfers need not be the same as in the optimal mechanism.[7] For instance, in the second-price auction, buyers bid their valuations, and the type-$\bar{\theta}$ buyer obtains rent $\underline{p}(\bar{\theta} - \underline{\theta})$, instead of $\underline{p}(\bar{\theta} - \underline{\theta})/2$, the rent that is optimal when $\underline{\theta} \geq \bar{p}\bar{\theta}$. The optimal revenue can be attained in this case by modifying the second-price auction so that, if one buyer bids $\underline{\theta}$ and the other bids $\bar{\theta}$, the high bidder receives the good at price $\underline{\theta} + (\bar{\theta} - \underline{\theta})/2$. Note that it is still an equilibrium for buyers to bid their valuations: If a high-value buyer bids $\bar{\theta}$, his expected profit is

$$\underline{p}(\bar{\theta} - (\underline{\theta} + (\bar{\theta} - \underline{\theta})/2)) = \underline{p}(\bar{\theta} - \underline{\theta})/2,$$

which equals his expected profit from bidding $\underline{\theta}$.

7.2 Mechanism Design and the Revelation Principle††

This section develops the general version of the mechanism-design problem and shows how it can be simplified using the revelation principle.

We suppose that there are $I + 1$ players: a principal (player 0) with no private information, and I agents ($i = 1, \ldots, I$) with types $\theta = (\theta_1, \ldots, \theta_I)$ in

6. The technical condition to be satisfied is that the distribution of buyers' types has a monotone hazard rate (see assumption A10). The continuous approximations to a discrete, two-point distribution do not have a monotone hazard rate.

7. This contrasts with the case of a continuum of types. There, incentive compatibility requires $\theta\dot{X}(\theta) - \dot{T}(\theta) = 0$ (see section 7.5) for all θ. Together with the equilibrium condition that the lowest type, $\underline{\theta}$, gets 0 utility, this implies that if $X(\cdot)$ is optimal, so is $T(\cdot)$.

some set Θ. For the time being, we can allow the probability distribution on Θ to be quite general, requiring only that expectations and conditional expectations of the utility functions be well defined.

The object of the mechanism built by the principal is to determine an *allocation* $y = \{x, t\}$. An allocation consists of a vector x, called a *decision*, belonging to a compact, convex, nonempty $\mathscr{X} \subset \mathbb{R}^n$, and a vector of monetary *transfers* $t = (t_1, \ldots, t_I)$ from the principal to each agent (which can be positive or negative).[8] In most applications $\mathscr{X}$ is taken large enough that we are ensured an interior solution; one exception is the auction example mentioned above.

Player i ($i = 0, 1, \ldots, I$) has a von Neumann-Morgenstern utility $u_i(y, \theta)$. We will assume that u_i ($i = 1, \ldots, I$) is strictly increasing in t_i, that u_0 is decreasing in each t_i, and that these functions are twice continuously differentiable.

Given a (type-contingent) allocation $\{y(\theta)\}_{\theta \in \Theta}$, agent i ($i = 1, \ldots, I$) with type θ_i has expected or "interim" utility

$$U_i(\theta_i) \equiv E_{\theta_{-i}}[u_i(y(\theta_i, \theta_{-i}), \theta_i, \theta_{-i}) \mid \theta_i]$$

and the principal has expected utility

$$E_\theta u_0(y(\theta), \theta).$$

In all applications developed in this chapter, agent i's utility depends on his own transfer t_i and type θ_i, but not on t_{-i} or θ_{-i}. (One situation where u_i depends on θ_{-i} is the common-value auction in which each bidder has private information about the quality of the good for sale.)

The interpretation of x and θ (up to sign adjustments) in the examples mentioned in the introduction is as follows:

Price discrimination x is the consumer's purchase, and t is the price paid to the monopolist; θ indexes the consumer's surplus from consumption.

Regulation x is the firm's cost or price or vector of cost and price, and t is the firm's income; θ is a technological parameter indexing the cost function.

Income tax x is the agent's income, and t is the amount of tax paid by the agent; θ is the agent's ability to earn money.

Public good x is the amount of public good supplied, and t_i is consumer i's monetary contribution to its financing; θ_i indexes consumer i's surplus from the public good.

8. In the price-discrimination and auction examples of section 7.1, the agents transferred money to the principal ($t_i = -T_i$).

Auctions $\mathcal{X}$ is the I-dimensional simplex, i.e., $x_i \geq 0$ for all i, and $\sum_{i=1}^{I} x_i \leq 1$. Here, x_i is the probability that consumer i buys the good, and t_i is the amount paid by consumer i; θ_i indexes consumer i's willingness to pay for the good that is auctioned off.

Bargaining x is the quantity sold by a seller to a buyer; t_1 is the transfer to the seller and t_2 is the (negative) transfer to the buyer, such that $t_1 + t_2 = 0$; $\theta_1 = c$ indexes the seller's cost of producing the good, and $\theta_2 = v$ indexes the buyer's willingness to purchase the good.

A *mechanism* or contract m defines a message space $\mathcal{M}_i$ for each agent i, and a game form ("step 3" in the introduction) to announce the messages, where $\mu = (\mu_1, \ldots, \mu_I)$ is the vector of all messages sent by the agents in the game form. Because types are private information, y can depend on θ only through the agents' messages; denote this function by $y_m: \mathcal{M} \to Y = \mathcal{X} \times \mathbb{R}^I$.

We can now derive the *revelation principle*, which states that the principal can content herself with "direct" mechanisms, in which the message spaces are the type spaces, all agents accept the mechanism in step 2 regardless of their types, and the agents simultaneously and truthfully announce their types in step 3. This principle has been enunciated by many researchers, including Gibbard (1973), Green and Laffont (1977), Dasgupta et al. (1979), and Myerson (1979).

Note that the game form associated with a mechanism in step 3, together with the acceptance decisions of step 2, defines a larger game among the agents. Without loss of generality, we can include the acceptance decision of the agents into their message $\mu_i(\cdot)$. Consider a Bayesian equilibrium of this larger game. Assume for notational simplicity that this is a pure-strategy equilibrium, which we can thus write $\mu_i^*(\theta_i)$.

Consider the new message space Θ_i for each agent i, so that each agent announces a type $\hat{\theta}_i$ (which may differ from the true value θ_i). Letting $\hat{\theta} \equiv (\hat{\theta}_1, \ldots, \hat{\theta}_I)$, define the new allocation rule $\bar{y}: \Theta \to Y$ by

$$\bar{y}(\hat{\theta}) = y_m(\mu^*(\hat{\theta})),$$

where

$$\mu^*(\hat{\theta}) = (\mu_1^*(\hat{\theta}_1), \ldots, \mu_I^*(\hat{\theta}_I)).$$

It is immediate that truthtelling, $\{\hat{\theta}_i = \theta_i\}$, is a Bayesian equilibrium of the new game, given that $\{\mu_i^*\}$ is a Bayesian equilibrium of the original game[9]: For all i and θ_i,

9. We discuss the revelation principle in a Bayesian context, but the same reasoning holds for equilibria in dominant strategies. (See subsection 7.4.2 for the definition of implementation in dominant strategies.)

$$\begin{aligned}
& E_{\theta_{-i}}[u_i(y(\theta), \theta_i, \theta_{-i}) | \theta_i] \\
& = E_{\theta_{-i}}[u_i(y_m(\mu^*(\theta)), \theta_i, \theta_{-i}) | \theta_i] \\
& = \sup_{\mu_i \in \mathscr{M}_i} E_{\theta_{-i}}[u_i(y_m(\mu_1^*(\theta_1), \ldots, \mu_i, \ldots, \mu_I^*(\theta_I)), \theta_i, \theta_{-i}) | \theta_i] \\
& \geq \sup_{\hat{\theta}_i \in \Theta_i} E_{\theta_{-i}}[u_i(\bar{y}(\theta_1 \ldots, \hat{\theta}_i, \ldots, \theta_I), \theta_i, \theta_{-i}) | \theta_i],
\end{aligned}$$

where the first equality results from the definition of the direct-revelation mechanism y, the second equality is the condition for Bayesian equilibrium in the original mechanism m, and the weak inequality expresses the fact that in the direct-revelation mechanism everything is as if agent i picked an announcement in the subset of messages $\{\mu_i^*(\hat{\theta}_i)\}_{\hat{\theta}_i \in \Theta_i}$ of $\mathscr{M}_i$ (the agent thus has, at most, as many possibilities for deviating as in the original game). When the σ_i^* are random, the same reasoning holds with $\bar{y}(\cdot)$ defined as the appropriate random function of $\hat{\theta}$.

Observation (revelation principle) Suppose that a mechanism with message spaces $\mathscr{M}_i$ and allocation function $y_m(\cdot)$ has a Bayesian equilibrium

$$\mu^*(\cdot) \equiv \{\mu_i^*(\theta_i)\}_{\substack{i=1,\ldots,I \\ \theta_i \in \Theta_i}}.$$

Then there exists a direct-revelation mechanism (namely, $\bar{y} = y_m \circ \mu^*$) such that the message spaces are the type spaces ($\mathscr{M}_i = \Theta_i$) and such that there exists a Bayesian equilibrium in which all agents accept the mechanism in step 2 and announce their types truthfully in step 3.

Caveat The direct-revelation game associated with $\bar{y}(\cdot)$ has one equilibrium that yields the same allocation as the original equilibrium. This equilibrium need not be unique. Ma et al. (1988), Mookherjee and Reichelstein (1988), Postlewaite and Schmeidler (1986), and Palfrey and Srivastava (1989) have derived conditions under which a Bayesian allocation can be implemented by a game in the sense of either being achieved by all equilibria of the game or (more strongly) being achieved by the unique equilibrium of the game.[10] The idea, as in Maskin 1977 and in Moore and Repullo 1989, is to use, instead of a direct mechanism, a mechanism where players report information in addition to their type. These additional "nontype" messages turn out to be superfluous in the equilibrium to be implemented, but serve to eliminate other equilibria of the reduced game in which players can only announce their types. The standard methodology is first to derive the principal's optimum, and then, if one is worried about multiple equilibria in the direct-revelation game, to see if the optimal allocation satisfies the sufficient conditions for unique implementation.

10. See also the work of Demski and Sappington (1984) and Ma, Moore, and Turnbull (1988) in more structured environments.

Remark We will be fairly casual about the distinction between a mechanism and an allocation. In a sense, the revelation principle, which we invoke systematically from now on, allows us to merge the two concepts.

7.3 Mechanism Design with a Single Agent††

The following methodology, first developed by Mirrlees (1971), was extended and applied to various contexts by Mussa and Rosen (1978), Baron and Myerson (1982), and Maskin and Riley (1984a), among others. The presentation, including the propositions, follows the general analysis of Guesnerie and Laffont (1984).[11]

Because there is a single agent, we omit the subscripts on transfer (t) and type (θ) in this section. We assume that the agent's type lies in an interval $[\underline{\theta}, \bar{\theta}]$. The agent knows θ, and the principal has the prior cumulative distribution function P ($P(\underline{\theta}) = 0, P(\bar{\theta}) = 1$), with differentiable density $p(\theta)$ such that $p(\theta) > 0$ for all θ in $[\underline{\theta}, \bar{\theta}]$. (Differentiability of the density is not necessary, but is assumed for convenience.) The type space is single dimensional,[12] but the decision space may be multidimensional. (Although we consider a multidimensional decision for completeness, the reader can grasp the main ideas from the case of a single-dimensional decision.) A (type-contingent) allocation is a function from the agent's type into an allocation:

$$\theta \to y(\theta) = (x(\theta), t(\theta)).$$

7.3.1 Implementable Decisions and Allocations

Definition 7.1 A decision function $x: \theta \to \mathscr{X}$ is *implementable* if there exists a transfer function $t(\cdot)$ such that the allocation $y(\theta) = (x(\theta), t(\theta))$ for $\theta \in [\underline{\theta}, \bar{\theta}]$ satisfies the incentive-compatibility constraint

$$\text{(IC)} \quad u_1(y(\theta), \theta) \geq u_1(y(\hat{\theta}), \theta) \text{ for all } (\theta, \hat{\theta}) \in [\underline{\theta}, \bar{\theta}] \times [\underline{\theta}, \bar{\theta}].$$

We will then say that the allocation $y(\cdot)$ is implementable.

Note that we ignore the individual-rationality constraint (that the agent be willing to participate in step 2) in this definition. Such a constraint, if any, must be reintroduced at the optimization stage.

Remark If $x(\cdot)$ is implementable through transfer $t(\cdot)$, there exists an "indirect" or "fiscal" mechanism $t = T(x)$, in which the agent chooses a decision x, rather than an announcement of his type, that implements the same allocation. Consider the following scheme:

11. See also Laffont 1989, chapter 10.
12. The case of a multi-dimensional type space is considerably harder. See Rochet 1985, Laffont, Maskin, and Rochet 1987, and McAfee and McMillan 1988.

$$T(x) \equiv \begin{cases} t & \text{if } \exists\hat{\theta} \text{ such that } t = t(\hat{\theta}) \text{ and } x = x(\hat{\theta}) \\ & \text{(if there exist several such } \hat{\theta}\text{, pick one)} \\ -\infty & \text{otherwise.} \end{cases}$$

Choosing an x is *de facto* equivalent to announcing a $\hat{\theta}$.

We restrict our attention to decision profiles $x(\cdot)$ that are piecewise continuously differentiable ("piecewise C^1").[13] We now derive a necessary condition for $x(\cdot)$ to be implementable.

Theorem 7.1 (necessity) A piecewise C^1 decision function $x(\cdot)$ is implementable only if

$$\sum_{k=1}^{n} \frac{\partial}{\partial\theta}\left(\frac{\partial u_1/\partial x_k}{\partial u_1/\partial t}\right)\frac{dx_k}{d\theta} \geq 0, \tag{7.1}$$

whenever $x = x(\theta)$, $t = t(\theta)$, and x is differentiable at θ.

Proof Type θ chooses an announcement $\hat{\theta}$ so as to maximize $\Phi(\hat{\theta},\theta) \equiv u_1(x(\hat{\theta}), t(\hat{\theta}), \theta)$. Because u_1 is C^2 and x is piecewise C^1, any transfer function t that implements x must be piecewise C^1 as well.[14] Maximizing at a point of differentiability yields a first-order condition and a local second-order condition at the optimum $\hat{\theta} = \theta$:

$$\frac{\partial\Phi}{\partial\hat{\theta}}(\theta,\theta) = 0 \text{ (truth telling or IC)} \tag{7.2}$$

and

$$\frac{\partial^2\Phi}{\partial\hat{\theta}^2}(\theta,\theta) \leq 0. \tag{7.3}$$

(To check that the second derivative in equation 7.3 exists except at a finite number of points, note that $\partial^2\Phi/\partial\hat{\theta}\partial\theta$ exists except at a finite number of points, because $dx/d\theta$ and $dt/d\theta$ do, and use identity 7.2.)

Differentiating equation 7.2 shows that (except perhaps at a finite number of points)

$$\frac{\partial^2\Phi}{\partial\hat{\theta}^2}(\theta,\theta) + \frac{\partial^2\Phi}{\partial\hat{\theta}\partial\theta}(\theta,\theta) = 0. \tag{7.4}$$

Therefore, the local second-order condition can be rewritten as

13. A piecewise-C^1 function admits a derivative except at a finite number of points. And when a derivative does not exist, the function still admits a left and a right derivative. Standard optimal control techniques (Hadley and Kemp 1971) require that functions be piecewise C^1. That the "piecewise" qualifier is needed becomes clear in our analysis of bunching in the appendix to the present chapter.

14. Perform a Taylor expansion of $u_1(x(\hat{\theta}), t(\hat{\theta}), \theta)$ to the left and to the right of $\hat{\theta} = \theta$ at a θ at which $dx/d\theta$ exists and is continuous, and use the fact that $\hat{\theta} = \theta$ is optimal for type θ.

$$\frac{\partial^2 \Phi}{\partial \hat{\theta} \partial \theta}(\theta, \theta) \geq 0, \tag{7.5}$$

or

$$\sum_{k=1}^{n} \frac{\partial}{\partial \theta}\left(\frac{\partial u_1}{\partial x_k}\right)\frac{dx_k}{d\theta} + \frac{\partial}{\partial \theta}\left(\frac{\partial u_1}{\partial t}\right)\frac{dt}{d\theta} \geq 0. \tag{7.6}$$

Rewriting equation 7.2 yields

$$\sum_{k=1}^{n} \frac{\partial u_1}{\partial x_k}\frac{dx_k}{d\theta} + \frac{\partial u_1}{\partial t}\frac{dt}{d\theta} = 0. \tag{7.7}$$

Using equation 7.7 to eliminate $dt/d\theta$ in equation 7.6 yields

$$\sum_{k=1}^{n} \left[\left[\frac{\partial}{\partial \theta}\left(\frac{\partial u_1}{\partial x_k}\right)\frac{\partial u_1}{\partial t} - \frac{\partial}{\partial \theta}\left(\frac{\partial u_1}{\partial t}\right)\frac{\partial u_1}{\partial x_k}\right]\bigg/\frac{\partial u_1}{\partial t}\right]\frac{dx_k}{d\theta} \geq 0, \tag{7.8}$$

which is equivalent to equation 7.1. ■

The interpretation of the necessary condition is particularly simple if we make the following assumption:

A1 For all $k \in \{1, \ldots, n\}$, either

$$(\mathrm{CS}^+) \quad \frac{\partial}{\partial \theta}\left(\frac{\partial u_1/\partial x_k}{\partial u_1/\partial t}\right) > 0$$

or

$$(\mathrm{CS}^-) \quad \frac{\partial}{\partial \theta}\left(\frac{\partial u_1/\partial x_k}{\partial u_1/\partial t}\right) < 0.$$

This is known as the *sorting* (or "constant sign" (CS), or "single crossing," or "Spence-Mirrlees") *condition.*

Note that by changing x_k into $-x_k$ if necessary, one can restrict attention to the case in which all the derivatives are positive if A1 holds. From now on we will assume that CS^+ holds for all k. A1 is very standard and is made in almost all applications of the theory. Note that

$$\frac{\partial u_1/\partial x_k}{\partial u_1/\partial t}$$

is the agent's marginal rate of substitution between decision k and transfer t. The condition asserts that the agent's type affects this marginal rate of substitution in a systematic way.

Suppose for instance that the decision is single dimensional ($n = 1$), and that $\partial u_1/\partial x < 0$, as is the case if x is an output supplied by the agent to the principal. In this case, under the sorting condition, inequality 7.1 is equivalent to the monotonicity of the decision in the agent's type. CS^+ means

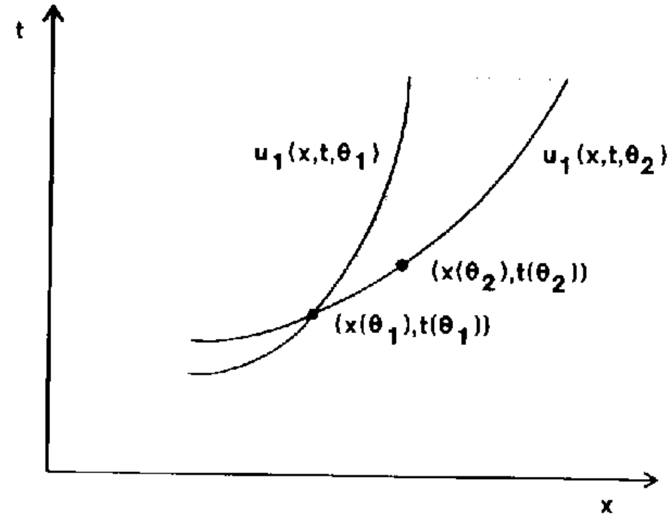


Figure 7.2

that the slope of the agent's indifference curve in the (x, t) space,

$$\left|\frac{\partial u_1/\partial x}{\partial u_1/\partial t}\right|,$$

decreases with the agent's type, θ. That is, a high-type (θ_2) agent must be compensated less than a low-type agent (θ_1) for a given increase in the decision x. This situation is depicted in figure 7.2.

In this case the interpretation of inequality CS^+ is straightforward. Let $y(\theta_1) = (x(\theta_1), t(\theta_1))$ and $y(\theta_2) = (x(\theta_2), t(\theta_2))$ denote the allocations of types θ_1 and θ_2. For the allocation to be incentive compatible, it must be the case that $y(\theta_2)$ lies below type θ_1's indifference curve through $y(\theta_1)$ and above type θ_2's indifference curve through $y(\theta_1)$. Hence, $y(\theta_2)$ must belong to the shaded region in figure 7.2. (We put $y(\theta_2)$ on the boundary of the shaded region in the figure because, as we show below, type θ_2's incentive-compatibility constraint is binding in the optimal mechanism.)

Throughout this chapter, we will make substantial use of the following theorem.

Theorem 7.2 (monotonicity) Assume that the decision space is single dimensional and that CS^+ holds. A necessary condition for $x(\cdot)$ to be implementable is that it be nondecreasing: $\theta_2 > \theta_1 \Rightarrow x(\theta_2) \geq x(\theta_1)$.

Of course, if CS^- held, the necessary condition would be that $x(\cdot)$ be nonincreasing. Note that, whereas theorem 7.1 implies monotonicity at points of differentiability, the proof of theorem 7.2 relies on the simple revealed-preference argument outlined in the discussion of figure 7.2 and has nothing to do with differentiability.

To obtain sufficient conditions for implementability, Guesnerie and Laffont (1984) make the sorting assumption A1 and add the following technical assumption, which guarantees the existence of a solution to a differential equation.

A2 The marginal rates of substitution between decisions and transfer do not increase too fast when the transfer goes to infinity: For all k, there exist K_0 and K_1 such that

$$\left|\frac{\partial u_1/\partial x_k}{\partial u_1/\partial t}\right| \leq K_0 + K_1|t| \text{ uniformly over } x, t, \text{ and } \theta.$$

Assumption A2 is satisfied, for instance, in the case of quasi-linear preferences (for which $\partial u_1/\partial t = 1$).

It turns out that monotonicity is also sufficient for implementability[15]:

Theorem 7.3 Under assumptions A1 (CS^+) and A2, any piecewise C^1 decision function $x(\cdot)$ satisfying $dx_k/d\theta \geq 0$ for all k is implementable. That is, there exists $t(\cdot)$ such that $(x(\cdot), t(\cdot))$ is incentive compatible.

Proof From the agent's first-order condition (equation 7.7), $t(\cdot)$ must satisfy

$$\frac{dt}{d\theta} = -\sum_{k=1}^{n}\left(\frac{\partial u_1/\partial x_k}{\partial u_1/\partial t}\right)\frac{dx_k}{d\theta}. \tag{7.9}$$

Assumption A2 guarantees the existence of a solution to equation 7.9.[16] We are left with showing that $(x(\cdot), t(\cdot))$ is incentive compatible. (By construction, the first-order condition of the agent's maximization with respect to $\hat{\theta}$ is satisfied; so is the local second-order condition in inequality 7.1, from CS^+ and $dx_k/d\theta > 0$. But this is not sufficient, as we must still prove that the global second-order condition for maximization is satisfied.) Suppose that truth telling is not optimal for type θ. That is, there exists $\hat{\theta}$ such that $\Phi(\hat{\theta}, \theta) - \Phi(\theta, \theta) > 0$ (recall that $\Phi(\hat{\theta}, \theta) \equiv u_1(x(\hat{\theta}), t(\hat{\theta}), \theta)$). Then

$$\int_{\theta}^{\hat{\theta}} \frac{\partial \Phi}{\partial a}(a, \theta)\, da > 0,$$

or

15. Thus, the incentive constraints are satisfied globally if they are satisfied locally and if each component of the decision is monotonic. Under some conditions on the utility functions and the distribution of types, the optimal mechanism subject only to the local incentive constraints is monotonic, and thus is optimal subject to the global incentive constraints. We develop this approach below for the case of a one-dimensional decision.

Alternatively, weaker assumptions on the utility functions can be made that, coupled with the hypothesis that all the "downward" incentive constraints are satisfied (that is, $u_1(y(\theta), \theta) \geq u_1(y(\theta'), \theta)$ for all $\theta' \leq \theta$), guarantee that the neglected upward incentive constraints are satisfied and therefore that global incentive compatibility holds. Moreover, one can obtain some properties of the optimal mechanism (without solving explicitly for the optimal mechanism) using the fact that only the downward constraints must be satisfied. This "nonlocal" approach is developed by Moore (1984, 1985, 1988) and by Matthews and Moore (1987).

16. As $|dt/d\theta| \leq (\sup_{\theta,k}|dx_k/d\theta|)(K_0 + K_1|t|)$. See Hurewicz 1958.

$$\int_{\theta}^{\hat{\theta}} \frac{\partial u_1}{\partial t}(x(a), t(a), \theta) \left(\sum_{k=1}^{n} \frac{\frac{\partial u_1}{\partial x_k}(x(a), t(a), \theta)}{\frac{\partial u_1}{\partial t}(x(a), t(a), \theta)} \frac{dx_k}{da}(a) + \frac{dt}{da}(a) \right) da > 0. \tag{7.10}$$

If $\hat{\theta} > \theta$, from the sorting condition CS^+,[17] equation 7.10 implies that

$$\int_{\theta}^{\hat{\theta}} \frac{\partial u_1}{\partial t}(x(a), t(a), \theta) \left(\sum_{k=1}^{n} \frac{\frac{\partial u_1}{\partial x_k}(x(a), t(a), a)}{\frac{\partial u_1}{\partial t}(x(a), t(a), a)} \frac{dx_k}{da}(a) + \frac{dt}{da}(a) \right) da > 0. \tag{7.11}$$

But equation 7.9 implies that the integrand in equation 7.11 is equal to 0 for all a, which is a contradiction.

If $\hat{\theta} < \theta$, the same reasoning shows that equation 7.11 cannot hold either. ∎

An important corollary of theorem 7.3 is that in the case of a single-dimensional decision, under the sorting conditions CS^+ or CS^-, a decision function is implementable if and only if it is monotone (nondecreasing under CS^+, nonincreasing under CS^-).

7.3.2 Optimal Mechanisms

Now that we have characterized the set of implementable allocations, we can determine the optimal one for the principal. To do so, we must re-introduce the individual-rationality constraint for the agent. An implementable allocation that satisfies the individual-rationality constraint is called *feasible*; the principal's problem is to choose the feasible allocation with the highest expected payoff. For simplicity, we assume that the agent's reservation utility (i.e., his expected utility when he rejects the principal's mechanism) is independent of his type.

A3 The reservation utility $\underline{u}$ is independent of type; i.e., the participation constraint is

(IR) $u_1(x(\theta), t(\theta), \theta) \geq \underline{u}$ for all θ.

Under this assumption, if u_1 increases with the type ($\partial u_1/\partial\theta > 0$), then IR can bind only at $\theta = \underline{\theta}$: Any type $\theta > \underline{\theta}$ can always announce $\hat{\theta} = \underline{\theta}$, which

17. That is,

$$\frac{\frac{\partial u_1}{\partial x_k}(x, t, \theta)}{\frac{\partial u_1}{\partial t}(x, t, \theta)} \leq \frac{\frac{\partial u_1}{\partial x_k}(x, t, a)}{\frac{\partial u_1}{\partial t}(x, t, a)} \quad \text{for } \theta \leq a.$$

gives him more than type $\underline{\theta}$'s utility, which is at least $\underline{u}$.[18] For notational simplicity, we normalize $\underline{u} = 0$.

We will also make the following assumptions.

A4 Quasi-linear utilities[19]:

$$u_0(x, t, \theta) = V_0(x, \theta) - t,$$

$$u_1(x, t, \theta) = V_1(x, \theta) + t,$$

where V_0 and V_1 are thrice differentiable and concave in x.

A5 $n = 1$: The decision is single dimensional, and CS$^+$ holds; that is, $\partial^2 V_1/\partial x \partial \theta \geq 0$.

A6 $\partial V_1/\partial \theta > 0$.

A7 $\partial^2 V_0/\partial x \partial \theta \geq 0$ (which is satisfied if V_0 does not depend on θ).

A8 $\partial^3 V_1/\partial x \partial \theta^2 \leq 0$ and $\partial^3 V_1/\partial x^2 \partial \theta \geq 0$.

A9 $\mathscr{X}$ is the interval $[0, \bar{x}]$, where $\bar{x} > \arg\max(V_0(x, \bar{\theta}) + V_1(x, \bar{\theta}))$.

We have little information about whether assumption A8 is likely to be satisfied, as it contains third derivatives. This assumption is a sufficient condition (together with the monotone-hazard-rate condition introduced below) for the optimal decision obtained by ignoring the monotonicity constraint to satisfy monotonicity. As can be seen from equation 7.13 below, assumption A8 is not necessary if "uncertainty about θ" is small, i.e., if the hazard rate is large.[20]

The principal maximizes her expected utility subject to the agent's IR and IC constraints:

$$\max_{\{x(\cdot), t(\cdot)\}} \mathrm{E}_\theta u_0(x(\theta), t(\theta), \theta)$$

18. There are no general results on mechanism design when the reservation utility is increasing with θ. The issue is that the participation constraint may be binding at points other than $\theta = \underline{\theta}$. For economic examples in which it is binding at $\bar{\theta}$, see Champsaur and Rochet 1989, Laffont and Tirole 1990a, and Lewis and Sappington 1989a; for examples in which it is binding in the middle of the interval, see Lewis and Sappington 1989b. For instance, Champsaur and Rochet (1989) and Laffont and Tirole (1990a) study price discrimination by a firm when the high-demand consumers can purchase an alternative (bypass) product (see also exercise 7.8). Similarly, in a labor market, a high-ability worker might have better outside opportunities, and therefore a higher reservation utility, than a low-ability one. Another common cause of type-dependent reservation utilities and of IR constraints that are binding for a good type is the existence of a prior contract between the principal and the agent (Laffont and Tirole 1990b; Caillaud, Jullien and Picard 1990). Even if, *ex ante*, all types have the same reservation utility before signing a contract with the principal, this contract, once signed, defines a *status quo* allocation in any future contract renegotiation. The *status quo* allocation is then type dependent.

19. Although quasi-linearity is a strong assumption, given quasi-linearity the fact that the coefficients of t are -1 and $+1$ is not an additional restriction, as one can always normalize V_0 and V_1 so that the payoff functions can be written as in assumption A4.

20. The intuition is that, if the uncertainty is small, allocations are close to the symmetric-information allocation, and the study of the symmetric-information case requires only assumptions on second derivatives.

subject to $x \in \mathscr{X}$ and

$$\text{(IC)} \quad u_1(x(\theta), t(\theta), \theta) \geq u_1(x(\hat{\theta}), t(\hat{\theta}), \theta) \quad \text{for all } (\theta, \hat{\theta})$$

$$\text{(IR)} \quad u_1(x(\theta), t(\theta), \theta) \geq \underline{u} = 0 \quad \text{for all } \theta.$$

For the moment, we ignore the constraint $x \in \mathscr{X}$. We will return to it at the end of the analysis; in most applications this constraint does not bind. We noted that assumptions A3 and A6 imply that IR need be satisfied only at $\theta = \underline{\theta}$. Furthermore, because transfers are costly to the principal, it is clear that IR is binding at $\theta = \underline{\theta}$:

$$\text{(IR}'\text{)} \quad u_1(x(\underline{\theta}), t(\theta), \theta) = \underline{u} = 0.$$

A useful trick, due to Mirrlees (1971), is to use the indirect utility function. This allows us to eliminate transfers in the above program. Let

$$U_1(\theta) \equiv \max_{\hat{\theta}} u_1(x(\hat{\theta}), t(\hat{\theta}), \theta) = u_1(x(\theta), t(\theta), \theta).$$

The envelope theorem implies that

$$\frac{dU_1}{d\theta} = \frac{\partial u_1}{\partial \theta} = \frac{\partial V_1}{\partial \theta},$$

which implies that

$$U_1(\theta) = u + \int_{\theta}^{\theta} \frac{\partial V_1}{\partial \tilde{\theta}}(x(\tilde{\theta}), \tilde{\theta})\, d\tilde{\theta}.$$

Furthermore, $u_0 = V_0 + V_1 - U_1$; that is, the principal's utility is equal to the social surplus minus the agent's utility. The principal's objective function is thus

$$\int_{\theta}^{\theta} \left[V_0(x(\theta), \theta) + V_1(x(\theta), \theta) - \int_{\underline{\theta}}^{\theta} \frac{\partial V_1}{\partial \tilde{\theta}}(x(\tilde{\theta}), \tilde{\theta})\, d\tilde{\theta} \right] p(\theta)\, d\theta$$

$$= \int_{\theta}^{\theta} \left[V_0(x(\theta), \theta) + V_1(x(\theta), \theta) - \frac{1 - P(\theta)}{p(\theta)} \frac{\partial V_1}{\partial \theta}(x(\theta), \theta) \right] p(\theta)\, d\theta$$

after an integration by parts.

Next, we argue that IC is equivalent to the conjunction of the condition that $dU_1/d\theta = \partial V_1/\partial\theta$ and the condition that $x(\cdot)$ be nondecreasing. Theorem 8.2 shows that IC implies these two conditions, and theorem 8.3 shows that the converse holds.

The principal's optimization program is thus

$$\max_{\{x(\cdot)\}} \int_{\theta}^{\theta} [V_0(x, \theta) + V_1(x, \theta) - \frac{1 - P(\theta)}{p(\theta)} \frac{\partial V_1}{\partial \theta}(x, \theta)] p(\theta)\, d\theta$$

s.t. (monotonicity) $x(\cdot)$ is nondecreasing;

we call this program I. Once the solution $x(\cdot)$ to program I is obtained, we can compute the agent's indirect utility,

$$U_1(\theta) \equiv \int_{\underline{\theta}}^{\theta} \frac{\partial V_1}{\partial \tilde{\theta}}(x(\tilde{\theta}), \tilde{\theta})\, d\tilde{\theta},$$

and the transfer,

$$t(\theta) \equiv U_1(\theta) - V_1(x(\theta), \theta)).$$

For the moment, let us ignore the monotonicity constraint in program I. The relaxed program is called program II. If the solution to program II turns out to be nondecreasing, then it is also a solution to the full program. Otherwise, one must introduce the monotonicity constraint.

The solution to the relaxed program is given by

$$\frac{\partial V_0}{\partial x} + \frac{\partial V_1}{\partial x} = \frac{1 - P(\theta)}{p(\theta)} \frac{\partial^2 V_1}{\partial x \partial \theta}. \tag{7.12}$$

Let $x^*(\cdot)$ denote a solution to equation 7.12. (From A4 and A8, the relaxed program is concave in x, so the second-order condition is satisfied.)

Interpretation of Equation 7.12 The principal faces a tradeoff between maximizing total surplus ($V_0 + V_1$) and appropriating the agent's informational rent (U_1). Consider a type θ. By increasing x over the interval $[\theta, \theta + d\theta]$ by δx, the total surplus is increased by

$$\left(\frac{\partial(V_0 + V_1)}{\partial x}\delta x\right) p(\theta)\, d\theta.$$

However, the rent of type $\theta + d\theta$ is increased by

$$\left[\frac{\partial}{\partial x}\left(\frac{\partial V_1}{\partial \theta}\right)\delta x\right] d\theta,$$

as is the rent of types in $[\theta + d\theta, \bar{\theta}]$ (which have weight $1 - P(\theta)$). At the optimum, the increase in total surplus must be equal to the expected increase in the agent's rent. Note that at $\theta = \bar{\theta}$ rent extraction is not a concern, and so $V_0 + V_1$ is maximized; this result is known as "no distortion at the top."

As a trivial illustration, consider monopoly pricing. Let $x \in [0, 1]$ denote the quantity purchased by a buyer with 0-1 demand. Let $V_0(x, \theta) = -cx$ (where c is marginal cost), and $V_1(x, \theta) = \theta x$ (so θ is the buyer's valuation for the good). The maximand in the principal's optimization program is linear in x, and the "bang-bang" solution is $x = 1$ if $\theta \geq \theta^* > c$, where $\theta^* = c + [1 - P(\theta^*)]/p(\theta^*)$, and $x = 0$ otherwise. This is the same solution that is obtained when the monopolist charges a price π and knows that a fraction $1 - P(\pi)$ of consumers have valuation exceeding π: The program

$\max_\pi (\pi - c)(1 - P(\pi))$ has solution $\pi = \theta^*$. This is perhaps the simplest example of gains from trade not being realized because the principal trades off efficiency against extraction of the agent's rent.

More generally, a simple revealed-preference argument shows that $x^*(\theta)$ is smaller than the level $\hat{x}(\theta)$ that maximizes total surplus ($V_0 + V_1$), which is the level that would prevail if the principal knew the agent's type. To see this, note that, by definition,

$$V_0(\hat{x}, \theta) + V_1(\hat{x}, \theta) \geq V_0(x^*, \theta) + V_1(x^*, \theta)$$

and

$$V_0(x^*, \theta) + V_1(x^*, \theta) - \frac{1 - P}{p}\frac{\partial V_1}{\partial \theta}(x^*, \theta)$$

$$\geq V_0(\hat{x}, \theta) + V_1(\hat{x}, \theta) - \frac{1 - P}{p}\frac{\partial V_1}{\partial \theta}(\hat{x}, \theta).$$

Adding up these two inequalities and using the sorting condition $\partial^2 V_1/\partial x \partial \theta \geq 0$ yields

$$x^*(\theta) \leq \hat{x}(\theta).$$

Inspection of program I suggests the following definition, due to Myerson (1981):

Definition 7.2 The agent's *virtual surplus* is

$$V_1(x, \theta) - \frac{1 - P(\theta)}{p(\theta)}\frac{\partial V_1}{\partial \theta}(x, \theta).$$

Thus, everything is as if the principal maximized total surplus, where the agent's surplus is replaced by his virtual surplus. Note that the principal's virtual surplus is equal to her surplus in the computation of the total virtual surplus. This is due to the fact that the principal has full information about herself.

When is it legitimate to focus on the relaxed program? One can ignore the monotonicity constraint if and only if the $x^*(\cdot)$ defined by equation 7.12 is nondecreasing. Let us assume for simplicity that the objective function in the relaxed program is strictly concave in x. Totally differentiating equation 7.12 yields

$$\left(\frac{\partial^2 V_0}{\partial x^2} + \frac{\partial^2 V_1}{\partial x^2} - \frac{1 - P(\theta)}{p(\theta)}\frac{\partial^3 V_1}{\partial x^2 \partial \theta}\right)\frac{dx^*}{d\theta}$$

$$= \frac{\partial^2 V_1}{\partial x \partial \theta}\left[\frac{d}{d\theta}\left(\frac{1 - P(\theta)}{p(\theta)}\right) - 1\right] - \frac{\partial^2 V_0}{\partial x \partial \theta} + \frac{1 - P(\theta)}{p(\theta)}\frac{\partial^3 V_1}{\partial x \partial \theta^2}. \tag{7.13}$$

Thus, under assumptions A5, A7, and A8, and using the second-order condition, $dx^*/d\theta$ is positive if

$$\frac{d}{d\theta}\left(\frac{1 - P(\theta)}{p(\theta)}\right) \leq 0.$$

Thus, one can legitimately focus on the relaxed program if the following assumption is satisfied:

A10 (monotone hazard rate) $$\frac{d}{d\theta}\left(\frac{p(\theta)}{1 - P(\theta)}\right) \geq 0.$$

To see why this is called the monotone-hazard-rate condition, interpret θ as the lifetime of a machine, and let $Q(\theta) = 1 - P(\theta)$ be the reliability function, which gives the probability that the machine lasts at least until time θ. The conditional probability that the machine fails over the interval $[\theta, \theta + d\theta]$, given that it lasts until time θ, is the "hazard rate"

$$\frac{p(\theta)}{Q(\theta)} = \frac{p(\theta)}{1 - P(\theta)}.$$

The monotone hazard rate thus indicates that the rate of failure increases as the machine grows older.

Since $1 - P(\theta)$ is decreasing in θ, a sufficient condition for the hazard rate to increase is that the density p is increasing. More generally, the monotone-hazard-rate condition is equivalent to the reliability function Q being log-concave (a function Q is log-concave if $\ln Q$ is concave). One can show that if p is log-concave on $[\underline{\theta}, \overline{\theta}]$, then the reliability function Q is log-concave on $[\underline{\theta}, \overline{\theta}]$ (Bagnoli and Berstrom 1989, theorem 2).[21] Applying these results shows that A10 holds (and thus that one can ignore the monotonicity constraint) if P is uniform, normal, logistic, chi-squared, exponential, Laplace, and, under some restrictions on the parameters, Weibull, gamma, or beta. (See Bagnoli and Bergstrom 1989 for a more complete list of distributions whose reliability function is log-concave.)

Finally, we reintroduce the constraint $x \in \mathscr{X}$, which we have ignored to this point. Because

$$x^*(\theta) = \arg\max_x \, [V_0(x, \overline{\theta}) + V_1(x, \overline{\theta})],$$

which is less than $\bar{x}$ by assumption A9, and $x^*(\theta) \leq x^*(\overline{\theta})$, the constraint $x(\theta) \leq x$ does not bind. If $x^*(\theta) < 0$ for some θ, the optimal allocation will have $x^*(\theta) = 0$ for a range of θ's. In monopoly pricing, for example, the monopolist will choose not to sell to all consumers whose willingness to pay is less than the monopoly price.

21. Note that if q is the density of Q, then $q'/q = p'/p$. But a sufficient condition for a strictly monotonic function on an interval $[\underline{\theta}, \overline{\theta}]$ taking value 0 at either $\underline{\theta}$ or $\overline{\theta}$ to be log-concave on this interval is that its derivative is log-concave (Prekova 1973; Bagnoli and Bergstrom 1989, theorem 1). Hence, if $(p'/p)' = (q'/q)' \leq 0$, $(q/Q)' \leq 0$.

Theorem 7.4 Under assumptions A1–A10, the optimal decision $x^*(\theta)$ is given by equation 7.12.

In the appendix to this chapter we analyze what to do if the monotonicity constraint is binding. We also study whether it might be desirable to use stochastic schemes.

7.4 Mechanisms with Several Agents: Feasible Allocations, Budget Balance, and Efficiency††

Now we turn to the case of mechanisms with several agents. We will distinguish between the case of a self-interested principal and that of a "benevolent" principal who maximizes the sum of the agents' welfare. Of course, the distinction is relevant only when the principal optimizes over feasible allocations (section 7.5). We will make the following assumptions in the rest of the chapter:

B1 Types are single dimensional. They are drawn from independent distributions P_i on $[\underline{\theta}_i, \overline{\theta}_i]$ with strictly positive and differentiable densities p_i. The distributions are common knowledge.

B2 (private values) Agent i's preferences depend only on the decision, his own type, and his own transfer: $u_i(x, t_i, \theta_i)$.

B3 Preferences are quasi-linear:

$$u_i(x, t_i, \theta_i) = V_i(x, \theta_i) + t_i, \quad i \in \{1, \ldots, I\}$$

and either

$$u_0(x, t, \theta) = V_0(x, \theta) - \sum_{i=1}^{I} t_i \quad \text{(self-interested principal)}$$

or

$$u_0(x, t, \theta) = \sum_{i=0}^{I} V_i(x, \theta) \quad \text{(benevolent principal),}$$

where $V_0(x, \theta) = B_0(x, \theta) - C_0(x)$, $C_0(x)$ is the principal's monetary cost from decision x (for example, for supplying a public good), and $B_0(x, \theta)$ is nonmonetary (representing, for example, side-benefits of the decision in other markets).

We will say that an allocation $y(\cdot)$ is (*ex post*) *efficient* if $x(\theta) \in \mathscr{X}$ for each θ and

(E) $x(\theta)$ maximizes $\sum_{i=0}^{I} V_i(x, \theta)$ over $\mathscr{X}$, for all θ.

The rest of this section proceeds as follows. Subsection 7.4.1 defines budget balance and explains that it may imply that no efficient allocation is implementable. Subsection 7.4.2 discusses the difference between Bayesian implementation and implementation in dominant strategies. Subsection 7.4.3 discusses the case where the reservation utilities of the agents are so low that efficient allocations can be implemented even when budget balance is imposed. Subsection 7.4.4 derives conditions that imply that any allocation that can be implemented under budget balance must be inefficient. Subsection 7.4.5 shows how this inefficiency can disappear with many agents in an exchange economy. Subsection 7.4.6 shows how the inefficiency actually becomes more severe with more agents when the decision is whether to produce a public good.

We should mention that there are other notions of individual rationality, incentive compatibility, budget balance, and efficiency than the ones defined here (see, e.g., Holmström and Myerson 1983). These concepts can be defined *ex ante* (when the agents have not yet received their information), interim (after the agents have received their private information, but before they report), and *ex post* (after announcements are observed, so all types are publicly known). In order not to confuse the reader, we have defined only the notions that we will use.

7.4.1 Feasibility under Budget Balance

In many mechanism problems with several agents, the "principal" is not allowed to be a net source of funds to the agents. Moreover, the principal must raise enough revenue from the transfers to cover her cost. (In some applications, this cost is identically 0.) This leads us to consider mechanisms that meet the additional constraint of budget balance:

$$\text{(BB)} \quad \sum_{i=1}^{I} t_i(\theta) \leq -C_0(x(\theta)) \text{ for all } \theta.$$ [22]

As in subsection 7.3.2, we say that an allocation $y = (x, t)$ is *feasible* if x is implementable through t and y is individually rational; y is *feasible under budget balance* if it satisfies BB as well.

One theme of this section will be that efficient allocations are typically not feasible under budget balance when there is incomplete information unless the individual-rationality constraints are very weak. (If budget balance is not required, individual-rationality constraints are irrelevant, as the principal can induce the agents to participate by giving them all very large positive transfers, and efficient allocations are usually feasible.) This kind of inefficiency is different from that in the monopoly-pricing example of

22. Either the principal is endowed with limited powers (for instance, a regulator is constitutionally allowed to set up a mechanism, but not to make transfers to the agents), or there is no principal and the analysis aims at characterizing potential outcomes of bargaining among the agents under asymmetric information (see subsections 7.4.4 and 7.5.2).

section 7.1. There, the competitive outcome where price equals the monopolist's cost is both feasible and efficient; the monopolist's optimal mechanism is inefficient because it is designed to maximize the monopolist's profit and not social welfare. In contrast, the inefficiency results of this section pertain to *all* the allocations that are feasible under the budget-balance constraint.

7.4.2 Dominant Strategy vs. Bayesian Mechanisms

This chapter emphasizes Bayesian mechanisms. Another popular concept is that of "dominant-strategy mechanisms," which are mechanisms where each agent's optimal announcement is independent of the announcements of the other agents. (Note that the two solution concepts are equivalent in the single-agent case.) Since the optimal announcement can be taken to be the truth from the revelation principle, the formal definition of a dominant-strategy mechanism is a function $y(\theta)$ such that, for each agent $i = 1, \ldots, I$ and for each θ_i, $\hat{\theta}_i$, and θ_{-i},

$$\text{(DIC)} \quad u_i(y(\theta_i, \theta_{-i}), \theta_i) \geq u_i(y(\hat{\theta}_i, \theta_{-i}), \theta_i).$$

That is, each agent is induced to tell the truth whatever the other agents' reports (or types—this is equivalent). The incentive-compatibility constraint (DIC) for dominant-strategy implementation is much more stringent than the incentive constraint under Bayesian implementation. In the latter, incentive compatibility is required to hold only *on average* over types θ_{-i}, where the expectation is taken over agent i's beliefs about θ_{-i} conditional on his type θ_i. Bayesian incentive compatibility thus pools the incentive constraints of dominant-strategy incentive compatibility. Also, the Bayesian conditions for each player suppose that all other players report truthfully. Thus, the Bayesian incentive-compatibility constraint is

$$\text{(IC)} \quad \mathrm{E}_{\theta_{-i}} u_i(y(\theta_i, \theta_{-i}), \theta_i) \geq \mathrm{E}_{\theta_{-i}} u_i(y(\hat{\theta}_i, \theta_{-i}), \theta_i).$$

When possible, a principal might prefer dominant-strategy implementation of her mechanism, because it is not sensitive to beliefs that players have about each other and it does not require players to compute Bayesian equilibrium strategies. However, focusing on dominant-strategy mechanisms restricts the set of mechanisms considerably. Thus, implementation in dominant strategies is a nice property to have if feasible, but it is not clear how much utility loss a principal should be willing to tolerate in order to have dominant strategies for the agents.

Mookherjee and Reichelstein (1989) identify a class of models in which dominant-strategy implementation involves no welfare loss relative to Bayesian implementation.[23] Suppose that the agents have quasi-linear

23. Their result generalizes a similar observation made by Laffont and Tirole (1987a) in the context of auctions of an incentive contract among firms.

preferences, and that, for $i = 1, \ldots, I$,

$$u_i(x, t, \theta) = V_i(x, \theta_i) + t_i,$$

where t_i is the principal's transfer to agent i. Mookherjee and Reichelstein allow x to be multi-dimensional, but require that V_i depend on x only through a one-dimensional statistic $h_i(x)$:

$$u_i(x, t, \theta) = V_i(h_i(x), \theta_i) + t_i.$$

They further assume that types are drawn independently, that the distribution $P_i(\cdot)$ of player i's type satisfies the monotone-hazard-rate condition ($p_i/(1 - P_i)$ nondecreasing) for each i, and that preferences satisfy the sorting assumption $\partial V_i/\partial\theta_i\partial h_i \geq 0$ and the condition that $\partial^2 V_i/\partial h_i \partial\theta_i$ is decreasing in θ_i. Under these assumptions, they show that an allocation that maximizes the principal's expected utility,

$$\mathrm{E}_\theta\left(V_0(x, \theta) - \sum_{i=1}^{I} t_i(\theta)\right),$$

subject to the constraints of Bayesian incentive compatibility (IC) and individual rationality,

$$\text{(IR)} \quad \mathrm{E}_{\theta_{-i}} u_i(y(\theta_i, \theta_{-i}), \theta_i) \geq 0 \text{ for all } \theta_i,$$

can be implemented in dominant strategies. That is, one can choose the transfer function in the above program such that DIC, and not only IC, is satisfied.

Though we will mention results about dominant-strategy implementation, we will take the incentive-compatibility and individual-rationality constraints facing the principal to be the *Bayesian* ones, IC and IR, unless we specify otherwise.

7.4.3 Efficiency Theorems

There are two basic results about the implementability of efficient allocations, both of which suppose that the agents' reservation utilities are arbitrarily low.

The Groves Mechanism

An early implementability result was the discovery by Groves (1973) and Clarke (1971) that any efficient provision of public goods can be implemented as long as budget balance is not required. Even more striking is the fact that efficient provision can be implemented in dominant strategies.

The idea is straightforward: Choose agent i's transfer so that agent i's payoff is the same as the total surplus of all parties up to a constant. Because agent i already internalizes his own surplus, it suffices to set the transfer equal to the total surplus minus his surplus. The transfers are thus "externality payments."

$$\sum_{i=1}^{I} t_i(\hat{\theta}) = 0.$$

Let

$$\mathscr{E}_i(\hat{\theta}_i) \equiv \mathrm{E}_{\theta_{-i}}\left(\sum_{j \neq i} V_j(x^*(\hat{\theta}_i, \theta_{-i}), \theta_j)\right)$$

denote the "expected externality" for agent i when he announces $\hat{\theta}_i$. $\mathscr{E}_i(\hat{\theta}_i)$ is the first part of the transfer to agent i; because $\tau_i(\cdot)$ is supposed not to depend on $\hat{\theta}_i$, $\mathscr{E}_i(\hat{\theta}_i)$ must be paid by the other agents. One can, for instance, have them share the payment, i.e., allocate $\mathscr{E}_i(\hat{\theta}_i)/(I-1)$ to each $\tau_j(\cdot)$, $j \neq i$. Thus, the following functions ensure budget balance[26]:

$$\begin{aligned}\tau_i(\hat{\theta}_{-i}) &= -\sum_{j \neq i} \mathscr{E}_j(\hat{\theta}_j) \Big/ (I-1) \\ &= -\frac{1}{I-1} \sum_{j \neq i} \mathrm{E}_{\theta_{-j}}\left(\sum_{k \neq j} V_k(x^*(\hat{\theta}_j, \theta_{-j}), \theta_k)\right). \end{aligned} \tag{7.18}$$

Now suppose that the principal incurs a cost $C_0(x)$ from any decision $x \neq 0$, so that budget balance requires

$$\sum_{i=1}^{I} t_i(\hat{\theta}) \leq -C_0(x(\hat{\theta})).$$

To implement the efficient decision under this constraint, we consider the "fictional problem" where the agents' utility functions are

$$\tilde{V}_i(x, \theta_i) \equiv V_i(x, \theta_i) - C_0(x)/I$$

and the principal's cost is $\tilde{C}_0(x) \equiv 0$. We then compute the transfers $\tilde{t}_i(\cdot)$ for this fictional problem, and set

$$t_i(\cdot) = \tilde{t}_i(\cdot) - C_0(x^*(\cdot))/I.$$

We claim that these transfers implement the efficient decision with budget balance in the original problem. Budget balance is trivial: Since $\sum_{i=1}^{I} \tilde{t}_i(\hat{\theta}) = 0$ for all $\hat{\theta}$,

$$\sum_{i=1}^{I} t_i(\hat{\theta}) = -C_0(x^*(\hat{\theta})).$$

As for incentive compatibility, note that

$$\tilde{V}_i(x^*(\hat{\theta}), \theta_i) + \tilde{t}_i(\hat{\theta}) = V_i(x^*(\hat{\theta}), \theta_i) + t_i(\hat{\theta})$$

26. Crémer and Riordan (1985) show that one can strengthen the AGV result by having a dominant strategy for $I-1$ agents. The first mover maximizes his expected payoff by announcing his true type, and it is a dominant strategy for the $I-1$ "Stackelberg followers" to announce truthfully also.

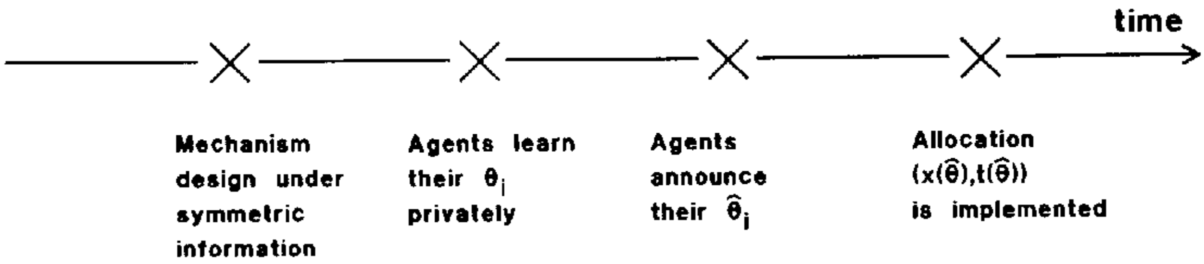


Figure 7.3

for every $\hat{\theta}$ and θ_i; thus, if reporting truthfully is an equilibrium in the fictional problem under transfers $\tilde{t}$, it is an equilibrium in the original problem under transfers t.

An implication of the AGV result is the following: Suppose that I agents meet and agree on a mechanism $(x(\cdot), t(\cdot))$ *ex ante*, i.e., before they receive their private information. The timing is then as shown in figure 7.3. We claim that if $x^*(\cdot)$ is the efficient-decision rule, the contract signed among the I agents *ex ante* will yield $x^*(\theta)$ for each realization of θ, even if the agents can refuse to sign the contract (i.e., $\underline{u}$ is not unboundedly low). Clearly, if $x^*(\cdot)$ is implementable, any mechanism $(x^*(\cdot), t(\cdot))$ that implements it is optimal, as it maximizes the "pie" to be divided among the agents, and agents have an incentive to maximize the pie and distribute it among themselves, possibly with *ex ante* compensatory transfers (recall that the contract is signed under symmetric information, so one expects *ex ante* gains from trade to be realized). But to implement $x^*(\cdot)$, it suffices to give *ex post* transfers $t_i(\cdot)$ specified by the AGV formulas 7.17 and 7.18.

7.4.4 Inefficiency Theorems

We saw in subsection 7.4.3 that, with quasi-linear preferences, efficiency is attainable if budget balance is not required or agents can be forced to participate. In contrast, inefficiency "tends to occur" when the principal must provide the agents with an exogenous reservation utility level and budget balance is required. There are two basic inefficiency results, one due to Laffont and Maskin (1979) and the other to Myerson and Satterthwaite (1983).[27] In this subsection, we first develop the Myerson-Satterthwaite analysis and then sketch that of Laffont and Maskin.

Myerson and Satterthwaite consider a two-agent trading game. The seller can supply one unit of a good at cost c drawn from distribution $P_1(\cdot)$ with differentiable, strictly positive density $p_1(\cdot)$ on $[\underline{c}, \bar{c}]$. The buyer has unit demand and valuation v drawn from distribution $P_2(\cdot)$ on $[\underline{v}, \bar{v}]$ with

27. As Eric Maskin pointed out to us, the inefficiency result can with hindsight be seen in Mirrlees' (1971) treatment of the optimal-taxation problem. Suppose that the social planner's objective is, as proposed by Rawls, to maximize the minimum utility in society. One can consider the minimum utility as a reservation utility and look for incentive-compatible, balanced-budget allocations. Mirrlees shows that there exists no such allocation that is efficient.

differentiable, strictly positive density $p_2(\cdot)$. Let $x(c, v) \in [0, 1]$ denote the probability of trade and $t(c, v)$ denote the transfer from the buyer to the seller (so $t_1 \equiv t$, $t_2 \equiv -t$, and $t_1 + t_2 = 0$). We do not specify how the two players end up with the type-contingent allocation $\{x(\cdot), t(\cdot)\}$. For instance, they might bargain as in the Chatterjee-Samuelson model described in chapter 6, or they could use a more complex sequential-bargaining process such as the ones described in chapter 10, or they might respond to a mechanism designed by a principal (see subsection 7.5.2 for more on this). Rather, the question is whether efficiency is consistent with equilibrium strategies (i.e., IC), individual rationality, and budget balance in general games.

Let

$$X_1(c) \equiv \mathrm{E}_v[x(c, v)]$$

and

$$X_2(v) \equiv \mathrm{E}_c[x(c, v)]$$

denote, respectively, the seller's and the buyer's probabilities of trading as a function of their type; let

$$T_1(c) \equiv \mathrm{E}_v[t(c, v)]$$

and

$$T_2(v) \equiv -\mathrm{E}_c[t(c, v)]$$

denote their expected transfers; and let

$$U_1(c) \equiv T_1(c) - c\, X_1(c)$$

and

$$U_2(v) \equiv v\, X_2(v) + T_2(v)$$

denote their expected utilities when they have types c and v, respectively. Because the sorting condition (see subsection 7.3.1) is satisfied, we conclude from theorem 7.2 that X_1 and X_2 must be monotonic if incentive compatibility obtains: X_1 is nonincreasing and X_2 is nondecreasing. Also, from section 7.3, we have

$$U_1(c) = U_1(\bar{c}) + \int_c^{\bar{c}} X_1(\gamma)\, d\gamma \tag{7.19}$$

and

$$U_2(v) = U_2(\underline{v}) + \int_{\underline{v}}^{v} X_2(\nu)\, d\nu. \tag{7.20}$$

Substituting for $U_1(c)$ and $U_2(v)$ and adding up equations 7.19 and 7.20 yields

$$T_1(c) + T_2(v) = c\,X_1(c) - v\,X_2(v) + U_1(c) + U_2(v) + \int_c^{\bar{c}} X_1(\gamma)\,d\gamma + \int_{\underline{v}}^{v} X_2(\nu)\,d\nu. \tag{7.21}$$

But budget balance ($t_1(c,v) + t_2(c,v) = 0$) in particular implies that

$$\mathrm{E}_c T_1(c) + \mathrm{E}_v T_2(v) = 0.$$

Or, using equation 7.21,

$$0 = \int_{\underline{c}}^{\bar{c}} \left(c\,X_1(c) + \int_c^{\bar{c}} X_1(\gamma)\,d\gamma \right) p_1(c)\,dc + U_1(\bar{c}) + \int_{\underline{v}}^{\bar{v}} \left(\int_{\underline{v}}^{v} X_2(\nu)\,d\nu - v\,X_2(v) \right) p_2(v)\,dv + U_2(\underline{v}). \tag{7.22}$$

Integrating by parts in equation 7.22 yields

$$U_1(\bar{c}) + U_2(\underline{v}) = -\int_{\underline{c}}^{\bar{c}} \left(c + \frac{P_1(c)}{p_1(c)} \right) X_1(c) p_1(c)\,dc + \int_{\underline{v}}^{\bar{v}} \left(v - \frac{1 - P_2(v)}{p_2(v)} \right) X_2(v)\,p_2(v)\,dv, \tag{7.23}$$

so that, by replacing X_1 and X_2 by their definitions, we get

$$U_1(\bar{c}) + U_2(\underline{v}) = \int_{\underline{c}}^{\bar{c}} \int_{\underline{v}}^{\bar{v}} \left[\left(v - \frac{1 - P_2(v)}{p_2(v)} \right) - \left(c + \frac{P_1(c)}{p_1(c)} \right) \right] x(c,v) p_1(c) p_2(v)\,dc\,dv. \tag{7.24}$$

Since individual rationality is equivalent to $U_1(\bar{c}) \geq 0$ and $U_2(\underline{v}) \geq 0$, a necessary condition for $x(\cdot)$ to be implementable is that the right-hand side of equation 7.24 be nonnegative.

Now, efficiency requires that $x(\cdot) = x^*(\cdot)$, where $x^*(c,v) = 1$ if $v \geq c$ and $= 0$ otherwise. One can check that equation 7.24 is not satisfied for $x(\cdot) = x^*(\cdot)$ if $\bar{c} > \underline{v}$ and $\underline{c} < \bar{v}$, which establishes the following result.

Theorem 7.5 (Myerson and Satterthwaite 1983) Suppose that the seller's cost and the buyer's valuation have differentiable, strictly positive densities on $[\underline{c}, \bar{c}]$ and $[\underline{v}, \bar{v}]$, that there is a positive probability of gains from trade ($c < v$), and that there is a positive probability of no gains from trade ($c > v$). Then there is no efficient trading outcome that satisfies individual rationality, incentive compatibility, and budget balance.[28]

28. The hypothesis that the distributions are represented by strictly positive densities is important. To see this, consider the following discrete example: $v = \underline{v}$ with probability $\underline{p}$ and $= \bar{v}$ with probability $\bar{p}$; $c = \underline{c}$ with probability $\underline{q}$ and $= \bar{c}$ with probability $\bar{q}$, where $\underline{p} + \bar{p} = \underline{q} + \bar{q} = 1$, $\underline{c} < \underline{v} < \bar{c} < \bar{v}$ and $\underline{v} - \underline{c} > \bar{p}(\bar{v} - \underline{c})$. And consider the following bargaining scheme, in which the seller makes a "take it or leave it" offer, which the buyer accepts

Equation 7.24 exhibits the now familiar virtual surpluses:

$$\left(v - \frac{1 - P_2(v)}{p_2(v)}\right)x$$

for the buyer and

$$-\left(c + \frac{P_1(c)}{p_1(c)}\right)x$$

for the seller.[29,30] Furthermore, as in section 7.3, incentive costs must be taken into account when evaluating gains from trade. This explains the inefficiency result. For example, take two types c and v such that $v = c + \varepsilon$ where ε is "small" (two such types exist as long as $\overline{v} > \overline{c} > \underline{v}$). While the buyer's valuation exceeds the seller's cost, the buyer's virtual valuation is lower than the seller's virtual cost, so that there are no "implementable gains from trade."

Note that the inefficiency result is as tight as possible. When it is common knowledge that there are gains from trade ($\underline{v} \geq \overline{c}$), there exist efficient mechanisms that satisfy IR, IC, and BB: "$x(\hat{c}, \hat{v}) = 1$ and $t(\hat{c}, \hat{v}) = t$ for all $(\hat{c}, \hat{v})$ where $\overline{c} \leq t \leq \underline{v}$."

Cramton, Gibbons, and Klemperer (1987) extend the work of Myerson and Satterthwaite (1983) by allowing arbitrary ownership patterns and more than two agents. In the seller-buyer example, the initial ownership pattern is $(\alpha_1 = 1, \alpha_2 = 0)$, where α_i is player i's share of the good; the bargaining is about transforming the ownership structure to $(\alpha_1' = 0, \alpha_2' = 1)$. More generally, suppose that there are I agents who initially hold shares $(\alpha_1, \ldots, \alpha_I)$ of a good, with $\sum_{i=1}^{I} \alpha_i = 1$. Suppose that the final shares are $(\alpha_1', \ldots, \alpha_I')$ with $\sum_{i=1}^{I} \alpha_i' = 1$, and that agent i's surplus is $V_i(\alpha_i, \theta_i) = \alpha_i \theta_i$, where the θ_i are independently drawn from some symmetric distribution $P(\cdot)$ on $[\underline{\theta}, \overline{\theta}]$. Cramton et al. show that if the initial shares are fairly evenly distributed (close to $(1/I, \ldots, 1/I)$), there exist efficient mechanisms that satisfy IC, IR, and BB.

Laffont and Maskin (1979, section 6) obtain an inefficiency result in a framework more general than that of Myerson and Satterthwaite: The decision variable x need not be binary but can take values in $\mathbb{R}^n$. Agents have quasi-linear utilities $u_i = V_i(x, \theta_i) + t_i$. Laffont and Maskin assume (i) that the efficient solution $x^*(\theta)$ that maximizes $\sum_{i=1}^{I} V_i(x, \theta_i)$ is continuously differentiable in θ and (ii) that the optimal expected transfers $t_i(\theta_i)$ are differentiable. Assumption i, although it does not allow for the discontin-

or rejects. Clearly, the $\overline{c}$-seller offers price $\overline{v}$ and sells if and only if $v = \overline{v}$. The $\underline{c}$-seller offers price $\underline{v}$ as $\underline{v} - \underline{c} > \overline{p}(\overline{v} - \underline{c})$, and always sells. The bargaining outcome, which satisfies IR, BB, and IC, is efficient.

29. $v - (1 - P_2)/p_2$ and $c + P_1/p_1$ can be called virtual valuation and virtual cost, respectively.

30. The relevant hazard rate for the seller is P_1/p_1 rather than $(1 - P_1)/p_1$. This comes from the fact that the seller dislikes, rather than likes, higher decisions.

uous x^* considered by Myerson and Satterthwaite (which can only be approximated by continuously differentiable x), is natural and of little concern. Assumption ii, which involves endogenous variables, seems more controversial. However, in most applications, incentive compatibility requires that t_i be monotonic: Decisions concerning the agents—e.g., the expected probability of trade, $X_i(\theta_i)$—can be shown to be monotonic, and t_i must be monotonic if X_i is (for instance, purchasing more at a lower price would not be incentive compatible). But a monotonic function is differentiable almost everywhere, and continuouly differentiable functions can be approximated by almost-everywhere-differentiable ones. Hence, Laffont and Maskin's assumption on differentiable transfers is satisfied in many applications of interest.[31]

7.4.5 Efficiency Limit Theorems†††

The Myerson-Satterthwaite result shows that a buyer and a seller are unable to exhaust gains from trade if they have incomplete information about each other and there is positive probability that there are no gains from trade. This strengthens our earlier observation that the Coase theorem may not extend to asymmetric-information bargaining. One would want to know whether inefficiency remains substantial when there are many buyers and many sellers. In particular, one would expect that, with a large number of traders, any one trader would be unable to have much influence on his terms of trade by misrepresenting his preferences, and therefore allocations that approximate Walrasian equilibria or Pareto optima could be implemented despite asymmetric information.

Confirming this intuition with a continuum of buyers and sellers is straightforward. Suppose for instance that each seller has one unit of the good for sale and has opportunity or production cost c drawn (independent from those of the other sellers and buyers) from the distribution P_1 on $[\underline{c}, \bar{c}]$; similarly, suppose that the buyers have unit demands and their valuations are drawn independently from the distribution P_2 on $[\underline{v}, \bar{v}]$.

31. There have been other extensions of the inefficiency result. Spier (1989) considers bargaining between a plaintiff and a defendant where both have private information about the outcome of the case if they go to court (so, unlike Myerson and Satterthwaite's model, this is a model of "common values"; that is, each agent cares directly about the other agent's information). Going to court involves a judicial cost for both parties. Settling out of court is the Pareto-superior outcome, as it avoids the judicial costs. Thus, it is common knowledge that there are gains from trade (i.e., gains from agreeing). Yet, Spier shows that if the judicial costs are small (but positive), efficiency is inconsistent with IR, IC, and BB. (The intuition is that efficiency requires that the probability of going to court be equal to 0 and, therefore, that all types of defendants pay and all types of plaintiffs receive the same monetary transfer. But if the judicial costs are small, it pays the plaintiff or the defendant to go to court when they have information that is very favorable to them.) Ledyard and Palfrey (1989) consider the case of a public-good mechanism which is designed by the principal. They show that the principal may choose a mechanism that does not maximize the sum of the agents' willingness to pay for the public good even if she faces no IR constraint, as long as the agents' private information is the marginal utility of income ($u_i \equiv x - t/\theta_i$) and the principal cares about income distribution (i.e., about $\sum_{i=1}^{I} u_i$).

With $c > v$ (not everyone ought to trade), the market-clearing price π is given by $P_1(\pi) = 1 - P_2(\pi)$ (if sellers and buyers are in equal numbers). Let $x_1 \in [0, 1]$ and $x_2 \in [0, 1]$ denote a seller's probability of selling and a buyer's probability of purchasing, respectively. A social planner can obtain the efficient outcome by offering the Walrasian mechanism: "$x_1(\hat{c}) = 1$ and $t_1(\hat{c}) = \pi$ if $\hat{c} \leq \pi$, and $x_1(\hat{c}) = t_1(\hat{c}) = 0$ otherwise; $x_2(\hat{v}) = 1$ and $t_2(\hat{v}) = -\pi$ if $\hat{v} > \pi$, and $x_2(\hat{v}) = t_2(\hat{v}) = 0$ otherwise."

With a large but finite number of traders, one cannot generally obtain efficient outcomes under IR. Indeed, Hurwicz (1972) shows that, in general economies, any mechanism that asks the traders to announce their preference orderings, that is efficient, and that satisfies the individual-rationality constraint that the traders prefer their assigned consumption bundle to their initial endowment vector must violate incentive compatibility for some preference orderings. Hurwicz's informational assumption is that traders know one another's preferences (this is called a "Nash environment" in the literature, to distinguish it from "Bayesian environments," where preferences are private information). Roberts and Postlewaite (1976) pursued this line of research and showed that, under some regularity conditions, a trader's gain in utility from distorting his announcement of preferences is bounded above by a number that tends to zero as the number of traders tends to infinity.

Wilson (1985) and Gresik and Satterthwaite (1989) (see also Cramton et al. 1987) perform a similar analysis in a Bayesian context. Suppose with Wilson that there are I_1 sellers, $i = 1, \ldots, I_1$, and I_2 buyers, $i = 1, \ldots, I_2$; that the sellers' costs and the buyers' valuations are drawn independently from distributions on $[c, c]$ and $[\underline{v}, \overline{v}]$; and that $\underline{c} \leq \underline{v} < \overline{c} \leq \overline{v}$, so that $v > c$ and $v < c$ have positive probability.

Wilson studies "double auctions," in which the sellers and the buyers make bids $\{\hat{c}_i\}_{i=1,\ldots,I_1}$ and $\{\hat{v}_i\}_{i=1,\ldots,I_2}$, respectively (bids are similar to announcements of costs or valuations). Without loss of generality, we can reorder the bids so that

$$\hat{c}_{I_1} \geq \hat{c}_{I_1 - 1} \geq \cdots \geq \hat{c}_1$$

and

$$\hat{v}_1 \geq \hat{v}_2 \geq \cdots \geq \hat{v}_{I_2}.$$

Then the number of units traded in a double auction is the largest k such that $\hat{v}_k \geq \hat{c}_k$, and those who trade are sellers 1 through k and buyers 1 through k. The transfer price π is an arbitrary price in $[\hat{c}_k, \hat{v}_k]$ (for instance, $(\hat{v}_k + \hat{c}_k)/2$). The other sellers and buyers do not trade and do not give or receive transfers. Note that if each player's bid equals his type, a double auction maximizes social surplus. Of course, traders have an incentive to misrepresent their preferences, and the equilibrium need not be efficient. Yet Wilson shows that, under some assumptions (existence of an equilibrium in symmetric strategies that are differentiable functions of private information and have uniformly bounded derivatives), a double auction (a very simple mechanism indeed) yields efficiency in the limit when I_1 and I_2 tend to infinity. Gresik and Satterthwaite (1989) provide results on the rate of convergence to Walrasian equilibria.

7.4.6 Strong Inefficiency Limit Theorems†††

The efficiency limit theorems for private goods mentioned in the previous subsection are in stark contrast with limit results by Rob (1989) and Mailath and Postlewaite (1990) for public goods when each agent has veto power. In the private-good case with a large number of traders, a trader has little influence on the price at which he trades, so he has little incentive to manipulate the announcement of his preferences to trade off more favorable prices against a lower probability of trading. The reverse holds for public goods with a large number of traders. A trader has a low probability of being pivotal, i.e., of influencing the decision of whether to produce the public good. Hence, the probability of "trade"—the probability of the public good being supplied—cannot be affected, but under some conditions (to be described) each agent can manipulate his "terms of trade" the amount of his contribution toward the provision of the public good.

Consider a fixed-sized public-good project with I agents. Agent i ($i = 1, \ldots, I$) has utility $u_i = \theta_i x + t_i$ where $x = 1$ if the public good is supplied and $x = 0$ otherwise (t_i is likely to be negative). Let the parameters θ_i be independently drawn from distributions P_i with positive density p_i on $[\underline{\theta}_i, \bar{\theta}_i]$. Assume further that the cost of realizing the project is a function $C(I)$ of the number of agents.

Let us look for mechanisms $m = \{x, t\}$ that satisfy the following properties:

$$x(\hat{\theta}) \in [0,1] \text{ for all } \hat{\theta},$$

$$\text{(IC)} \quad \mathrm{E}_{\theta_{-i}}[x(\theta_i, \theta_{-i})\theta_i + t_i(\theta_i, \theta_{-i})] \geq \mathrm{E}_{\theta_{-i}}[x(\hat{\theta}_i, \theta_{-i})\theta_i + t_i(\hat{\theta}_i, \theta_{-i})] \quad \text{for all } (i, \theta_i, \hat{\theta}_i),$$

$$\text{(IR)} \quad \mathrm{E}_{\theta_{-i}}[x(\theta_i, \theta_{-i})\theta_i + t_i(\theta_i, \theta_{-i})] \geq 0 \text{ for all } (i, \theta_i),$$

$$\text{(BB)} \quad \sum_{i=1}^{I} t_i(\theta) + x(\theta)C(I) \leq 0 \text{ for all } \theta.$$

The Rob-Mailath-Postlewaite result is that, in the limit with a large number of traders, IC, IR, and BB imply that *no* gains from trade are realized if $C(I)$ is proportional to I and $C(I)/I > \underline{\theta}_i$ for all i.

Actually, an apparently stronger result will be proved, by replacing BB (which is an *ex post* concept) by *ex ante* budget balance:

$$\text{(EABB)} \quad \mathrm{E}_\theta\left(\sum_{i=1}^{I} t_i(\theta) + x(\theta)C(I)\right) \leq 0.$$

Of course, BB implies EABB.[32]

Let

$$U_i(\theta_i) \equiv \mathrm{E}_{\theta_{-i}}[x(\theta_i, \theta_{-i})\theta_i + t_i(\theta_i, \theta_{-i})]$$

denote agent i's expected utility when he has type θ_i, and let

$$X_i(\theta_i) \equiv \mathrm{E}_{\theta_{-i}}[x(\theta_i, \theta_{-i})]$$

denote the probability that the good is supplied. The analysis of section 7.3 implies that

$$U_i(\theta_i) = U_i(\underline{\theta}_i) + \int_{\underline{\theta}_i}^{\theta_i} X_i(\tilde{\theta}_i)\, d\tilde{\theta}_i. \tag{7.25}$$

The expected total surplus, W, which is equal to the expectation of the sum of the budget surplus and the agents' utilities, is then

$$\begin{aligned} W &= \mathrm{E}_\theta\left(\sum_i [-t_i(\theta)] - C(I)x(\theta) + \sum_i U_i(\theta_i)\right) \\ &= \mathrm{E}_\theta\left(\sum_i [-t_i(\theta)] - C(I)x(\theta) + \sum_i U_i(\underline{\theta}_i)\right) \\ &\quad + \sum_i \mathrm{E}_{\theta_i}\left[\left(\frac{1 - P_i(\theta_i)}{p_i(\theta_i)}\right) X_i(\theta_i)\right], \end{aligned} \tag{7.26}$$

where

$$\int_{\underline{\theta}_i}^{\bar{\theta}_i} \int_{\underline{\theta}_i}^{\theta_i} X_i(\tilde{\theta}_i)\, d\tilde{\theta}_i p_i(\theta_i)\, d\theta_i$$

has been integrated by parts. Now,

$$\mathrm{E}_\theta\left(\sum_i [-t_i(\theta)] - C(I)x(\theta)\right) \geq 0$$

by *ex ante* budget balance, and $U_i(\underline{\theta}_i) \geq 0$ for all i by individual rationality. Because

$$0 \leq \mathrm{E}_\theta\left[\sum_i U_i(\theta_i)\right] = \mathrm{E}_\theta\left[\sum_i (t_i(\theta) + \theta_i x(\theta))\right] \leq \mathrm{E}_\theta\left[\sum_i \theta_i x(\theta) - C(I)x(\theta)\right]$$

(using the budget constraint), an integration by parts yields

32. But Mailath and Postlewaite show that if EABB, IC, and IR are satisfied, one can choose the transfers $t_i(\cdot)$ such that BB, IC, and IR are satisfied as well.

$$\mathrm{E}_\theta\left\{\left[\sum_i\left(\theta_i-\frac{1-P_i(\theta_i)}{p_i(\theta_i)}-\frac{C(I)}{I}\right)\right]x(\theta)\right\}\geq 0. \tag{7.27}$$

We will use the following lemma:

Lemma 7.1 The expectation of the virtual valuation is equal to the lower bound of the interval.

Proof Integrating by parts,

$$\begin{aligned}&\int_{\underline{\theta}_i}^{\bar{\theta}_i}\left(\theta_i-\frac{1-P_i(\theta_i)}{p_i(\theta_i)}\right)p_i(\theta_i)\,d\theta_i\\&=\int_{\underline{\theta}_i}^{\bar{\theta}_i}\theta_i p_i(\theta_i)\,d\theta_i-\{[1-P_i(\theta_i)]\theta_i\}_{\underline{\theta}_i}^{\bar{\theta}_i}-\int_{\underline{\theta}_i}^{\bar{\theta}_i}\theta_i p_i(\theta_i)\,d\theta_i\\&=\underline{\theta}_i.\end{aligned}$$

■

We next assume that the per-capita cost of supplying the public good is constant, $C(I)/I=c$, and that $c>\underline{\theta}_i$ for all i. Let us further assume (for simplicity) that all θ_i are drawn from the same distribution $P(\cdot)$ on $[\underline{\theta},\bar{\theta}]$, so that $c>\underline{\theta}$.

Note that the left-hand side of equation 7.27 is maximized by $x(\theta)=1$ if

$$\sum_i\left(\theta_i-\frac{1-P(\theta_i)}{p(\theta_i)}-c\right)\geq 0$$

and by $x(\theta)=0$ otherwise.

With a continuum of agents, the realized distribution of types in the population of agents coincides with the prior distribution. As $\underline{\theta}<c$, and (from the lemma) the expected total virtual surplus is equal to $\mathrm{E}_\theta[(\underline{\theta}-c)x(\theta)]$, in order for expected surplus to be nonnegative, x must be equal to 0 with probability 1. With a large but finite number of agents, the law of large numbers suggests that the same result holds approximately. Technical work is needed to make this intuition precise, but Rob (1989) and Mailath and Postlewaite (1990) show that, as I tends to ∞, the probability that the public project is implemented tends to 0 if $c>\underline{\theta}$, and IR, IC, and BB are required.[33]

Thus, with a large number of agents it becomes very hard to reach agreement. The inefficiency involved can be large. For, suppose that $P(c)$ is very small, so that, with probability close to 1, each agent's valuation for the public good exceeds the per-capita cost of supplying the public good.

33. A similar result is obtained by Roberts (1976) for dominant-strategy mechanisms rather than Bayesian ones. In contrast, Green and Laffont (1979) show that, in the absence of the IR constraint, efficiency can be obtained in the limit when the number of agents becomes large, with dominant strategies and budget balance.

Then there are gains from trade with probability close to 1, yet gains from trade are realized with probability close to 0.

The intuition for this result is straightforward. The probability of being pivotal (changing x through a change in $\hat{\theta}_i$) is very small with many agents, and is 0 with a continuum of them.[34] Thus, agent i's objective is simply to maximize his expected transfer, i.e., minimize his expected contribution to the public good. This expected contribution cannot exceed $\underline{\theta}$, because that would violate individual rationality for type $\underline{\theta}$, and agent i can always report type $\underline{\theta}$. But, if the expected contribution is at most $\underline{\theta}$, the cost of realizing the project cannot be covered, contradicting budget balance.

To avoid inefficiencies, one needs subsidies from an external source (such as a "government"). Mailath and Postlewaite show that the per-capita subsidy to implement the efficient provision of the public good (i.e., $x = 1$ if and only if $\sum_i \theta_i \geq cI$) is asymptotically equal to $c - \underline{\theta}$, as one would suspect.

7.5 Mechanism Design with Several Agents: Optimization††

In section 7.4 we looked at general properties of the implementable allocations. We now look at the *optimal* choice of a mechanism for two allocation problems. In the first, the auction example, a self-interested principal sells a good to one of several buyers with private information about their willingness to pay for the good. In the second, the bilateral-trade example, a seller and a buyer with private information about their cost and valuation may trade an object. In both cases we will assume that the mechanism is designed by an uninformed party to maximize her objective function, which will allow us to abstract from issues arising from information leakages through contract design (see subsection 7.6.3). In the auction example this corresponds to the assumption that the seller has no private information and maximizes her expected revenue. In the bilateral-trade example the interpretation is more difficult. There, it will be assumed that a benevolent third party maximizes the expected gains from trade between the buyer and the seller; as we will discuss, the existence of this third party is mysterious and the main point of the analysis is to supply an upper bound on the efficiency of bilateral exchange under asymmetric information.

7.5.1 Auctions

Suppose that a seller (the principal) has $\hat{x}$ units of a good for sale. There are I potential buyers (agents): $i = 1, \ldots, I$. All parties have quasi-linear preferences:

34. A similar idea underlies the "paradox of voting"—with a large number of voters, the probability of affecting the outcome of an election is infinitesimal. Palfrey and Rosenthal 1985 is a recent paper on this topic.

$$u_i = V_i(x_i, \theta_i) + t_i \text{ for } i = 0, 1, \dots, I$$

where $x_i \in [0, \hat{x}]$ is the amount consumed by party i and t_i is his (or her) income (in this section, $t_0 = -\sum_{i=1}^{I} t_i$). We assume that V_i is increasing in x_i, and that the sorting condition holds:

$$\frac{\partial^2 V_i}{\partial x_i \partial \theta_i} \geq 0,$$

that is, the marginal utility of the good increases in θ_i.

The seller's parameter θ_0 is common knowledge. In contrast, the buyers' types θ_i are independently drawn from cumulative distributions $P_i(\cdot)$ with strictly positive densities $p_i(\cdot)$ on $[\underline{\theta}, \overline{\theta}]$.

The seller attempts to maximize her expected utility. From the revelation principle, she can restrict attention to direct revelation mechanisms $\{x(\cdot), t(\cdot)\}$. Thus, she maximizes her expected (net) revenue:

$$R = \mathrm{E}_\theta \left[V_0 \left(\hat{x} - \sum_{i=1}^{I} x_i(\theta), \theta_0 \right) - \sum_{i=1}^{I} t_i(\theta) \right]$$

subject to

$$\text{(IC)} \quad \mathrm{E}_{\theta_{-i}}[V_i(x_i(\theta_i, \theta_{-i}), \theta_i) + t_i(\theta_i, \theta_{-i})] \geq \mathrm{E}_{\theta_{-i}}[V_i(x_i(\hat{\theta}_i, \theta_{-i}), \theta_i) + t_i(\hat{\theta}_i, \theta_{-i})] \quad \text{for all } (i, \theta_i, \hat{\theta}_i),$$

$$\text{(IR)} \quad \mathrm{E}_{\theta_{-i}}[V_i(x_i(\theta_i, \theta_{-i}), \theta_i) + t_i(\theta_i, \theta_{-i})] \geq 0 \text{ for all } (i, \theta_i),$$

and

$$x_i(\theta) \geq 0 \text{ and } \sum_{i=1}^{I} x_i(\theta) \leq \hat{x} \text{ for all } \theta.$$

Let

$$U_i(\theta_i) \equiv \mathrm{E}_{\theta_{-i}}[V_i(x_i(\theta_i, \theta_{-i}), \theta_i) + t_i(\theta_i, \theta_{-i})]$$

denote buyer i's expected utility when he has type θ_i. The seller's objective function can be rewritten as a function of the buyers' expected utilities by substituting for the transfers:

$$R = \mathrm{E}_\theta \left[V_0 \left(\hat{x} - \sum_{i=1}^{I} x_i(\theta), \theta_0 \right) + \sum_{i=1}^{I} V_i(x_i(\theta), \theta_i) \right] - \sum_{i=1}^{I} \mathrm{E}_{\theta_i} U_i(\theta_i). \tag{7.28}$$

But, from the envelope theorem,

$$\frac{dU_i}{d\theta_i} = \mathrm{E}_{\theta_{-i}} \left(\frac{\partial V_i}{\partial \theta_i}(x_i(\theta_i, \theta_{-i}), \theta_i) \right) \tag{7.29}$$

or

$$U_i(\theta_i) = U_i(\underline{\theta}) + \int_{\underline{\theta}}^{\theta_i} \mathrm{E}_{\theta_{-i}}\left(\frac{\partial V_i}{\partial \theta_i}(x_i(\tilde{\theta}_i, \theta_{-i}), \tilde{\theta}_i)\right) d\tilde{\theta}_i. \tag{7.30}$$

At the optimum, $U_i(\underline{\theta}) = 0$, as the seller does not want to leave unnecessary rents to the buyers. Substituting equation 7.30 into equation 7.28 and integrating by parts yields

$$R \equiv E_\theta\left[V_0\left(\hat{x} - \sum_{i=1}^{I} x_i(\theta), \theta_0\right) + \sum_{i=1}^{I}\left(V_i(x_i(\theta), \theta_i) - \frac{1 - P_i(\theta_i)}{p_i(\theta_i)}\frac{\partial V_i}{\partial \theta_i}(x_i(\theta), \theta_i)\right)\right]. \tag{7.31}$$

The optimal auction defines an allocation $x_i(\cdot)$ of the good so as to maximize R subject to the agents' incentive compatibility. Rather than give a comprehensive study of incentive compatibility, we content ourselves with a full treatment of a special case. Assume that

$$V_i(x_i, \theta_i) = \theta_i x_i, \quad i = 0, 1, \ldots, I$$

and

$$\hat{x} = 1.$$

We know from theorem 7.2 that incentive compatibility for agent i is equivalent to equation 7.30 plus the condition that $X_i(\theta_i) \equiv \mathrm{E}_{\theta_{-i}} x_i(\theta_i, \theta_{-i})$ be nondecreasing.

Hence, the optimal auction solves

$$\mathrm{Max}\ \mathrm{E}_\theta\left[\sum_{i=1}^{I}\left(\theta_i - \frac{1 - P_i(\theta_i)}{p_i(\theta_i)}\right)x_i(\theta) + \theta_0\left(1 - \sum_{i=1}^{I} x_i(\theta)\right)\right] \tag{7.32}$$

subject to

$$\sum_{i=1}^{I} x_i(\theta) \leq 1,\ x_i(\theta) \geq 0 \text{ for all } \theta \tag{7.33}$$

and

$$X_i(\cdot) \text{ nondecreasing.} \tag{7.34}$$

The expected transfers associated with the optimal auction are obtained by computing $U_i(\theta_i)$ and using the definition of U_i:

$$T_i(\theta_i) = \mathrm{E}_{\theta_{-i}} t_i(\theta_i, \theta_{-i}) = -\theta_i X_i(\theta_i) + \int_{\underline{\theta}}^{\theta_i} X_i(\tilde{\theta}_i)\, d\tilde{\theta}_i. \tag{7.35}$$

Note that maximizing equation 7.35 determines only *expected* transfers $T_i(\cdot)$, so there is a lot of leeway in defining the *ex post* transfers $t_i(\cdot)$. We will see that this leeway translates into a multiplicity of ways of implementing the optimal auction.

Let

$$J_i(\theta_i) \equiv \theta_i - \frac{1 - P_i(\theta_i)}{p_i(\theta_i)}$$

denote the virtual valuation of buyer i, and let $J_0(\theta_0) \equiv \theta_0$ denote the seller's valuation. We first maximize expression 7.32, ignoring the incentive-compatibility constraint 7.34. This yields

$$x_i(\theta) = 1 \text{ iff } J_i(\theta_i) = \max_{j \in \{0,\ldots,I\}} J_j(\theta_j).$$

(We ignore the cases in which the maximum is reached for at least two players. Such cases have probability 0.)

If $J_i(\cdot)$ is nondecreasing for all i (which is true in particular if the monotone-hazard-rate condition holds—see section 7.3), then, if $x_i(\theta_i, \theta_{-i}) = 1$,

$$x_i(\theta_i', \theta_{-i}) = 1 \text{ for all } \theta_i' > \theta_i.$$

Hence, $X_i(\cdot)$ is nondecreasing, and the ignored incentive-compatibility constraint is automatically satisfied. If $J_i(\cdot)$ decreases over some interval, one must proceed along the lines of the analysis of bunching in the appendix to this chapter. (See Myerson 1981 for details.) In the following, we will assume that $J_i(\cdot)$ is nondecreasing.

We now examine the implications of this analysis.

First, note that the relevant comparison concerns the parties' *virtual valuations*, and not their valuations. The seller's virtual valuation is equal to her true valuation θ_0, because the seller has full information about herself and therefore needs not introduce the incentive cost of revelation of information.

Second, all auctions that yield the same decision $x_i(\cdot)$ and give zero surplus to type $\underline{\theta}$ of each buyer yield the same revenue. We will shortly give an implication of this fact, known as the *revenue equivalence theorem*.

Third, the analysis yields a number of standard results in the symmetric case ($P_i(\cdot) = P(\cdot)$). In this case, the good goes to the highest-valuation buyer if it is sold at all. The good is sold if and only if

$$\max_{i \in \{1,\ldots,I\}} \theta_i \geq \theta^*,$$

where $\theta^* > \theta_0$ is defined by

$$\theta^* - \frac{1 - P(\theta^*)}{p(\theta^*)} \equiv \theta_0.^{35}$$

35. Again, this result generalizes the monopoly-pricing paradigm. Note that if $\theta_0 < \max_{i \in \{1,\ldots,I\}} \theta_i < \theta^*$, gains from trade are not realized. The seller distorts the auction in her favor.

Furthermore, all auctions that give the good to the highest bidder (i.e., $X_i(\theta_i) \equiv [P(\theta_i)]^{I-1}$ if $\theta_i \geq \theta^*$, $\equiv 0$ otherwise) and yield zero surplus to a bidder with valuation θ^* (or, equivalently, to valuation $\underline{\theta}$ from equation 7.30) yield the same revenue to the seller.

In particular, a first-price auction (see chapter 6) and a second-price auction (see chapter 1), each with minimum or reservation price θ^*, yield the same revenue and are optimal (Vickrey 1961; Myerson 1981; Riley and Samuelson 1981). Although the first- and second-price auctions yield the same x_i and T_i, they yield different t_i: When bidder i wins, his payment depends only on his bid and therefore only on θ_i in a first-price auction, and depends only on the second bid ($\max_{j \neq i; j \in \{1,\ldots,I\}} \theta_j$) in a second-price auction. This illustrates the leeway one has in building the *ex post* transfers t_i to implement an optimal auction. Note also that the two-type example in section 7.1 shows that neither the first- nor the second-price auction is optimal when the distribution of types is discrete. The problem with both of these auctions is that the high-valuation type receives an unnecessarily high rent. Starting from the second-price auction, for example, the seller can increase her revenue while still inducing buyers to bid their valuations if she specifies that when one buyer bids $\bar{\theta}$ and the other bids $\underline{\theta}$ the high bidder receives the good at price $\underline{\theta} + (\bar{\theta} - \underline{\theta})/2$.

In the asymmetric case, the auction does not necessarily allocate the good to the bidder with the highest willingness to pay (Myerson 1981; McAfee and McMillan 1987b). In particular, suppose that there are two bidders ($i = 1, 2$) and that, for all θ,

$$\frac{1 - P_1(\theta)}{p_1(\theta)} \geq \frac{1 - P_2(\theta)}{p_2(\theta)}.$$

That is, bidder 1 is "on average" more eager to buy than bidder 2. Then the auction should be biased in favor of bidder 2. There exist θ_1 and θ_2 such that $\theta_1 > \theta_2$ but $x_2(\theta_1, \theta_2) = 1$, while there exist no θ_1 and θ_2 such that $\theta_2 > \theta_1$ and $x_1(\theta_1, \theta_2) = 1$.

7.5.2 Efficient Bargaining Processes†††

Consider now a single buyer and a single seller. The seller has one unit for sale and has private information about his cost c of supplying the unit. The buyer has unit demand and has private information about his willingness to pay or valuation v for the unit. Thus, $\theta_1 \equiv c$, $\theta_2 \equiv v$, and $\theta \equiv (c, v)$. c and v are independently drawn from cumulative distributions $P_1(\cdot)$ and $P_2(\cdot)$ on $[\underline{c}, \bar{c}]$ and $[\underline{v}, \bar{v}]$, with strictly positive densities $p_1(\cdot)$ and $p_2(\cdot)$. The two parties are risk neutral.

A balanced-budget mechanism is a probability $x(c, v) \in [0, 1]$ that the traders exchange the good given that their types are c and v and a payment $w(c, v)$ (or, equivalently, given that the parties are risk neutral, an expected payment $w(c, v)$) from the buyer to the seller (in our previous notation, $t_1(\theta) - w(c, v) = -t_2(\theta)$). Let

$$X_1(c) \equiv E_v\, x(c, v);\; X_2(v) \equiv E_c\, x(c, v);$$

$$W_1(c) \equiv E_v\, w(c, v);\; W_2(v) \equiv E_c\, w(c, v);$$

$$U_1(c) \equiv -c\, X_1(c) + W_1(c);\; U_2(v) \equiv v\, X_2(v) - W_2(v).$$

The mechanism is individually rational if $U_1(c) \geq 0$ for all c and $U_2(v) \geq 0$ for all v. It is incentive compatible if

$$U_1(c) \geq -c\, X_1(\hat{c}) + W_1(\hat{c}) \text{ for all } (c, \hat{c})$$

and

$$U_2(v) \geq v\, X_2(\hat{v}) - W_2(\hat{v}) \text{ for all } (v, \hat{v}).$$

Consider a benevolent principal trying to maximize expected social surplus $E_{\{c,v\}}[(v - c)x(c, v)]$, and suppose that she is able to design a (balanced-budget) mechanism to which the seller and the buyer must comply as long as it is individually rational and incentive compatible. The role of the principal here is difficult to interpret. She might stand for a government, but then it is not clear why the mechanism must satisfy the individual-rationality constraints, since governments have coercive powers. Another potential interpretation is that the parties appeal to a mediator (the principal) to design an efficient mechanism. This interpretation also is often questionable. If the parties appeal to the mediator once they have received their private information (at the "interim stage"), the bargaining over whether to have a mediator and over which objective function to give to the mediator is likely to reveal information about the cost and the valuation; the IR and IC constraints are then misspecified in that the mechanism is played under posterior beliefs that differ from the prior beliefs $P_1(\cdot)$ and $P_2(\cdot)$. If the two parties decide to use a mediator *before* they receive their private information (at the "*ex ante*" stage), they may be able to commit themselves to use the mechanism once they learn their valuations, and so the interim IR constraint may not be relevant. Such commitments can sometimes be accomplished by contractually specified damages for "opting out" or "breach of contract."[36] If parties can commit to use the mechanism, they typically prefer to do so, as binding interim IR constraints generally create inefficiency, whereas in the absence of these constraints AGV mechanisms can be built that implement the *ex post* efficient outcome (i.e., $x = 1$ if $v \geq c$, $= 0$ if $v < c$).

36. However, these commitment options may be limited: If we interpret the "seller" as a worker who is providing labor to a firm, workers are not allowed to agree to fines for quitting; however, firm might still be able to commit. This suggests a hybrid model with only one player subject to an individual-rationality constraint.

Because of these reservations about interpretations where a principal designs the mechanism, the best interpretation of the model may be as a characterization of utilities that can be achieved by equilibria of noncooperative bargaining games. Suppose that the seller and the buyer bargain over whether to trade and over the price. The bargaining process can be a simultaneous sealed-bid auction (à la Chatterjee and Samuelson—see chapter 6) or a more complex, sequential bargaining game (see chapter 10). It has been known for a while as part of the profession's folklore that any (Bayesian) equilibrium of a bargaining process gives rise to an allocation that can be interpreted as a mechanism that satisfies IC and IR, as long as the two traders have identical time preferences.[37] This is a straightforward application of the revelation principle: Suppose that bargaining starts at date 0 and that both traders discount the future at interest rate $r > 0$ (we allow for either discrete-time bargaining—at dates $t = 0, 1, 2, \ldots$—or continuous-time bargaining). Let agreement to trade between the seller with cost c and the buyer with valuation v be reached at time $\tau(c, v)$ at price $z(c, v)$ (we assume that τ and z are deterministic; the reasoning extends straightforwardly to stochastic τ and z). $\tau = +\infty$ corresponds to the case in which agreement is never reached. One can then define

$$x(c, v) \equiv e^{-r\tau(c,v)} \in [0, 1],$$

$$w(c, v) \equiv e^{-r\tau(c,v)} z(c, v),$$

$$U_1(c) \equiv \mathrm{E}_v[w(c, v) - c\,x(c, v)],$$

$$U_2(v) \equiv \mathrm{E}_c[v\,x(c, v) - w(c, v)].$$

Note that delay in reaching agreement ($\tau > 0$) amounts to a probability that exchange does not take place ($x < 1$) in the mechanism reinterpretation.

Observe that the mechanism $\{x(\cdot,\cdot), w(\cdot,\cdot)\}$ satisfies IR, IC, and BB. It is individually rational because each trader can always refuse to trade (by making outrageous demands, and rejecting all offers), and thus get 0. By definition of a Bayesian equilibrium, it satisfies incentive compatibility: A type θ_i of player i cannot adopt the strategy of type $\hat{\theta}_i$ of the same player and obtain a higher expected payoff. Budget balance follows from the absence of a third party.

Viewed from this perspective, the program of computing the highest expected social surplus that can be obtained through individually rational, incentive-compatible, balanced-budget mechanisms can be interpreted as deriving an upper bound on the efficiency of unmediated bilateral bargaining.

37. If the traders have different rates of time preference, then having the more patient trader make loans to the less patient one allows the attainment of utility levels that are not feasible in the static problem.

Remark In the same spirit, one can derive the set of allocations that can be implemented by a mediator. The question is then whether any element in this set may arise as *an* equilibrium of *some* unmediated bargaining game. This line of research will be discussed in chapter 10.

Let us now derive the mechanism that maximizes expected gains from trade,

$$E_{c,v}[(v-c)x(c,v)], \tag{7.36}$$

subject to IR, IC, and BB. We saw in subsection 7.4.4 that IR, IC, and BB imply

$$E_{c,v}\{[J_2(v) - J_1(c)]x(c,v)\} \geq 0, \tag{7.37}$$

where

$$J_1(c) \equiv c + \frac{P_1(c)}{p_1(c)}$$

and

$$J_2(v) = v - \frac{1 - P_2(v)}{p_2(v)}.$$

Conversely, if the function $x(\cdot,\cdot)$ maximizes expression 7.36 subject to inequality 7.37, there exists a transfer function $t(\cdot,\cdot)$ that satisfies BB (by definition), satisfies IR, and satisfies IC as long as $X_1(c) = E_v x(c,v)$ is nonincreasing and $X_2(v) = E_c x(c,v)$ is nondecreasing. With $\mu \geq 0$ denoting the multiplier of equation 7.37, the Lagrangian for the above program is

$$\mathscr{L} = E_{c,v}(\{(v - c) + \mu[J_2(v) - J_1(c)]\}x(c,v)). \tag{7.38}$$

The first-order condition is thus

$$x(c,v) = \begin{cases} 1 \text{ if } v + \mu J_2(v) \geq c + \mu J_1(c) \\ 0 \text{ otherwise.} \end{cases} \tag{7.39}$$

Thus, trade occurs if and only if

$$v - \left(\frac{\mu}{1+\mu}\right)\frac{1 - P_2(v)}{p_2(v)} \geq c + \left(\frac{\mu}{1+\mu}\right)\frac{P_1(c)}{p_1(c)}. \tag{7.40}$$

Equation 7.40 does not quite yet define the solution, as the coefficient $\alpha \equiv \mu/(1+\mu) \in [0,1)$ must still be specified. To this purpose, it suffices to note that equation 7.37 must be satisfied with equality if $\bar{c} > \underline{v}$.[38] (Ideally, one would want the trading rule to come as close as possible to the first-best

38. If the inequality is strict in equation 7.37, $\mu = 0$ and equation 7.40 is the first-best rule. But we know from subsection 7.4.4 that, as long as $\bar{c} > \underline{v}$, efficient trade is inconsistent with IR, IC, and BB.

trading rule (trade if and only if $v \geq c$); i.e., one would want μ (or α) to be as small as possible. Equation 7.37 has been relaxed as much as is consistent with IR, IC, and BB by imposing $U_1(\bar{c}) = U_2(\underline{v}) = 0$.)

Note again that if the monotone-hazard-rate conditions hold ($p_2/(1 - P_2)$ nondecreasing, p_1/P_1 nonincreasing), equation 7.40 yields monotonic $X_1(\cdot)$ and $X_2(\cdot)$, so the optimal trading rule has indeed been obtained.

Myerson and Satterthwaite apply equation 7.40 to the case of uniform densities on $[0, 1]$ ($P_1(c) = c$ and $P_2(v) = v$ for $(c, v) \in [0, 1]^2$). Equation 7.40 then yields

$$v - c \geq \frac{\alpha}{1 + \alpha}. \tag{7.41}$$

Substituting into equation 7.37 yields

$$\int_0^{1-(\alpha/(1+\alpha))} \left(\int_{c+(\alpha/(1+\alpha))}^{1} [(2v - 1) - 2c]\, dv \right) dc = 0, \tag{7.42}$$

which has solution $\alpha/(1 + \alpha) = \frac{1}{4}$. In the optimal trading rule, trade occurs if and only if the buyer's valuation exceeds the seller's cost by at least one-fourth. Thus, in the uniform case, the linear equilibrium of the Chatterjee-Samuelson double auction exhibited in chapter 6 yields the optimal amount of trade constrained by IR, IC, and BB![39]

7.6 Further Topics in Mechanism Design†††

The bare-bones analysis of this chapter has ignored many of the recent extensions of the mechanism-design paradigm. In this concluding section, we give the flavor of a few of these extensions.

7.6.1 Correlated Types

Section 7.5 assumed that the agents' types were independent. Maskin and Riley (1980), Crémer and McLean (1985, 1988), McAfee, McMillan, and Reny (1989), Johnson, Pratt, and Zeckhauser (1990), and d'Aspremont, Crémer, and Gérard-Varet (1990a,b) have shown in various environments that, when preferences are quasi-linear (risk neutrality) and the agents' types are correlated, the principal can implement the same allocation she would implement if she knew the agents' types.[40] Thus, IC is not binding under risk neutrality and correlated types.

39. This result is not robust. Satterthwaite and Williams (1989) show that optimal trading allocations cannot be implemented by double auctions for "generic" pairs of prior distributions.

40. Recall from subsection 7.4.3 that correlation is not needed when the principal wants to maximize the sum of the agents' utilities. The result here is interesting when there is a conflict between the objectives of the principal and the agents.

To get some intuition about why this is so, suppose that the agents' types are perfectly correlated. Then each knows the others' types. Let the principal organize a "shoot them all" mechanism: The principal asks the agents to announce the vector of the I types simultaneously. If all announcements coincide, the principal implements the optimal full-information allocation corresponding to the announced types (which may or may not satisfy IR constraints, depending on the case); if they do not coincide, the principal "shoots all agents": $t_i = -\infty$ for all i. Clearly, if all other agents announce the true vector of types, it is in the interest of the remaining agent to announce the true vector of types as well. Hence, the principal can costlessly obtain the agents' information and *de facto* has full information.[41]

This idea generalizes to the case of (even small) imperfect correlation of the agents' types. One can use the fact that an agent's information yields the best predictor of the other agents' information[42] to "shoot the agent stochastically" if he misreports his type. Because the agents and the principal are risk neutral, using transfers that depend not only on the agent's type but also on the other agents' types and therefore impose risk on the agent creates no social loss in terms of risk bearing.

The papers in the literature make a full-rank assumption. Assume that there are a *finite* number of types per agent. Let $p(\theta_{-i}|\theta_i)$ denote the probability of types θ_{-i} for players other than i conditional on player i's having type θ_i. Let $p_i^{\theta_i}$ denote the vector of

$$\{p(\theta_{-i}|\theta_i)\}_{\theta_{-i}\in\Theta_{-i}}.$$

The full-rank condition is satisfied if, for each i, the vectors

$$\{p_i^{\theta_i}\}_{\theta_i\in\Theta_i}$$

are linearly independent. That is, there do not exist an agent i, a type θ_i,

41. Note that there are many other equilibria in the "shoot them all" mechanism. For instance, all agents could announce the same incorrect vector of types. This multiplicity is precisely what gave rise to a large literature on unique Nash implementation, starting with Maskin 1977 (see Moore 1990 for a survey and a list of references). Some authors, including Maskin and Riley (1980), have also looked at equilibrium uniqueness in the imperfect-correlation case (see also the more general literature mentioned in section 7.2). Crémer and McLean (1985, 1988) obtain results on dominant-strategy as well as Bayesian implementation.

42. One formalization of the notion that "an agent's information yields the best predictor of the other agents' information" is obtained by considering the "proper scoring rules" familiar in the statistics literature: Suppose that agent i is asked to reveal his type $\hat{\theta}_i$, and is given transfer $\tau_i(\hat{\theta}) = \ln p(\hat{\theta}_{-i}|\hat{\theta}_i)$ when the other agents announce $\hat{\theta}_{-i}$. Suppose in a first step that no decision x is at stake, so that agent i aims at maximizing his expected transfer. It is easily checked that, if the other agents announce truthfully, it is in the interest of agent i to announce his type truthfully, and strictly so if the vectors of conditional probabilities differ.

When there is a payoff-relevant decision x, such as allocating a good among bidders or supplying a public good, agent i's payoff function depends on the decision as well as his report, and the above proper scoring rule τ_i may no longer induce truthful revelation. However, one can "scale up" τ_i by multiplying by a large positive constant K. Then, any misreport of type implies substantial losses in the transfer $K\tau_i$, which swamps any effect on V_i of misreporting the type. Johnson et al. (1990) use such inflated proper scoring rules (to which they add further terms to meet other constraints such as budget balance).

and a vector of positive numbers $\rho_i(\theta_i')$ such that

$$p_i^{\theta_i} = \sum_{\theta_i' \neq \theta_i} \rho_i(\theta_i') p_i^{\theta_i'}.$$

In words, the full-rank condition means that the vectors of agent i's conditional probabilities about the other agents' types can be told apart.

Crémer and McLean (1985) show that the principal can implement any decision rule $x^*(\cdot)$ and agents' utilities $U_i^*(\cdot)$ under risk neutrality and full rank, even if the principal does not know θ. We illustrate their construction in the case of two agents and two types per agent. Agent i can have type θ_i or θ_i. Let q_{11} and q_{12} denote the conditional probabilities that $\theta_2 = \underline{\theta}_2$ and $\theta_2 = \theta_2$ when $\theta_1 = \theta_1$; the conditional probabilities when $\theta_1 = \bar{\theta}_1$ are q_{21} and q_{22}. The full-rank condition for player 1 is $q_{11}q_{22} \neq q_{21}q_{12}$. Let t_{11} and t_{12} denote the transfers to agent 1 when he announces $\underline{\theta}_1$ and agent 2 announces $\underline{\theta}_2$ and $\bar{\theta}_2$, respectively. And similarly for t_{21} and t_{22}. The decisions and utilities are indexed in the same way. To yield the desired utilities, the transfers must satisfy, for some constants A_1 and A_2 determined by the data of the problem,[43]

$$q_{11}t_{11} + q_{12}t_{12} = A_1 \tag{7.43}$$

and

$$q_{21}t_{21} + q_{22}t_{22} = A_2. \tag{7.44}$$

The transfers must also ensure incentive compatibility for player 1 with type θ_1 or θ_1. That is,

$$q_{11}(t_{11} - t_{21}) + q_{12}(t_{12} - t_{22}) \geq A_3 \tag{7.45}$$

and

$$q_{21}(t_{21} - t_{11}) + q_{22}(t_{22} - t_{12}) \geq A_4, \tag{7.46}$$

where A_3 and A_4 are constants determined by the data.[44]

Substituting equations 7.43 and 7.44 into equations 7.45 and 7.46 yields

$$(q_{11}q_{22} - q_{21}q_{12})t_{11} \geq A_5 \equiv A_1 q_{22} + (A_4 - A_2)q_{12} \tag{7.47}$$

43. Where

$$A_1 \equiv q_{11}(U_{11}^* - V_1(x_{11}^*, \theta_1)) + q_{12}(U_{12}^* - V_1(x_{12}^*, \underline{\theta}_1))$$

and

$$A_2 \equiv q_{21}(U_{21}^* - V_1(x_{21}^*, \bar{\theta}_1)) + q_{22}(U_{22}^* - V_1(x_{22}^*, \bar{\theta}_1)).$$

44. The reader will check that

$$A_3 \equiv q_{11}(V_1(x_{21}^*, \underline{\theta}_1) - V_1(x_{11}^*, \underline{\theta}_1)) + q_{12}(V_1(x_{22}^*, \underline{\theta}_1) - V_1(x_{12}^*, \underline{\theta}_1)).$$

and

$$A_4 \equiv q_{21}(V_1(x_{11}^*, \bar{\theta}_1) - V_1(x_{21}^*, \bar{\theta}_1)) + q_{22}(V_1(x_{12}^*, \bar{\theta}_1) - V_1(x_{22}^*, \theta_1)).$$

and

$$(q_{11}q_{22} - q_{21}q_{12})t_{21} \leq A_6 \equiv -A_2 q_{12} - (A_3 - A_1)q_{22}. \tag{7.48}$$

Transfers satisfying equations 7.47, 7.48, 7.43, and 7.44 yield the desired allocation for the principal, and such transfers always exist under the full-rank condition. Note, however, that, as types become less correlated, $(q_{11} - q_{21})$ and $(q_{12} - q_{22})$ both converge to 0, so $q_{11}q_{22} - q_{21}q_{12}$ converges to 0, and so the transfers required to satisfy inequalities 7.47 and 7.48 become very large. Transfers for agent 2 can be constructed in a similar manner (given the full-rank condition for player 2). More generally, with an arbitrary number of types (and players), Farkas's lemma (which gives conditions for a system of linear inequalities and equalities to have a solution—see, e.g., section 22 of Rockafellar 1970) and the full-rank condition can be used to prove the existence of appropriate transfers.[45]

Of course, the result that the principal can use any arbitrarily small amount of correlation to achieve the full-information outcome while she usually suffers from the asymmetry of information under independent distributions of types is extreme. The point is that the credibility of risk neutrality is stretched by the very large transfers required for small correlations.

7.6.2 Risk Aversion

Most of the literature on mechanism design has focused on the case of quasi-linear preferences. We saw in sections 7.4 and 7.5 that in this case optimal mechanism design with several agents is a simple extension of mechanism design with a single agent. With risk-averse agents, one still makes heavy use of the single-agent framework and its optimal-control techniques, but things become harder.

To illustrate the issues, consider the problem of designing an optimal auction for one unit of a good when the buyers are risk averse, have the same preferences, and have types that are independently drawn from the same distribution $P(\cdot)$ on $[\underline{\theta}, \bar{\theta}]$ (the theory was developed by Maskin and Riley (1984) and Matthews (1983)). To allow for the case in which the agents have utility functions that are not separable in income and consumption, one must consider two transfers, $t_i(\hat{\theta})$ and $\tilde{t}_i(\hat{\theta})$, according to whether the agent wins or loses in the auction (for simplicity, assume that these transfers are deterministic). Let $u(t_i(\hat{\theta}), \theta_i)$ and $w(\tilde{t}_i(\hat{\theta}))$ denote the utilities of agent i when he wins and when he loses the auction. Let

$$t_i(\hat{\theta}_i) \equiv \mathrm{E}_{\theta_{-i}} t_i(\hat{\theta}_i, \theta_{-i})$$

and

45. With a continuum of types, an agent can approximate the true conditional probability distribution arbitrarily closely by lying. One must then solve "Fredholm equations" (see McAfee et al. 1989, Caillaud et al. 1986, and Melumad and Reichelstein 1989).

$$\hat{t}_i(\hat{\theta}_i) \equiv \mathrm{E}_{\theta_{-i}} \tilde{t}_i(\hat{\theta}_i, \theta_{-i}).$$

Eliminating the dependence of transfers on the other agents' announcements reduces the agent's risk and raises his utility. Doing so,[46] assuming a symmetric auction and eliminating subscripts under t_i and $\tilde{t}_i$, yields utility function for an agent of type θ_i:

$$U(\theta_i) = \max_{\hat{\theta}_i} \{X(\hat{\theta}_i)u(t(\hat{\theta}_i), \theta_i) + [1 - X(\hat{\theta}_i)]w(\tilde{t}(\hat{\theta}_i))\}, \tag{7.49}$$

where $X(\hat{\theta}_i) \equiv \mathrm{E}_{\theta_{-i}} x(\hat{\theta}_i, \theta_{-i})$ is the probability that the agent wins the auction. Let

$$U(\theta_i) = X(\theta_i)u(t(\theta_i), \theta_i) + [1 - X(\theta_i)]w(\tilde{t}(\theta_i)). \tag{7.50}$$

The envelope theorem implies that

$$\frac{dU}{d\theta_i} = X(\theta_i)\frac{\partial u}{\partial \theta_i}(t(\theta_i), \theta_i). \tag{7.51}$$

The principal maximizes her expected revenue per buyer,

$$\mathrm{Max} \int_{\underline{\theta}}^{\bar{\theta}} \{X(\theta_i)t(\theta_i) + [1 - X(\theta_i)]\tilde{t}(\theta_i)\} p(\theta_i)\, d\theta_i, \tag{7.52}$$

subject to equation 7.50, equation 7.51, (IR) $U(\underline{\theta}) \geq 0$, and "consistency."

The "consistency" constraint arises from the fact that, if equation 7.52 is maximized subject to only equation 7.50, equation 7.51, and IR, nothing guarantees that, given $X(\cdot)$, one can find a decision function $x(\cdot) \in [0, 1]$ such that

$$X(\theta_i) = \mathrm{E}_{\theta_{-i}}[x(\theta_i, \theta_{-i})] \text{ for all } (i, \theta_i). \tag{7.53}$$

In other words, analyzing isolated single-buyer problems ignores the constraint that there is a single unit of the good to be distributed among all buyers. The consistency constraint means that one must restrict attention to probabilities $X(\cdot)$ such that there exist a function $x(\cdot)$ satisfying equation 7.53.

In the case of an auction, there is fortunately a characterization of consistent $X(\cdot)$ that preserves the simple structure of an optimal-control problem. (This characterization is due to Maskin and Riley (1984) and Matthews (1983) and finds its most general formulation in Matthews 1984.) Namely, if $X(\cdot)$ is nondecreasing and satisfies

46. Further analysis is needed to prove that it is indeed optimal to eliminate this dependence. Assumptions on preferences must be made so that the agent's incentive-compatibility constraint is not relaxed through the use of a random scheme. (Even if these conditions on preferences are not met and optimal auctions involve random transfers, the optimal randomness has in general little to do with that created by the uncertainty about θ_{-i}.)

$$\int_{\theta}^{\bar{\theta}} [P(\tilde{\theta})^{I-1} - X(\tilde{\theta})]p(\tilde{\theta})\,d\tilde{\theta} \geq 0 \text{ for all } \theta \in [\underline{\theta}, \bar{\theta}], \tag{7.54}$$

then it is consistent.

It is easy to see that equation 7.54 is a necessary condition for consistency: The probability that a buyer with a valuation in $[\theta, \bar{\theta}]$ wins,

$$I \int_{\theta}^{\bar{\theta}} X(\tilde{\theta})p(\tilde{\theta})\,d\tilde{\theta},$$

cannot exceed the total probability that at least one buyer has valuation in $[\theta, \bar{\theta}]$,

$$1 - P(\theta)^I.$$

As

$$\frac{1 - P(\theta)^I}{I} = \int_{\theta}^{\bar{\theta}} P(\tilde{\theta})^{I-1} p(\tilde{\theta})\,d\tilde{\theta},$$

this yields equation 7.54. The difficult part of the characterization is to prove that equation 7.54 is sufficient for consistency.

7.6.3 Informed Principal

In this chapter we have assumed that the agents perfectly know the principal's preferences. It may be that the principal (the mechanism designer) also has private information. For instance, she may have information about the cost of supplying a public good, about her private cost of departing with the object in an auction, or about her willingness to pay for a good purchased from the agent.

Once the principal has private information, it must be recognized that the very proposal of a mechanism by the principal will reveal information about her type, as Myerson (1983) pointed out. Whereas Myerson analyzes this situation from a cooperative-game viewpoint, Maskin and Tirole (1989, 1990) keep the three-stage structure described in the introduction and used throughout the chapter and apply noncooperative game theory. (They use perfect Bayesian equilibrium rather than Bayesian equilibrium—see the next chapter. The concept mainly adds the extra requirement that, after observing the principal's contract offer, the agents update their beliefs about her type using Bayes' rule.)

One must distinguish between two situations. In the "private values" case, the principal's type does not enter the agents' preferences (but the agents' types are allowed to enter the principal's preferences). With y denoting the allocation and θ_0 the principal's type, the principal's utility is $u_0(y, \theta, \theta_0)$ and agent i's utility is $u_i(y, \theta)$. In contrast, if θ_0 affects some agent's utility, we have "common values." The difference between private and common values is that in the former case the agents care about the principal's type only to the extent that it affects the principal's behavior in the implementation of the mechanism, whereas in the latter case the agents care about her type *per se*. The three examples given at the beginning of this subsection exhibit private values. In contrast, if, in an auction, the seller's cost of departing with the good is correlated with an unknown-to-the-buyers quality of the good, we have common values.

A simple observation is that under *private values* the principal can guarantee herself the expected payoff she would obtain if the agents knew her type: It suffices that the principal offer the mechanism that is optimal for her when the agents know her type. Because the principal is not a player in the third stage (implementation of the mechanism), nothing is altered by the asymmetry of information about θ_0. The issue is then whether the principal can do better when her type is unknown to the agents than when it is common knowledge. Clearly, to do better the principal must participate in the third stage—for instance, by announcing her private information at the same time that the agents announce theirs. By delaying revelation of her information until after the proposal of the contract, the principal may be able to pool the agents' (IR or IC) constraints across her types. Indeed, Maskin and Tirole (1990) show that any equilibrium of the mechanism-design game can be computed as a Walrasian equilibrium of a fictitious economy. In this economy, the traders are the different types of principals, in proportions equal to those of the prior beliefs about θ_0, the goods traded are the slack variables on the agents' (IC and IR) constraints, and the traders have zero initial endowments of the goods.[47]

When preferences are quasi-linear, it turns out that the multipliers associated with the agents' IR and IC constraints do not depend on θ_0 when the agents know θ_0. Hence, the different types of principal do not gain by pooling when they offer a mechanism, as they do not gain by pooling constraints and trading slack. This implies that the unique equilibrium is the same as when the agents know θ_0. Thus, the single-agent theory of section 7.3 and the multi-agent theory of section 7.5 remain valid when the principal has private information, values are private, and preferences are quasi-linear.

In contrast, the analysis of this chapter must be amended when preferences are not quasi-linear. Generically, the multipliers of the agents' constraints do not coincide for different types of principal, and these types gain by trading slack on the constraints. In equilibrium, the principal does not reveal any of her information in the first step (contract proposal) and waits until the third step (contract implementation) to do so. And she does strictly better than when the agents know her type.

47. The paper considers a single agent, but the ideas extend to multiple agents, as this chapter would suggest.

The case of *common values* is more complex. For one thing, the principal may no longer be able to guarantee herself the same payoff as when agents know θ_0. The point is that the optimal mechanism when the agents know θ_0 need no longer be accepted by the agents if they draw the wrong inference about θ_0, as their utilities are directly affected by θ_0. Maskin and Tirole (1989) consider the restrictive case in which there is a single agent and this agent has no private information (and generalize their results to bilateral asymmetric information only in the case of quasi-linear preferences). The mechanism-design game is then similar to the standard signaling game we describe in section 8.2, except that the "sender" (the principal) has a large strategy space (the space of all contracts). The set of equilibria can be fully characterized, has a unique element for a subset of the agents' prior beliefs about θ_0, and has a continuum of elements for the complementary subset of beliefs.

7.6.4 Dynamic Mechanism Design

The static analysis of this chapter can be used to characterize repeated mechanism design as long as the principal and the agents can commit intertemporally (see, e.g., Baron and Besanko 1984a). Consider a multi-period problem, with periods $\tau = 0, 1, \ldots, T$. Suppose for instance that there is a single agent, with preferences

$$\sum_{\tau=0}^{T} \delta^{\tau} u_1(y_\tau, \theta)$$

where $y_\tau = (x_\tau, t_\tau)$ is the allocation at date τ and δ is the discount factor. The principal has preferences

$$\sum_{\tau=0}^{T} \delta^{\tau} u_0(y_\tau, \theta).$$

Note that we assume that the agent's type is invariant.[48]

Let $y^*(\theta)$ denote the optimal allocation for the principal subject to the agent's IR and IC constraints in a one-period context (see section 7.3). We claim that the allocation $y_\tau(\theta) = y^*(\theta)$ for all τ is optimal (i.e., the optimal allocation is the $(T+1)$ replica of the static one). To see this, suppose that the principal could do better than replicate the optimal static allocation. That is, assume that there exists an allocation $\{y_\tau(\cdot)\}_{\tau=0,\ldots,T}$ that satisfies the agent's multi-period IR and IC constraints,

$$\text{(multi-period IR)} \quad \sum_{\tau=0}^{T} \delta^{\tau} u_1(y_\tau(\theta), \theta) \geq \sum_{\tau=0}^{T} \delta^{\tau} \underline{u}_1(\theta) \text{ for all } \theta$$

(where $u_1(\theta)$ is the invariant per-period reservation utility of type θ), and

48. See Baron and Besanko 1984a for the case of a type that changes over time.

$$\text{(multi-period IC)} \quad \sum_{\tau=0}^{T} \delta^\tau u_1(y_\tau(\theta), \theta) \geq \sum_{\tau=0}^{T} \delta^\tau u_1(y_\tau(\hat{\theta}), \theta) \text{ for all } (\theta, \hat{\theta}),$$

and that yields more expected utility to the principal than y^* repeated $T + 1$ times:

$$\mathrm{E}_\theta\left(\sum_{\tau=0}^{T} \delta^\tau u_0(y_\tau(\theta), \theta)\right) > (1 + \delta + \cdots + \delta^T)(\mathrm{E}_\theta[u_0(y^*(\theta), \theta)]). \tag{7.55}$$

Now consider the random static mechanism that, for an announcement $\hat{\theta}$, gives the agent allocation $y_0(\hat{\theta})$ with probability $1/(1 + \cdots + \delta^T)$, $y_1(\hat{\theta})$ with probability $\delta/(1 + \cdots + \delta^T)$, ..., $y_T(\hat{\theta})$ with probability $\delta^T/(1 + \cdots + \delta^T)$. Dividing (multiperiod IR), (multiperiod IC), and equation 7.55 by $(1 + \cdots + \delta^T)$, this random allocation satisfies the (static) IR and IC constraints and yields more expected utility than $y^*(\cdot)$, a contradiction. Hence, the optimal static allocation remains optimal in a dynamic context with commitment.[49]

To implement the dynamic optimum, the principal asks the agent to reveal his type $\hat{\theta}$ at date 0, and then implements allocation $y^*(\hat{\theta})$ repeatedly until the end of the horizon. Note that it is important that the principal can commit. Otherwise we face the time-consistency problem studied in chapter 3. We saw in section 7.3 that (if $u_1(\underline{\theta}) = \underline{u}$ and u_1 is increasing in θ), except "at the bottom" ($\theta = \underline{\theta}$), the agent enjoys a rent associated with his private information ($u_1(y^*(\theta), \theta) > \underline{u}$). At the end of period 0, the principal has learned the agent's type and would want to put the agent at his IR level at dates $\tau = 1, \ldots, T$. That is, the principal would want to renege on her commitment to keep the same allocation over time once she had learned the agent's type.

Actually, the ability to commit to a long-term contract that any of the parties (principal or agent) can have enforced by a court if she or he wants to is not sufficient for the optimal static mechanism repeated $T + 1$ times to be feasible, as was demonstrated by Dewatripont (1989). To see this, recall from section 7.3 that $y^*(\cdot)$ involves (under the assumptions made there) a distortion except "at the top" (at $\theta = \bar{\theta}$). The principal trades off efficiency and rent extraction. Now, if at the end of period 0 the principal knows the agent's type to be θ, it is common knowledge that the two parties can improve upon $y^*(\cdot)$ at dates $1, \ldots, T$ to their mutual benefit. They will then renegotiate the initial contract. Thus, the commitment assumption underlying the result that the dynamic allocation is the replicated static one must be taken to mean that the parties not only sign an enforceable long-term contract at date 0 but also can commit never to renegotiate the contract in the future, even if it is in their interest to do so. When the parties cannot

49. The above notation implicitly assumes that $y^*(\cdot)$ is deterministic, but the same reasoning clearly holds when the optimal static allocation is random.

commit not to renegotiate, dynamic mechanism design does not boil down to a static one, and the dynamic equilibrium notions developed in chapter 8 must be employed. Hart and Tirole (1988) and Laffont and Tirole (1990b) show that, in the quasi-linear case, the dynamics of the equilibrium allocation $y_\tau(\cdot)$ coincides with the Coasian dynamics of the durable-good models analyzed in chapter 10.[50]

Besides these two paradigms, "full commitment" and "commitment and renegotiation," economists have considered a third one, called "noncommitment." Suppose that the parties are unable to sign long-term contracts, either for transactional reasons or for legal ones (as is sometimes the case when the principal is a government). One can then consider the repeated version of the three-step game of section 7.3. In each period τ, the principal offers a mechanism $y_\tau(\cdot)$ that applies only to that period.[51] A main issue in such a situation is the "ratchet effect." Suppose for instance that the agent reveals his type in period 0. The continuation game from date 1 on is then a symmetric-information one, and, in the unique subgame equilibrium of this continuation game, the principal offers in each period an allocation that puts the agent at his IR level. Thus, revealing one's type is very costly in a dynamic setting without commitment, and the different types of agent will have a tendency to "pool." We will not give an analysis of the ratchet problem, which requires the tools of dynamic games of incomplete information developed in chapter 8.

7.6.5 Common Agency

In some situations an agent may serve several principals. For example, a distributor may carry the products of several manufacturers, a firm may be regulated by several government agencies, and a consumer may buy from several producers. Martimort (1990) and Stole (1990a) have developed a theory of common agency.[52]

Suppose that there are two principals, A and B. Principal i, $i = A, B$, is interested in decision $x_i \in \mathbb{R}$, and has utility

$$u_i = V_i(x_i, \theta) - t_i.$$

The agent has utility

$$u_1 = V_1(x_A, x_B, \theta) + t_A + t_B.$$

50. The issue of contract renegotiation under asymmetric information also arises in moral-hazard models of the principal-agent relationship. Once the agent has chosen his effort, this effort, if private information, becomes a type for the agent. (See Fudenberg and Tirole 1990.)

51. See Freixas et al. 1985 and Laffont and Tirole 1987b, 1988. See Baron and Besanko 1987 for an approach using a different solution concept.

52. An early example of common agency is found in Baron 1985. Other examples are found in Gal-Or 1989, where the two principals' decisions do not interact in the agent's utility function ($\partial^2 V_1/\partial x_A \partial x_B = 0$), and in Laffont and Tirole 1990c, where the decisions are perfect complements ($\partial^2 V_1/\partial x_A \partial x_B = +\infty$).

A Nash equilibrium in contracts is a pair

$$\{t_A(x_A), t_B(x_B)\},$$

or

$$\{(t_A(\hat{\theta}_A), x_A(\hat{\theta}_A)), (t_B(\hat{\theta}_B), x_B(\hat{\theta}_B))\}$$

where $\hat{\theta}_i$ is the agent's announcement of type to principal i, such that each principal, given the other principal's contract and the agent's optimal reaction to contract offers, maximizes her expected payoff. Note that principal i observes only the report $\hat{\theta}_i$ (or equivalently, the decision x_i) meant for her.

A natural generalization of equation 7.12 to a common-agency differentiable equilibrium (if one exists—see below) is, for all $i = A, B$,

$$\frac{\partial V_i}{\partial x_i} + \frac{\partial V_1}{\partial x_i}$$
$$= \frac{1 - P(\theta)}{p(\theta)} \left(\frac{\partial^2 V_1}{\partial x_i \partial \theta} + \frac{\partial^2 V_1}{\partial x_j \partial \theta} x_j'(\theta) \frac{\dfrac{\partial^2 V_1}{\partial x_i \partial x_j}}{\dfrac{\partial^2 V_1}{\partial x_j \partial \theta} + \dfrac{\partial^2 V_1}{\partial x_i \partial x_j} x_i'(\theta)} \right). \tag{7.56}$$

Equation 7.56 coincides with equation 7.12 except for the second (interaction) term on the right-hand side. When principal i induces an increase dx_i in $x_i(\theta)$, she changes the marginal utility of decision x_j. The resulting change in decision x_j is

$$dx_j = dx_i x_j'(\theta) \frac{\partial^2 V_1}{\partial x_i \partial x_j} \Big/ \left(\frac{\partial^2 V_1}{\partial x_j \partial \theta} + \frac{\partial^2 V_1}{\partial x_i \partial x_j} x_i'(\theta) \right).$$

(To obtain this, differentiate the first-order condition for x_j totally with respect to x_i and $\hat{\theta}_j$ to get an expression for $\partial \hat{\theta}_j / \partial x_i$ and note that $dx_j = x_j'(\theta)(\partial \hat{\theta}_j / \partial x_i)\, dx_i$.) The change dx_i thus has both a direct $((\partial^2 V_1 / \partial x_i \partial \theta)\, dx_i)$ and an indirect $((\partial^2 V_1 / \partial x_j \partial \theta)\, dx_j)$ effect on the rate of growth of the agent's rent, which yields equation 7.56.

Contract complements $(\partial^2 V_1 / \partial x_i \partial x_j > 0)$ lead to a double rent extraction, with the reduction in x_i by principal i making a reduction in x_j more desirable by principal j. The distortion in the decisions thus exceeds that under cooperative contracting by the principals (i.e., that of the single-principal case). In contrast, in a symmetric equilibrium, the decisions lie between the cooperative-contracting decisions and the full-information (or first-best) ones for contract substitutes $(\partial^2 V_1 / \partial x_i \partial x_j < 0)$.

The analysis focuses on finding sufficient conditions for implementability. In the single-principal case, and under the sorting condition, monotonicity is sufficient for local- and global-second-order conditions to be satisfied (theorem 7.3). With two principals, if the agent does not announce his type truthfully to principal i, he may also lie to principal j, and perhaps in a different way. That is, misreporting of θ occurs in a two-dimensional space instead of a single-dimensional one. Martimort and Stole derive sufficient conditions for implementation, and are thus able to prove the existence of a differentiable equilibrium. There is a unique symmetric differentiable equilibrium for contract substitutes and quadratic payoff functions. There is a continuum of symmetric equilibria for contract complements, but the one involving the smallest distortions Pareto dominates the others for the principals and the agent.[53]

Appendix[†††]

What to Do if the Monotonicity Constraint Is Binding

When $x^*(\cdot)$ given by equation 7.12 is not nondecreasing everywhere, one must analyze the full program. There are then two subsets of $[\underline{\theta}, \bar{\theta}]$, both composed of a set of disconnected intervals. In the first subset, the monotonicity constraint is not binding and thus $x(\theta) = x^*(\theta)$. Note that this subset is never empty, because for θ close to $\bar{\theta}$, $p/(1 - P)$ is necessarily increasing.[54] In particular, the "no distortion at the top" result is a general result and does not depend on the monotone-hazard-rate assumption.

In the second subset, the monotonicity constraint is binding and therefore $x(\cdot)$ is constant on each interval in this subset.

We first derive a characterization of the bunching levels, i.e., of decisions x that are chosen by more than one θ. We then sketch an algorithm to obtain the bunching regions. Consider an interval $[\theta_1, \theta_2]$ over which there is "bunching" so that $x(\theta) = \hat{x}$ for all $\theta \in [\theta_1, \theta_2]$, but such that the monotonicity constraint is not binding just outside the interval.

Maximize the principal's expected payoff, and replace the monotonicity constraint by

$$\frac{dx}{d\theta} = \gamma(\theta) \tag{7.57}$$

and

$$\gamma(\theta) \geq 0. \tag{7.58}$$

If $\nu(\theta)$ and $\lambda(\theta)$ denote the shadow prices of equations 7.57 and 7.58, the

53. Another difference with the single-principal case is the treatment of the agent's IR constraint. This treatment depends on whether the agent can accept zero, one, or two contracts (as is the case for a consumer), or whether he can accept zero or two contracts only (as is the case for a regulated firm). For instance, in the second case, the individual transfers for the lowest type, $t_A(\theta)$ and $t_B(\theta)$, are not uniquely defined (but their sum is).

54. Recall that we assumed that p is continuous and strictly positive on the whole interval.

Hamiltonian for program I is then

$$\mathrm{H} = \left(V_0 + V_1 - \frac{1-P}{p}\frac{\partial V_1}{\partial \theta}\right)p + \nu\gamma + \lambda\gamma,$$

where x is taken as a state variable and γ as a control variable. The Pontryagin conditions are

$$\frac{\partial \mathrm{H}}{\partial \gamma} = 0 = \nu + \lambda \tag{7.59}$$

and

$$\frac{d\nu}{d\theta} = -\frac{\partial \mathrm{H}}{\partial x} = -\left(\frac{\partial V_0}{\partial x} + \frac{\partial V_1}{\partial x} - \frac{1-P}{p}\frac{\partial^2 V_1}{\partial x \partial \theta}\right)p. \tag{7.60}$$

Now we exploit the assumption that the monotonicity constraint is not binding at the two boundaries of the interval. Thus, $\nu(\theta_1) = \nu(\theta_2) = 0$, and equation 7.60 can be rewritten as

$$\int_{\theta_1}^{\theta_2} \left(\frac{\partial V_0}{\partial x} + \frac{\partial V_1}{\partial x} - \frac{1-P}{p}\frac{\partial^2 V_1}{\partial x \partial \theta}\right)p\, d\theta = 0. \tag{7.61}$$

That is, the average distortion of the total virtual surplus is equal to 0 over the interval. Together, equation 7.61 and the condition $x^*(\theta_1) = x^*(\theta_2)$ (which results from the boundary conditions $x(\theta_1) = x^*(\theta_1)$ and $x(\theta_2) = x^*(\theta_2)$ and the fact that $x(\theta_1) = x(\theta_2)$) yield two equations with two unknowns. Figure 7.4 depicts the case in which A10 is not satisfied.

Using this characterization of the bunching regions, we now determine where such regions are located. From our assumptions, x^* is continuously differentiable. Let us assume that the curve x^* has a finite number of interior peaks on $[\underline{\theta}, \overline{\theta}]$.

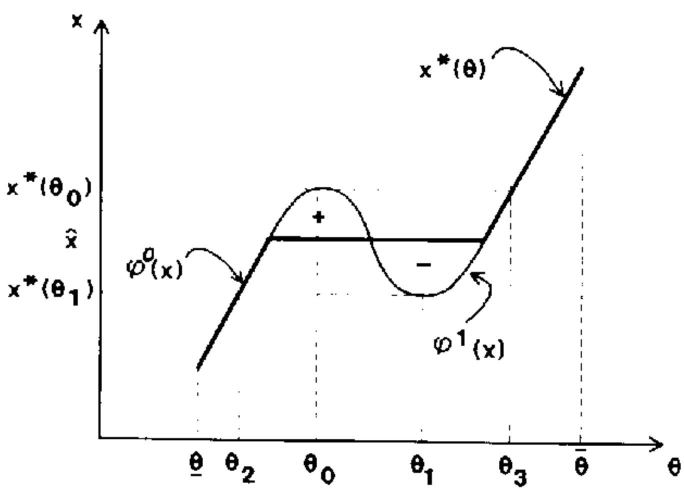


Figure 7.4

If there is no interior peak, x^* is nondecreasing (recall that $x^*(\bar{\theta}) \geq x^*(\theta)$ for all θ) and is therefore the solution to program I. If there is a single interior peak θ_0, then there is also a single interior trough θ_1 (see figure 7.4). The inverse image of the interval $[x^*(\theta_1), x^*(\theta_0)]$ is composed of two intervals, $[\theta_2, \theta_0]$ and $[\theta_1, \theta_3]$, over which $x^*(\cdot)$ is increasing (if there is no $\theta_2 < \theta_0$ such that $x^*(\theta_2) = x^*(\theta_1)$, let $\theta_2 \equiv \underline{\theta}$), and one interval, $[\theta_0, \theta_1]$, over which $x^*(\cdot)$ is decreasing. Let $\varphi^0(x)$ and $\varphi^1(x)$ denote the inverse functions of x over the intervals $[\theta_2, \theta_0]$ and $[\theta_1, \theta_3]$. Last, for each $x \in [x^*(\theta_1), x^*(\theta_0)]$, define

$$\Delta(x) \equiv \int_{\varphi^0(x)}^{\varphi^1(x)} \left(\frac{\partial V_0}{\partial x}(x, \theta) + \frac{\partial V_1}{\partial x}(x, \theta) - \frac{1 - P(\theta)}{p(\theta)} \frac{\partial^2 V_1}{\partial x \partial \theta}(x, \theta) \right) d\theta.$$

Note that at $x = x^*(\theta_0)$, $\varphi^0(x) = \theta_0$ and $\varphi^1(x) = \theta_3$ and $\Delta(x) < 0$ as $x > x^*(\theta)$ for all $\theta \in (\theta_0, \theta_3)$ and the objective function

$$V_0 + V_1 - \frac{1 - P}{p} \frac{\partial^2 V_1}{\partial x \partial \theta}$$

is strictly concave in x. Similarly, at $x = x^*(\theta_1)$, $\varphi^0(x) = \theta_2$ and $\varphi^1(x) = \theta_1$, and if $\theta_2 > \underline{\theta}$, $\Delta(x) > 0$ as $x < x^*(\theta)$ for all $\theta \in (\theta_2, \theta_1)$. Furthermore, with x optimal at $\varphi^0(x)$ and $\varphi^1(x)$,

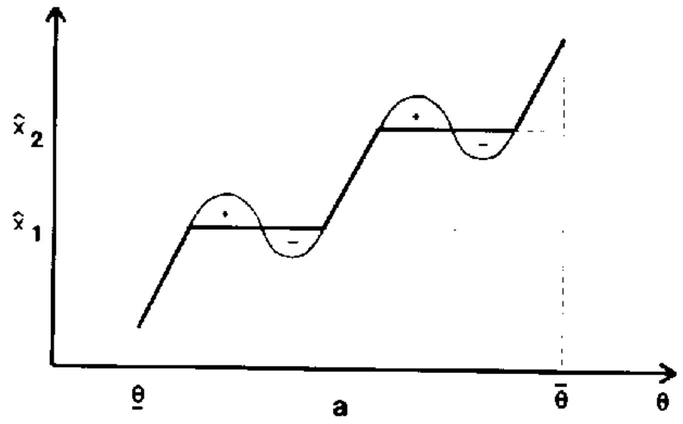
$$\Delta'(x) = \int_{\varphi^0(x)}^{\varphi^1(x)} \left(\frac{\partial^2 V_0}{\partial x^2}(x, \theta) + \frac{\partial^2 V_1}{\partial x^2}(x, \theta) - \frac{1 - P(\theta)}{p(\theta)} \frac{\partial^3 V_1}{\partial x^2 \partial \theta} \right) d\theta < 0.$$

If $\theta_2 > \theta$, then the intermediate-value theorem shows that there exists a (unique) $\hat{x} \in [x^*(\theta_1), x^*(\theta_0)]$ such that $\Delta(\hat{x}) = 0$. From our previous characterization, the bunching interval is $[\varphi^0(\hat{x}), \varphi^1(\hat{x})]$, so the solution is $x^*(\theta)$ for $\theta \notin [\varphi^0(\hat{x}), \varphi^1(\hat{x})]$ (see the bold curve in figure 7.4).[55]

Now suppose there are two interior peaks. Intuitively, if we can independently design two bunching levels $\hat{x}_1$ and $\hat{x}_2$ as in figure 7.5a, such that $\hat{x}_1 \leq \hat{x}_2$ and $\hat{x}_1, \hat{x}_2$ and the associated boundaries of the two bunching intervals satisfy the property that the average distortion over each bunching interval is equal to 0, we have the solution (represented by the bold curve in figure 7.5a). If treating the two bunching regions separately yields $\hat{x}_1 > \hat{x}_2$, the resulting decision schedule is not monotonic and therefore not incentive compatible (see the broken segments in figure 7.5b). We must then merge the two into a single bunching interval at some level $\hat{x}_3$ such that the average distortion over the interval $[\theta_5, \theta_6]$ in figure 7.5b is equal to 0.

55. If $\theta_2 = \underline{\theta}$, there may or may not exist such an $\hat{x}$. More precisely, if $\Delta(x^*(\underline{\theta})) \geq 0$, there exists such an $\hat{x}$ and the answer is as above. If $\Delta(x^*(\underline{\theta})) < 0$, then the bunching interval is $[\theta, \theta_4]$, where $\theta_4 \in [\theta_1, \theta_3]$ and where

$$\int_{\theta}^{\theta_4} \left(\frac{\partial V_0}{\partial x} + \frac{\partial V_1}{\partial x} - \frac{1 - P}{p} \frac{\partial^2 V_1}{\partial x \partial \theta} \right) d\theta = 0.$$



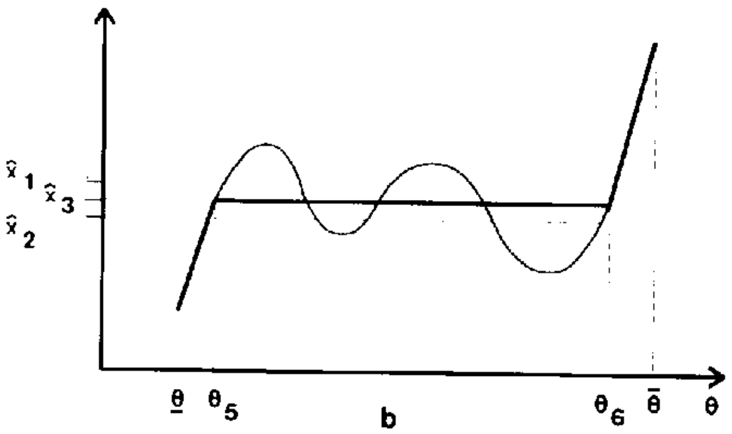


Figure 7.5

We leave it to the reader to construct an algorithm to obtain the solution with more than two peaks.

Assumption A9 implies that the constraint $x(\theta) \leq \bar{x}$ is never binding. First, monotonicity implies that $x(\theta) \leq x(\bar{\theta})$ for all θ. Second, we saw that there is no distortion at the top, so that $x(\bar{\theta}) = x^*(\bar{\theta})$. But $x^*(\bar{\theta}) \leq \bar{x}$ from A9.

When Is It Legitimate to Focus on Deterministic Mechanisms?

We restricted our attention to mechanisms in which the decision x and the transfer t are deterministic functions of the announced type $\hat{\theta}$. More generally, one can allow x and t to have random values $\tilde{x}(\hat{\theta})$ and $\tilde{t}(\hat{\theta})$. It is clear that with quasi-linear utilities there is no gain to be had from introducing random transfers, as the principal and the agent care only about the expectation $t(\hat{\theta}) \equiv \mathscr{E}\tilde{t}(\hat{\theta})$. (In this discussion, the expectations are with respect to the random variable underlying the stochastic allocation, and not with respect to type. To distinguish between the two, we denote the new expectations by $\mathscr{E}(\cdot)$.) Thus, only random decisions need be considered.

In many applications, the functions V_0 and V_1 are concave in x, which we assume in the following discussion. Then, V_0 and V_1 can be increased by replacing the random variable $\tilde{x}$ by its expectation $x(\hat{\theta}) \equiv \mathscr{E}\,\tilde{x}(\hat{\theta})$. Increasing V_0 benefits the principal directly; increasing V_1 helps her indirectly by allowing her to reduce the agent's income. Thus, if there is any benefit to introducing randomness in the decision, it must be the case that the randomness relaxes the incentive constraint. Recall that the incentive constraint can be expressed by the speed at which the agent's rent or utility increases with his type (together with the condition that the decision be monotonic in the agent's type, if the sorting condition holds). For a random scheme, the envelope theorem yields

$$\dot{U}_1(\theta) = \mathscr{E}\left[\frac{\partial V_1}{\partial \theta}(\tilde{x}(\theta), \theta)\right].$$

Suppose, for instance, that u_1 increases with θ. Then, to minimize the slope of the $U_1(\cdot)$ function, the principal wants to minimize $\mathscr{E}[\partial V_1/\partial\theta]$. If $\partial V_1/\partial\theta$ is convex in x ($\partial^3 V_1/\partial\theta\partial x^2 \geq 0$, which is part of assumption A8), Jensen's inequality implies that

$$\mathscr{E}\left[\frac{\partial V_1}{\partial \theta}(\tilde{x}(\theta), \theta)\right] \geq \frac{\partial V_1}{\partial \theta}(\mathscr{E}(\tilde{x}(\theta)), \theta) = \frac{\partial V_1}{\partial \theta}(x(\theta), \theta).$$

That is, $\dot{U}_1(\theta)$ can be reduced by using the deterministic decision $x(\theta)$ instead of the random decision $\tilde{x}(\theta)$. Because random schemes reduce V_0 and V_1, and raise $\dot{U}_1$, they yield less utility to the principal:

$$\begin{aligned} &E_\theta\left[\mathscr{E}\,V_0(\tilde{x}(\theta), \theta) + \mathscr{E}\,V_1(\tilde{x}(\theta), \theta) - \int_{\underline{\theta}}^{\theta} \mathscr{E}\,\frac{\partial V_1}{\partial \eta}(\tilde{x}(\eta), \eta)\,d\eta\right] \\ &\quad \leq E_\theta\left[V_0(\mathscr{E}(\tilde{x}(\theta)), \theta) + V_1(\mathscr{E}(\tilde{x}(\theta)), \theta) - \int_{\underline{\theta}}^{\theta} \frac{\partial V_1}{\partial \eta}(\mathscr{E}(\tilde{x}(\eta)), \eta)\,d\eta\right]. \end{aligned}$$

Turning things around, transforming a deterministic decision $x(\theta)$ into a random one $\tilde{x}(\theta)$ with the same mean for each θ reduces the principal's welfare. We thus conclude that if the agent's incentive-compatibility constraint for the deterministic allocation is fully characterized by the equation $\dot{U}_1(\theta) = \partial V_1(x(\theta), \theta)/\partial\theta$, as it is under the assumptions of theorem 7.4, the principal cannot gain by using a random mechanism.

In contrast, if $\partial V_1/\partial\theta$ is strictly concave in x (that is, $\partial^3 V_1/\partial\theta\partial x^2 < 0$), the principal can reduce the agent's rent by using stochastic decisions. The principal must then trade off the costs (the reduction in efficiency, i.e., in $V_0 + V_1$) and the benefits (the reduction in the agent's rent U_1) of random schemes. For more on random mechanisms, see Maskin 1981.[56]

56. Maskin and Riley (1984b) give a sufficient condition for random incentive schemes not to be optimal in the case of non-quasi-linear utilities.

Exercises

Exercise 7.1** Maskin and Riley (1984b) and Matthews (1983) show that with risk-averse bidders an optimal auction may require payments from the bidders even when they lose. The idea is that the seller can exploit the difference in the marginal utilities of income when a bidder wins or loses. Suppose that there are two bidders, $i = 1, 2$. Each bidder can have one of two valuations: $\underline{\theta}$ (with probability $\underline{p}$) or $\bar{\theta}$ (with probability $\bar{p}$), where $0 \le \underline{\theta} < \bar{\theta}$. Let $\underline{W}$ and $\underline{L}$ denote the transfers to the seller when the buyer wins or loses and the bidder has announced $\underline{\theta}$ (define $\bar{W}$ and $\bar{L}$ similarly for type $\bar{\theta}$). A bidder has utility $u(\theta - W)$ when winning and paying W, and $u(-L)$ when losing and paying L. Solve for the optimal symmetric auction and show that, at the optimum, $\bar{L} < 0$ (and the $\bar{\theta}$ type is perfectly insured), and $\underline{L} > 0$. (Hint: Proceed as in section 7.1 in your selection of IR and IC constraints. Letting $\bar{X}$ and $\underline{X}$ denote the probabilities of getting the good when $\bar{\theta}$ and $\underline{\theta}$, note that $\frac{1}{2} \ge \underline{p}\underline{X} + \bar{p}\bar{X}$ and $\bar{X} \le \underline{p} + \bar{p}/2$. Solve and check that the ignored constraints are satisfied.)

Exercise 7.2** Consider the problem of inducing firms to reveal information about their cost of reducing their pollution.

(a) Take the case of a single firm. The damage created by an amount of pollution x is $D(x)$. The production cost for the firm is $C(x, \theta)$, where θ is a private-information parameter and C is decreasing in x. Assume that $C_\theta > 0$, $C_{x\theta} < 0$, $C_{xx} \ge 0$, $D'' \ge 0$, and $C_{\theta xx} \ge 0$. Show that if the government has coercive power, it can obtain the socially optimal amount of pollution $x^*(\theta)$ by giving the firm a transfer equal to a constant minus the damage cost $D(x)$. How does this scheme link with the Groves scheme in section 7.3?

(b) Still in the single-firm context, suppose that the firm can refuse to participate (it has property rights and is free to pollute if it wants to). Can the first-best outcome still be implemented if the government cares about the sum of consumer surplus and producer profit? Next, suppose that the government faces a shadow cost of public funds $\lambda > 0$, so that its objective function is

$$W = \quad D(x) - (1 + \lambda)t + (t - C(x, \theta))$$

(up to a constant). Derive the optimal incentive scheme (Note: The IR level may be type dependent. Perform the analysis as if it were type independent and check *ex post* that everything is fine.)

(c) Assume there are I firms, with production costs $C_i(x_i, \theta_i)$, and that total damage is $D(x_1, \ldots, x_I)$. Coming back to question a's assumption that the government faces no individual rationality constraint, derive a d'Aspremont Gérard-Varet scheme for this model.