# Chapter 2
# Foundations of Mechanism Design

This chapter forms the first part of the monograph and presents key concepts and results in mechanism design. The second part of the monograph explores application of mechanism design to contemporary problems in network economics. The chapter comprises 21 sections that can be logically partitioned into four groups. Sections 2.1 through 2.5 constitute Group 1, and they set the stage by describing essential aspects of game theory for understanding mechanism design. The five sections deal with strategic form games, dominant strategy equilibria, pure strategy Nash equilibria, mixed strategy Nash equilibria, and Bayesian games. Sections 2.6 through 2.12 constitute the next group of sections, and they deal with fundamental notions and results of mechanism design. The sections include a description of the mechanism design environment, social choice functions, implementation of social choice functions by mechanisms, incentive compatibility and revelation theorem, properties of social choice functions, the Gibbard–Satterthwaite impossibility theorem, and the Arrow's impossibility theorem. Following this, the sections in the third group (Sections 2.13 - 2.20) present useful mechanisms that provide the building blocks for solving mechanism design problems. The sections here include: The quasilinear environment, Groves mechanisms, Clarke mechanisms, examples of VCG mechanisms, the dAGVA mechanism, Bayesian mechanisms in linear environment, revenue equivalence theorem, and optimal auctions. Finally, in Section 2.21, we provide a sprinkling of further key topics in mechanism design. The chapter uses a fairly large number of stylized examples of network economics situations to illustrate the notions and the results.

## 2.1 Strategic Form Games

Game theory may be defined as the study of mathematical models of interaction between rational, intelligent decision makers [1]. The interaction may include both *conflict* and *cooperation*. The theory provides general mathematical techniques for analyzing situations in which two or more individuals (called players or agents) make decisions that influence one another's welfare. There are many categories of games that have been proposed and discussed in game theory. We introduce here a class of games called *strategic form games* or *normal form games*, which are most appropriate for the discussions in this monograph. We start with the definition of a strategic form game.

**Definition 2.1 (Strategic Form Game).** A strategic form game $\Gamma$ is defined as a tuple $\langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$, where $N = \{1, 2, \ldots, n\}$ is a finite set of players; $S_1, S_2, \ldots, S_n$ are the strategy sets of the players $1, \ldots, n$, respectively; and $u_i : S_1 \times S_2 \times \cdots \times S_n \to \mathbb{R}$ for $i = 1, 2, \ldots, n$ are mappings called the utility functions or payoff functions.

The strategies are also called *actions* or *pure strategies*. We denote by $S$, the Cartesian product $S_1 \times S_2 \times \cdots \times S_n$. The set $S$ is the collection of all strategy profiles of the players. Note that the utility of an agent depends not only on its own strategy but also on the strategies of the rest of the agents. Every profile of strategies induces an *outcome* in the game. A strategic form game is said to be *finite* if $N$ and all the strategy sets $S_1, \ldots, S_n$ are finite.

The idea behind a strategic form game is to capture each agent's decision problem of choosing a strategy that will counter the strategies adopted by the other agents. Each player is faced with this problem and therefore the players can be thought of as simultaneously choosing their strategies from the respective sets $S_1, S_2 \ldots, S_n$. We can view the play of a strategic game as follows: each player simultaneously writes down a chosen strategy on a piece of paper and hands it over to a referee who then computes the outcome and the utilities. Several examples of strategic form games will be presented in Section 2.1.2.

## *2.1.1 Key Notions*

There are certain key notions underlying game theory. We discuss these notions and a few related issues.

### 2.1.1.1 Utilities

Utility theory enables the preferences of the players to be expressed in terms of payoffs in some utility scale. Utility theory is the science of assigning numbers to outcomes in a way that reflects the preferences of the players. The theory is an important contribution of von Neumann and Morgenstern, who stated and proved in [2] a crucial theorem called the *expected utility maximization theorem*. This theorem establishes for any rational decision maker that there must exist a way of assigning utility numbers to different outcomes in a way that the decision maker would always choose the option that maximizes his expected utility. This theorem holds under quite weak assumptions about how a rational decision maker should behave.

**John von Neumann** (1903 - 1957) is regarded as one of the foremost mathematicians of the 20th century. In particular, he is regarded as the founding father of game theory. He was born in Budapest, Hungary on December 28, 1903. He was a mathematical genius from early childhood, but interestingly his first major degree was in chemical engineering from the Swiss Federal Institute of Technology in Zurich. In 1926, he earned a Doctorate in Mathematics from the University of Budapest, working with Professor Leopold Fezer. During 1926 to 1930, he taught in Berlin and Hamburg, and from 1930 to 1933, he taught at Princeton University. In 1933, he became the youngest of the six professors of the School of Mathematics at the Institute of Advanced Study in Princeton. Albert Einstein and Godël were his colleagues at the center. During his glittering scientific career, von Neumann created several intellectual currents, two of the major ones being game theory and computer science. The fact that these two disciplines have converged during the 1990s and 2000s, almost sixty years after von Neumann brilliantly created them, is a testimony to his visionary genius. In addition to game theory and computer science, he made stunning contributions to a wide array of disciplines including set theory, functional analysis, quantum mechanics, ergodic theory, continuous geometry, numerical analysis, hydrodynamics, and statistics. He is best known for his minimax theorem, utility theory, von Neumann algebras, von Neumann architecture, and cellular automata.

In game theory, von Neumann's first significant contribution was the minimax theorem, which proves the existence of a randomized saddle point in two player zero sum games. His collaboration with Oskar Morgenstern at the Center for Advanced Study resulted in the classic book *The Theory of Games and Economic Behavior*, which to this day continues to be an excellent source for early game theory results. This, classic work develops many fundamental notions of game theory such as utilities, saddle points, coalitional games, bargaining sets, etc.

Von Neumann was associated with the development of the first electronic computer in the 1940s. He wrote a widely circulated paper entitled the First Draft of a Report on the EDVAC in which he described a computer architecture (which is now famously called the von Neumann architecture). He is also credited with the development of the notions of a computer algorithm and algorithm complexity.

#### 2.1.1.2 Rationality

The first key assumption in game theory is that the players are rational. An agent is said to be rational if the agent makes decisions consistently in pursuit of its own objectives. It is assumed that each agent's objective is to maximize the expected value of its own payoff measured in some utility scale. The above notion of rationality (maximization of expected utility) was initially proposed by Bernoulli (1738) and later formalized by von Neumann and Morgenstern (1944) [2]. A key observation here would be that *self-interest* is essentially an implication of rationality.

It is important to note that maximizing expected utility is not necessarily the same as maximizing expected monetary returns. For example, a given amount of money may have significant utility to a person desperately in need of the money. The same amount of money may have much less utility to a person who is already rich. In general, utility and money are nonlinearly related.

When there are two or more players, it would be the case that the rational solution to each player's decision problem depends on the others' individual problems and vice-versa. None of the problems may be solvable without understanding the solutions of the other problems. When such rational decision makers interact, their decision problems must be analyzed together, like a system of simultaneous equations. Game theory, in a natural way, deals with such analysis.

#### 2.1.1.3 Intelligence

Another key notion in game theory is that of intelligence of the players. This notion connotes that each player in the game knows everything about the game that a game theorist knows, and the player can make any inferences about the game that a game theorist can make. In particular, an intelligent player is *strategic*, that is, would fully take into account his knowledge or expectation of behavior of other agents in determining what his best response strategy should be. Each player is assumed to have enough computational resources to perform the required computations involved in determining a best response strategy.

Myerson [1] and several other authors provide the following convincing explanation to show that the assumptions of rationality and intelligence are indeed reasonable. The assumption that all individuals are rational and intelligent may not exactly be satisfied in a typical real-world situation. However, any theory that is not consistent with the assumptions of rationality and intelligence is fallible for the following reason: If a theory predicts that some individuals will be systematically deceived into making mistakes, then such a theory will lose validity when individuals learn through mistakes to understand the situations better. On the other hand, a theory based on rationality and intelligence assumptions would be credible.

#### 2.1.1.4 Common Knowledge

The notion of common knowledge is an important implication of *intelligence*. Aumann (1976) [3] defined *common knowledge* as follows: A fact is common knowledge among the players if every player knows it, every player knows that every player knows it, and so on. That is, every statement of the form "every player knows that every player knows that $\cdots$ every player knows it" is true ad infinitum. If it happens that a fact is known to all the players, without the requirement of all players knowing that all players know it, etc., then such a fact is called *mutual knowledge*. In game theory, analysis often requires the assumption of common knowledge to be true; however, sometimes, the assumption of mutual knowledge suffices for the analysis. A player's *private information* is any information that the player has that is not common knowledge among all the players.

**Robert Aumann** is a versatile game theorist who has stamped his authority with creative contributions in a wide range of topics in game theory such as repeated games, correlated equilibria, bargaining theory, cooperative game theory, etc. It was Aumann, who provided in 1976 [3], a convincing explanation of the notion of common knowledge in game theory, in a classic paper entitled *Agreeing to disagree* (which appeared in the Annals of Statistics). Aumann's work in the 1960s on repeated games clarified the difference between infinitely and finitely repeated games. With Bezalel Peleg in 1960, Aumann formalized the notion of a coalitional game without transferable utility (NTU), a significant result in cooperative game theory. With Michael Maschler (1963) he introduced the concept of a *bargaining set*, an important solution concept in cooperative game theory. In 1974, Aumann went on to define and formalize the notion of *correlated equilibrium* in Bayesian games. In 1975, Aumann proved a convergence theorem for the Shapley value. In 1976, in an unpublished paper with Lloyd Shapley, Aumann provided the perfect folk theorem using the limit of means criterion. All of these contributions have advanced game theory in significant ways. His books on *Values of Non-Atomic Games* (1984) co-authored with Lloyd Shapley and on *Repeated Games with Incomplete Information* (1995) co-authored with Michael Maschler are considered game theory classics.

Aumann was born in Frankfurt am Main, Germany on June 8, 1930. He earned an MSc Degree in Mathematics in 1952 from the Massachusetts Institute of Technology where he also received his Ph D Degree in 1955. His doctoral adviser at MIT was Professor George Whitehead Jr. and his doctoral thesis was on knot theory. He has been a professor at the Center for Rationality in the Hebrew University of Jerusalem, Israel, since 1956, and he also holds a visiting appointment with Stonybrook University, USA.

Robert Aumann and Thomas Schelling received the 2005 Nobel Prize in Economic Sciences for their contributions toward a clear understanding of conflict and cooperation through game theory analysis.

The intelligence assumption means that whatever a game theorist may know or understand about the game must be known or understood by the players of the game. Thus the model of the game is also known to the players. Since all the players know the model and they are intelligent, they also know that they all know the model. Thus the model is common knowledge.

In a *strategic form game with complete information*, the set $N$, the strategy sets $S_1, \ldots, S_n$, and the utility functions $u_1, \ldots, u_n$ are common knowledge, that is every player knows them, every player knows that every player knows them, etc. We will be studying strategic form games with complete information in this and the next three sections. We will study games with incomplete information in Section 2.5.

### 2.1.2 Examples of Strategic Form Games

We now provide several examples of game theoretic situations and formulate them as strategic form games.

*Example 2.1 (Matching Companies Game).* This example is developed on the lines of the famous matching pennies game, where there are two players who simultaneously put down a coin each, heads up or tails up. Each player is unaware of the move made by the other. If the two coins match, player 1 wins; otherwise, player 2 wins. In the version that we develop here, there are two companies, call them 1 and 2. Each company is capable of producing two products A and B, but at any given time, a company can only produce one product, owing to high setup and switchover costs. Company 1 is known to produce superior quality products but company 2 scores over company 1 in terms of marketing power and advertising innovations.

| | 2 | |
|---|---|---|
| 1 | A | B |
| A | +1, −1 | −1, +1 |
| B | −1, +1 | +1, −1 |

**Table 2.1** Payoff matrix for the matching companies game

If both the companies produce the same product (A or B), it turns out that company 1 makes all the profits and company 2 loses out, because of the superior quality of products produced by company 1. This is reflected in our model with a payoff of $+1$ for company 1 and a payoff of $-1$ for company 2, corresponding to the strategy profiles (A,A) and (B,B).

On the other hand, if one company produces product A and the other company produces product B, it turns out that because of the marketing skills of company 2 in differentiating the product offerings A and B, company 2 captures all the market, resulting in a payoff of $+1$ for company 2 and a payoff of $-1$ for company 1.

The two companies have to simultaneously decide (each one does not know the decision of the other) which product to produce. This is the strategic decision facing the two companies. This situation is captured by a strategic form game $\Gamma = \langle N, S_1, S_2, u_1, u_2 \rangle$, where $N = \{1,2\}$; $S_1 = S_2 = \{A, B\}$, and the utility functions are as described in Table 2.1.

*Example 2.2 (Battle of Companies Game).* This game is developed on the lines of the famous Battle of Sexes problem. Consider two companies, 1 and 2. As in Example 2.1, each company can produce two products A and B, but at any given time, a company can only produce one type of product, owing to high setup and switchover costs. The products A and B are competing products. Product A is a niche product of company 1 while product B is a niche product of company 2. If both the companies produce product A the consumers are compelled to buy product A and would naturally prefer to buy it from company 1 rather than from 2. Assume that company 1 will capture two thirds of the market. We will reflect this fact by saying that the payoff to company 1 is twice as much as for company 2. If both the companies produce product B, the reverse situation will prevail and company 2 will make twice as much payoff as company 1.

On the other hand, if the two companies decide to produce different products, then the market gets segmented, and each company tries to outwit the other through increased spending on advertising. In fact, their competition may actually benefit a third company, and effectively, neither of the original companies 1 or 2 makes any payoff. Table 2.2 depicts the payoff structure for this game.

| | 2 | |
|---|---|---|
| 1 | A | B |
| A | 2, 1 | 0,0 |
| B | 0,0 | 1,2 |

**Table 2.2** Payoff matrix for the battle of companies game

*Example 2.3 (Company's Dilemma Problem).* This game is modeled on the lines of the popular prisoner's dilemma problem. Here again, we have two companies 1 and 2, each of which can produce two competing products A and B, but only one at a time. The companies are known for product A rather than for product B. Environmentalists have launched a negative campaign on product A branding it as non-eco friendly.

If both the companies produce product A, then, in spite of the negative campaign, their payoff is quite high since product A happens to be a niche product of both the companies. On the other hand, if both the companies produce product B, they still make some profit, but not as much as they would if they both produced product A.

On the other hand, if one company produces product A and the other company produces product B, then because of the negative campaign about product A, the company producing product A makes zero payoff while the other company captures all the market and makes a high payoff.

Table 2.3 depicts the payoff structure for this game. In our next example, we describe the classical *prisoner's dilemma problem*.

| | 2 | |
|---|---|---|
| 1 | A | B |
| A | 6,6 | 0,8 |
| B | 8,0 | 3,3 |

**Table 2.3** Payoff matrix for the company's dilemma problem

*Example 2.4 (Prisoner's Dilemma Problem).* This is one of the most extensively studied problems in game theory, with many interesting interpretations cutting across disciplines. Two individuals are arrested for allegedly committing a crime and are lodged in separate prisons. The district attorney interrogates them separately. The attorney privately tells each prisoner that if he is the only one to confess, he will get a light sentence of 1 year in jail while the other would be sentenced to

10 years in jail. If both players confess, they would get 5 years each in jail. If neither confesses, then each would get 2 years in jail. The attorney also informs each prisoner what has been told to the other prisoner. Thus the payoff matrix is common knowledge. See Table 2.4.

| | 2 | |
|---|---|---|
| 1 | NC | C |
| NC | −2, −2 | −10, −1 |
| C | −1, −10 | −5, −5 |

**Table 2.4** Payoff matrix for the prisoner's dilemma problem

How would the prisoners behave in such a situation? They would like to play a strategy that is best response to a (best) response strategy that the other player may adopt, the latter player also would like to play a best response to the other player's best response, and so on. First observe that C is each player's best strategy regardless of what the other player plays:

$$u_1(C,C) > u_1(NC,C); \quad u_1(C,NC) > u_1(NC,NC)$$

$$u_2(C,C) > u_2(C,NC); \quad u_2(NC,C) > u_2(NC,NC)$$

Thus (C,C) is a natural prediction for this game. However, the outcome (NC, NC) is the best outcome jointly for the players. Prisoner's Dilemma is a classic example of a game where rational, intelligent behavior does not lead to a socially optimal result. Also, each prisoner has a negative effect or externality on the other. When a prisoner moves away from (NC, NC) to reduce his jail term by 1 year, the jail term of the other player increases by 8 years.

*Example 2.5 (A Sealed Bid Auction).* There is a seller who wishes to allocate an indivisible item to one of $n$ prospective buyers in exchange for a payment. Here, $N = \{1,2,\ldots,n\}$ represents the set of buying agents. Let $v_1, v_2, \ldots, v_n$ be the valuations of the players for the object. The $n$ buying agents submit sealed bids and these bids need not be equal to the valuations. Assume that the sealed bid from player $i$ ($i = 1,\ldots,n$) could be any real number greater than 0. Then the strategy sets of the players are: $S_i = (0,\infty)$ for $i = 1,\ldots,n$. Assume that the object is awarded to the agent with the lowest index among those who bid the highest. Let $b_1,\ldots,b_n$ be the bids from the $n$ players. Then the allocation function will be:

$$\begin{aligned} y_i(b_1,\ldots,b_n) &= 1 \text{ if } \quad b_i > b_j \text{ for } j = 1,2,\ldots,i-1 \text{ and} \\ &\qquad\qquad b_i \geq b_j \text{ for } j = i+1,\ldots,n \\ &= 0 \text{ else} \end{aligned}$$

In the first price sealed bid auction, the winner pays an amount equal to his bid, and the losers do not pay anything. In the second price sealed bid auction, the winner pays an amount equal to the highest bid among the players who do not win, and as usual the losers do not pay anything. The payoffs or utilities to the bidders in these two auctions are of the form:

$$u_i(b_1,\ldots,b_n) = y_i(b_1,\ldots,b_n)(v_i - t_i(b_1,\ldots,b_n))$$

where $t_i(b_1,\ldots,b_n)$ is the amount to be paid by bidder $i$ in the auction. Suppose $n = 4$, and suppose the values are $v_1 = 20$; $v_2 = 20$; $v_3 = 16$; $v_4 = 16$, and the bids are $b_1 = 10$; $b_2 = 12$; $b_3 = 8$; $b_4 = 14$. Then for both first price and second price auctions, we have the allocation $y_1(.) = 0$; $y_2(.) = 0$; $y_3(.) = 0$; $y_4(.) = 1$. The payments for the first price auction are $t_1(.) = 0$; $t_2(.) = 0$; $t_3(.) = 0$; $t_4(.) = 14$ whereas the payments for the second price auction would be: $t_1(.) = 0$; $t_2(.) = 0$; $t_3(.) = 0$; $t_4(.) = 12$. The utilities can be easily computed from the values and the payments.

An important question is: What will the strategies of the bidders be in these two auctions. This question will be discussed at length in forthcoming sections.

*Example 2.6 (A Bandwidth Sharing Game).* This problem is based on an example presented by Tardos and Vazirani [4]. There is a shared communication channel of maximum capacity 1. There are $n$ users of this channel, and user $i$ wishes to send $x_i$ units of flow, where $x_i \in [0,1]$. We have

$$N = \{1,2,\ldots,n\}$$
$$S_1 = S_2 = \ldots = S_n = [0,1].$$

If $\sum_{i\in N} x_i \geq 1$, then the transmission cannot happen since the capacity is exceeded, and the payoff to each player is zero. If $\sum_{i\in N} x_i < 1$, then assume that the following is the payoff to user $i$:

$$u_i = x_i(1 - \sum_{j\in N} x_j)$$

The above expression models the fact that the payoff to a player is proportional to the flow sent by the player but is negatively impacted by the total flow. The second term captures the fact that the quality of transmission deteriorates with the total bandwidth used. The above defines an $n$-player infinite game.

*Example 2.7 (A Duopoly Pricing Game).* This game model is due to Bertrand (1883) [5]. Bertrand competition is a model of competition in a duopoly (that is an economic environment with two competing companies), named after Joseph Louis Franois Bertrand (1822-1900). There are two companies 1 and 2 that produce homogeneous products and that do not cooperate. The companies obviously wish to maximize their profits. The quantity demanded for the product as a function of price $p$ is given by a continuous and strictly decreasing function $x(p)$. The cost for producing each unit of product $= c > 0$. The companies simultaneously choose their prices $p_1$ and $p_2$ and compete solely on price. We can see that the amount of sales for each company is given by:

$$
\begin{aligned}
x_1(p_1,p_2) &= x(p_1) && \text{if } p_1 < p_2\\
&= \tfrac{1}{2}x(p_1) && \text{if } p_1 = p_2\\
&= 0 && \text{if } p_1 > p_2
\end{aligned}
$$

$$
\begin{aligned}
x_2(p_1,p_2) &= x(p_2) && \text{if } p_2 < p_1\\
&= \tfrac{1}{2}x(p_2) && \text{if } p_2 = p_1\\
&= 0 && \text{if } p_2 > p_1.
\end{aligned}
$$

It is assumed that the companies incur production costs only for an output level equal to their actual sales. Given prices $p_1$ and $p_2$, the utilities of the two companies would be:

$$
\begin{aligned}
u_1(p_1,p_2) &= (p_1 - c)\, x_1(p_1,p_2)\\
u_2(p_1,p_2) &= (p_2 - c)\, x_2(p_1,p_2).
\end{aligned}
$$

Note that for this game, $N = \{1,2\}$ and $S_1 = S_2 = (0,\infty)$.

*Example 2.8 (A Procurement Exchange Game).* This example is adapted from an example presented by by Tardos and Vazirani [4]. Imagine a procurement exchange where buyers and sellers meet to match supply and demand for a particular product. Suppose that there are two sellers 1 and 2 and three buyers $A$, $B$, and $C$. Because of certain constraints such as logistics, assume that

- $A$ can only buy from seller 1.
- $C$ can only buy from seller 2.
- $B$ can buy from either seller 1 or seller 2.
- Each buyer has a maximum willingness to pay of 1 and wishes to buy one item.
- The sellers have enough items to sell.
- Each seller announces a price in the range $[0,1]$.

Let $s_1$ and $s_2$ be the prices announced. It is easy to see that buyer $A$ will buy an item from seller 1 at price $s_1$ and buyer $C$ will buy an item from seller 2 at price $s_2$. If $s_1 < s_2$, then buyer $B$ will buy an item from seller 1; otherwise buyer $B$ will buy from seller 2. Assume that buyer $B$ will buy from seller 1 if $s_1 = s_2$. The game can now be defined as follows:

$$
\begin{aligned}
N &= \{1,2\}\\
S_1 &= S_2 = [0,1]\\
u_1(s_1,s_2) &= 2s_1 \;\text{ if } s_1 \le s_2\\
&= s_1 \;\text{ if } s_1 > s_2\\
u_2(s_1,s_2) &= 2s_2 \;\text{ if } s_1 > s_2\\
&= s_2 \;\text{ if } s_1 \le s_2.
\end{aligned}
$$

We now start analyzing strategic form games by looking at their equilibrium behavior. First we discuss the notion of dominant strategy equilibria. Next we introduce the notion of Nash equilibrium. The notation we use is summarized in Table 2.5.

| | |
|---|---|
| $N$ | A set of players, $\{1,2,\ldots,n\}$ |
| $S_i$ | Set of actions or pure strategies of player $i$ |
| $S$ | Set of all action profiles $S_1 \times \ldots \times S_n$ |
| $s$ | A particular action profile, $s = (s_1,\ldots,s_n) \ \in S$ |
| $S_{-i}$ | Set of action profiles of all agents other than $i = S_1 \times \ldots S_{i-1} \times S_{i+1} \times \ldots \times S_n$ |
| $s_{-i}$ | A particular action profile of agents other than $i$, $s_{-i} = (s_1,\ldots,s_{i-1},s_{i+1},\ldots,s_n) \ \in S_{-i}$ |
| $(s_i,s_{-i})$ | Another representation for strategy profile $(s_1,\ldots,s_n)$ |
| $s_i^*$ | Equilibrium strategy of player $i$ |
| $u_i$ | Utility function of player $i$; $u_i : S \to \mathbb{R}$ |

**Table 2.5** Notation for a strategic form game

## 2.2 Dominant Strategy Equilibria

There are two notions of dominance that are aptly called strong dominance and weak dominance.

### *2.2.1 Strong Dominance*

**Definition 2.2 (Strongly Dominated Strategy).** Given a game $\Gamma = \langle N,(S_i),(u_i)\rangle$, a strategy $s_i \in S_i$ is said to be strongly dominated if there exists another strategy $s_i' \in S_i$ such that

$$u_i(s_i',s_{-i}) > u_i(s_i,s_{-i}) \ \ \forall s_{-i} \in S_{-i}.$$

We also say strategy $s_i'$ strongly dominates strategy $s_i$.

**Definition 2.3 (Strongly Dominant Strategy).** A strategy $s_i^* \in S_i$ is said to be a strongly dominant strategy for player $i$ if it strongly dominates every other strategy $s_i \in S_i$. That is, $\forall s_i \neq s_i^*$,

$$u_i(s_i^*,s_{-i}) > u_i(s_i,s_{-i}) \ \ \forall s_{-i} \in S_{-i}.$$

**Definition 2.4 (Strongly Dominant Strategy Equilibrium).** A profile of strategies $(s_1^*,s_2^*,\ldots,s_n^*)$ is called a strongly dominant strategy equilibrium of the game $\Gamma = \langle N,(S_i),(u_i)\rangle$ if $\forall i = 1,2,\ldots,n$, the strategy $s_i^*$ is a strongly dominant strategy for player $i$.

*Example 2.9 (Dominant Strategies in the Prisoner's Dilemma Problem).* Recall the prisoner's dilemma problem. Observe that the strategy NC is strongly dominated by C for player 1 since

$$u_1(C,NC) > u_1(NC,NC); \quad u_1(C,C) > u_1(NC,C).$$

Similarly, the strategy NC is strongly dominated by C for player 2 also, since

$$u_2(NC,C) > u_2(NC,NC); \quad u_2(C,C) > u_2(C,NC).$$

Thus $C$ is a strongly dominant strategy for both the players. Therefore $(C,C)$ is a strongly dominant strategy equilibrium for this game.

*Note 2.1.* If a player has a strongly dominant strategy then we should expect him to play it. On the other hand, if a player has a strongly dominated strategy, then we should expect him to not play it.

*Note 2.2.* A strongly dominant strategy equilibrium, if one exists, will be unique. The proof of this result is fairly straightforward.

### 2.2.2 Weak Dominance

**Definition 2.5 (Weakly Dominated Strategy).** Given a game $\Gamma = \langle N, (S_i), (u_i) \rangle$, a strategy $s_i \in S_i$ is said to be weakly dominated by a strategy $s_i' \in S_i$ for player $i$ if

$$u_i(s_i', s_{-i}) \geq u_i(s_i, s_{-i}) \ \ \forall s_{-i} \in S_{-i} \ \text{ and } \ u_i(s_i', s_{-i}) > u_i(s_i, s_{-i}) \ \text{ for some } \ s_{-i} \in S_{-i}.$$

The strategy $s_i'$ is said to weakly dominate strategy $s_i$.

**Definition 2.6 (Weakly Dominant Strategy).** A strategy $s_i^*$ is said to be a weakly dominant strategy for player $i$ if it weakly dominates every other strategy $s_i \in S_i$.

**Definition 2.7 (Weakly Dominant Strategy Equilibrium).** Given a game $\Gamma = \langle N, (S_i), (u_i) \rangle$, a strategy profile $(s_1^*, \ldots, s_n^*)$ is called a weakly dominant strategy equilibrium if for $i = 1, \ldots, n$, the strategy $s_i^*$ is a weakly dominant strategy for player $i$.

*Example 2.10 (A Modified Prisoner's Dilemma Problem).* Consider the payoff matrix of a slightly modified version of the prisoner's dilemma problem, shown in Table 2.6. Observe that the strategy $C$ is weakly dominant for player 1 since

| | 2 | |
|---|---|---|
| 1 | NC | C |
| NC | −2, −2 | −10, −2 |
| C | −2, −10 | −5, −5 |

**Table 2.6** Payoff matrix of a modified prisoner's dilemma problem

$$u_1(C,NC) = u_1(NC,NC); \quad u_1(C,C) > u_1(NC,C).$$

Also, $C$ is weakly dominant for player 2 since

$$u_2(NC,C) = u_2(NC,NC); \quad u_2(C,C) > u_2(C,NC).$$

Therefore the strategy profile $(C,C)$ is a weakly dominant strategy equilibrium.

*Note 2.3.* It is to be noted that there could exist multiple weakly dominant strategies for a player, and therefore there could exist multiple weakly dominant strategy equilibria in a strategic form game.

*Example 2.11 (Second Price Sealed Bid Auction with Complete Information).* Consider the second price sealed bid auction for selling a single indivisible item discussed in Example 2.5. Let $b_1, b_2, \ldots, b_n$ be the bids (strategies), and we shall denote a bid profile (strategy profile) by $b = (b_1, b_2, \ldots, b_n)$. Assume that $v_i, b_i \in (0,\infty)$ for $i = 1,2,\ldots,n$. Recall that the item is awarded to the bidder who has the lowest index among all the highest bidders. Recall the allocation function:

$$\begin{aligned} y_i(b_1,\ldots,b_n) &= 1 \text{ if } \quad b_i > b_j \text{ for } j = 1,2,\ldots,i-1 \text{ and} \\ &\qquad\qquad b_i \geq b_j \text{ for } j = i+1,\ldots,n \\ &= 0 \text{ else.} \end{aligned}$$

The payoff for each bidder is given by:

$$u_i(b_1,\ldots,b_n) = y_i(b_1,\ldots,b_n)(v_i - t_i(b_1,\ldots,b_n))$$

where $t_i(b_1,\ldots,b_n)$ is the amount paid by the winning bidder. Being second price auction, the winner pays only the next highest bid. We now show that the strategy profile $(b_1,\ldots,b_n) = (v_1,\ldots,v_n)$ is a weakly dominant strategy equilibrium for this game.

**Proof**: Consider bidder 1. His value is $v_1$ and bid is $b_1$. The other bidders have bids $b_2,\ldots,b_n$ and valuations $v_2,\ldots,v_n$. We consider the following cases.

**Case 1**: $v_1 \geq \max(b_2,\ldots,b_n)$. There are two sub-cases here: $b_1 \geq \max(b_2,\ldots,b_n)$ and $b_1 < \max(b_2,\ldots,b_n)$.
**Case 2**: $v_1 < \max(b_2,\ldots,b_n)$. There are two sub-cases here: $b_1 \geq \max(b_2,\ldots,b_n)$ and $b_1 < \max(b_2,\ldots,b_n)$.

We analyze these cases separately below.

**Case 1:** $v_1 \geq \max(b_2,\ldots,b_n)$.

We look at the following scenarios.

- Let $b_1 \geq \max(b_2,\ldots,b_n)$. This implies that bidder 1 is the winner, which implies that $u_1 = v_1 - \max(b_2,\ldots,b_n) \geq 0$.
- Let $b_1 < \max(b_2,\ldots,b_n)$. This means that bidder 1 is not the winner, which in turn means $u_1 = 0$.
- Let $b_1 = v_1$, then since $v_1 \geq \max(b_2,\ldots,b_n)$, we have $u_1 = v_1 - \max(b_2,\ldots,b_n)$.

Therefore, if $b_1 = v_1$, the utility $u_1$ is greater than or equal to the maximum utility obtainable. Thus, whatever the values of $b_2, \ldots, b_n$, it is a best response for player 1 to bid $v_1$. Thus $b_1 = v_1$ is a weakly dominant strategy for a bidder 1.

**Case 2:** $v_1 < \max(b_2, \ldots, b_n)$.

As before, we look at the following scenarios.

- Let $b_1 \geq \max(b_2, \ldots, b_n)$. This implies that bidder 1 is the winner and the payoff is given by:
$$u_1 = v_1 - \max(b_2, \ldots, b_n) \quad < 0.$$
- Let $b_1 < \max(b_2, \ldots, b_n)$. This means bidder 1 is not the winner. Therefore $u_1 = 0$.
- If $b_1 = v_1$, then bidder 1 is not the winner and therefore $u_1 = 0$.

From the above analysis, it is clear that $b_1 = v_1$ is a best response strategy for player 1 in Case 2 also. Combining our analysis of Case 1 and Case 2, we have that

$$u_1(v_1, b_2, \ldots, b_n) \geq u_1(\hat{b_1}, b_2, \ldots, b_n) \ \forall \hat{b_1} \in S_1 \quad \forall b_2 \in S_2, \ \ldots, \ b_n \in S_n$$

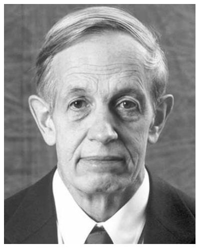
Also, we can show (and this is left as an exercise) that, for any $b_1' \neq v_1$, we can always find $b_2 \in S_2, b_3 \in S_3, \ldots, b_n \in S_n$, such that

$$u_1(v_1, b_2, \ldots b_n) > u_1(b_1', b_2, \ldots, b_n).$$

Thus $b_1 = v_1$ is a weakly dominant strategy for a bidder 1. Using almost similar arguments, we can show that $b_i = v_i$ is a weakly dominant strategy for bidder $i$ where $i = 2, 3, \ldots, n$. Therefore $(v_1, \ldots, v_n)$ is a weakly dominant strategy equilibrium.

## 2.3 Pure Strategy Nash Equilibrium

Dominant strategy equilibria (strongly dominant, weakly dominant), if they exist, are very desirable but rarely do they exist because the conditions to be satisfied are too demanding. A dominant strategy equilibrium requires that each player's choice be a best response against all possible choices of all the other players. If we only insist that each player's choice is a best response against the best response strategies of the other players, we get the notion of Nash equilibrium. This solution concept derives its name from John Nash, one of the most celebrated game theorists of our times. In this section, we introduce and discuss the notion of pure strategy Nash equilibrium. In the following section, we discuss the notion of mixed strategy Nash equilibrium.



**John F. Nash, Jr.** is described by many as one of the most original mathematicians of the 20th Century. He was born in 1928 in Bluefield, West Virginia, USA. He completed his BS and MS in the same year (1948) at Carnegie Mellon University, majoring in Mathematics. He became a student of Professor Albert Tucker at Princeton University and completed his Ph.D. in Mathematics in 1950. His doctoral thesis (which had exactly 28 pages) proposed the brilliant notion of Nash Equilibrium, which helped expand the scope of game theory beyond two player zero sum games. His main result in his doctoral work settled the question of existence of a mixed strategy equilibrium in finite strategic form games. During his doctoral study, Nash also wrote a remarkable paper on the two player bargaining problem. He showed using a highly imaginative axiomatic approach that there exists a unique solution to the two person bargaining problem.

He worked as a professor of Mathematics at MIT in the Department of Mathematics where he did path-breaking work on algebraic geometry. He is also known for the Nash embedding theorem, which proves that any abstract Riemannian manifold can be isometrically realized as a sub-manifold of the Euclidean space. He is also known for his fundamental contributions to nonlinear parabolic partial differential equations.

The life and achievements of John Nash are fascinatingly captured in his biography *A Beautiful Mind* authored by Sylvia Nasar. This was later made into a popular movie with the same title. Professor Nash is currently at the Princeton University.

In 1994, John Nash was awarded the Nobel Prize in Economic Sciences, jointly with Professor John C. Harsanyi, University of California, Berkeley, CA, USA, and Professor Dr. Reinhard Selten, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany, for their pioneering analysis of equilibria in the theory of non-cooperative games.

**Definition 2.8 (Pure Strategy Nash Equilibrium).** Given a strategic form game $\Gamma = \langle N, (S_i), (u_i) \rangle$, the strategy profile $s^* = (s_1^*, s_2^*, \ldots, s_n^*)$ is said to be a pure strategy Nash equilibrium of $\Gamma$ if,

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*), \quad \forall s_i \in S_i, \quad \forall i = 1, 2, \ldots, n.$$

That is, each player's Nash equilibrium strategy is a best response to the Nash equilibrium strategies of the other players.

**Definition 2.9 (Best Response Correspondence for Player $i$).** Given a game $\Gamma = \langle N, (S_i), (u_i) \rangle$, the best response correspondence for player $i$ is the mapping $B_i : S_{-i} \to 2^{S_i}$ defined by

$$B_i(s_{-i}) = \{s_i \in S_i : u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i}) \;\; \forall s_i' \in S_i\}.$$

That is, given a profile $s_{-i}$ of strategies of the other players, $B_i(s_{-i})$ gives the set of all best response strategies of player $i$.

**Definition 2.10 (An Alternative Definition of Nash Equilibrium).** Given a strategic form game $\Gamma = \langle N, (S_i), (u_i) \rangle$, the strategy profile $(s_1^*, \ldots, s_n^*)$ is a Nash equilibrium iff,

$$s_i^* \in B_i(s_{-i}^*), \quad \forall i = 1, \ldots, n.$$

*Note 2.4.* Given a strategic form game $\Gamma = \langle N, (S_i), (u_i) \rangle$, a strongly dominant strategy equilibrium or a weakly dominant strategy equilibrium $(s_1^*, \ldots s_n^*)$ is also a Nash equilibrium. This can be shown easily. The intuitive explanation for this is as follows. In a dominant strategy equilibrium, the equilibrium strategy of each player is a best response irrespective of the strategies of the rest of the players. In a pure strategy Nash equilibrium, the equilibrium strategy of each player is a best response against the Nash equilibrium strategies of the rest of the players. Thus, the Nash equilibrium is a much weaker version of a dominant strategy equilibrium. It is also fairly obvious to note that a Nash equilibrium need not be a dominant strategy equilibrium.

### 2.3.1 Pure Strategy Nash Equilibria: Examples

*Example 2.12 (Battle of Companies).* Consider Example 2.2 from the previous section. There are two Nash equilibria here, namely (A,A) and (B,B). The profile (A,A) is Nash equilibrium because

$$u_1(A,A) > u_1(B,A);\ \ u_2(A,A) > u_2(A,B).$$

The profile $(B,B)$ is a Nash equilibrium because

$$u_1(B,B) > u_1(A,B);\ \ u_2(B,B) > u_2(B,A).$$

The best response sets are given by:

$$B_1(A) = \{A\};\ \ B_1(B) = \{B\};\ \ B_2(A) = \{A\};\ \ B_2(B) = \{B\}.$$

Since $A \in B_1(A)$ and $A \in B_2(A)$, $(A,A)$ is a Nash equilibrium. Similarly since $B \in B_1(B)$ and $B \in B_2(B)$, $(B,B)$ is a Nash equilibrium.

*Example 2.13 (Prisoner's Dilemma).* Consider the prisoner's dilemma problem introduced in Example 2.4. Note that $(C,C)$ is the unique pure strategy Nash equilibrium here. To see why, we look at the best response sets:

$$B_1(C) = \{C\};\ \ B_1(NC) = \{C\};\ \ B_2(C) = \{C\};\ \ B_2(NC) = \{C\}.$$

Since $s_i^* \in B_1(s_2^*)$ and $s_2^* \in B_2(s_1^*)$ for a Nash equilibrium, the only possible pure strategy Nash equilibrium here is $(C,C)$. In fact as already seen, this is a strongly dominant strategy equilibrium. Note that the strategy profile $(NC,NC)$ is not a Nash equilibrium, though, it is jointly the most desirable outcome for the two prisoners. Quite often, Nash equilibrium profiles are not necessarily the the best outcomes.

*Example 2.14 (Bandwidth Sharing Game).* Recall the bandwidth sharing game discussed in Example 2.6. We compute a Nash equilibrium for this game in the following way. Let $x_i$ be the amount of flow that player $i$ ($i = 1,2,\dots,n$) wishes to transmit on the channel and assume that

$$\sum_{i \in N} x_i < 1.$$

Consider player $i$ and define:

$$t = \sum_{j \neq i} x_j.$$

The payoff for the player $i$ is equal to

$$x_i(1 - t - x_i).$$

In order to maximize the above payoff, we have to choose

$$\begin{aligned} x_i^* &= \arg \max_{x_i \in [0,1]} x_i(1 - t - x_i) \\ &= \frac{1 - t}{2} \\ &= \frac{1 - \sum_{j \neq i} x_j^*}{2}. \end{aligned}$$

If this has to be satisfied for all $i \in N$, then we end up with $n$ simultaneous equations

$$x_i^* = \frac{1 - \sum_{j \neq i} x_j^*}{2} \quad i = 1, 2, \ldots, n.$$

A Nash equilibrium of this game is any solution to the above $n$ simultaneous equations. It can be shown that the above set of simultaneous equations has the unique solution:

$$x_i^* = \frac{1}{1 + n} \quad i = 1, 2, \ldots, n.$$

The profile $(x_1^*, \ldots, x_n^*)$ is thus a Nash equilibrium. The payoff for player $i$ in the above Nash equilibrium

$$= \left(\frac{1}{n+1}\right)\left(\frac{1}{n+1}\right).$$

Therefore the total payoff to all players combined

$$= \frac{n}{(n+1)^2}.$$

As shown below, the above is not a very happy situation. Consider the following non-equilibrium profile

$$\left(\frac{1}{2n}, \frac{1}{2n}, \ldots, \frac{1}{2n}\right).$$

This profile gives each player a payoff

$$\begin{aligned} &= \frac{1}{2n}\left(1 - \frac{n}{2n}\right) \\ &= \frac{1}{4n}. \end{aligned}$$

Therefore the total payoff to all the players

$$= \frac{1}{4} > \frac{n}{(n+1)^2}.$$

Thus a non-equilibrium payoff $\left(\frac{1}{2n}, \frac{1}{2n}, \ldots, \frac{1}{2n}\right)$ provides more payoff than a Nash equilibrium payoff. This is referred to as a *tragedy of the commons*. In general, like in the prisoner's dilemma problem, the equilibrium payoffs may not be the best possible outcome for the players individually and also collectively. This lack of Pareto optimality is a property that Nash equilibrium payoffs often suffer from.

*Example 2.15 (Duopoly Pricing Game).* Recall the pricing game discussed in Example 2.7. There are two companies 1 and 2 that wish to maximize their profits by choosing their prices $p_1$ and $p_2$. The utilities of the two companies are:

$$u_1(p_1, p_2) = (p_1 - c)\, x_1(p_1, p_2)$$
$$u_2(p_1, p_2) = (p_2 - c)\, x_2(p_1, p_2).$$

Note that $u_1(c,c) = 0$ and $u_2(c,c) = 0$. Also, it can be easily noted that

$$u_1(c,c) \geq u_1(p_1, c) \;\; \forall p_1 \in S_1$$

$$u_2(c,c) \geq u_2(c, p_2) \;\; \forall p_2 \in S_2.$$

Therefore the strategy profile $(c,c)$ is a pure strategy Nash equilibrium. The implication of this result is that in the equilibrium, the companies set their prices equal to the marginal cost. The intuition behind this result is to imagine what would happen if both the companies set equal prices above marginal cost. Then the two companies would get half the market at a higher than marginal cost price. However, by lowering prices just slightly, a firm could gain the whole market, so both firms are tempted to lower prices as much as they can. It would be irrational to price below marginal cost, because the firm would make a loss. Therefore, both firms will lower prices until they reach the marginal cost limit.

*Example 2.16 (Game without a Pure Strategy Nash Equilibrium).* Recall the matching companies game (Example 2.1) and the payoff matrix for this game:

| | 2 | |
|---|---|---|
| 1 | A | B |
| A | +1, −1 | −1, +1 |
| B | −1, +1 | +1, −1 |

It is easy to see that this game does not have a pure strategy Nash equilibrium. This example shows that there is no guarantee that a pure strategy Nash equilibrium will exist. Later on in this section, we will state sufficient conditions under which a given strategic form game is guaranteed to have a pure strategy Nash equilibrium. In the next section, we will show that this game has a mixed strategy Nash equilibrium. We will now study another example that does not have a pure strategy Nash equilibrium.

*Example 2.17 (Procurement Exchange Game).* Recall Example 2.8. Let us investigate if this game has a pure strategy Nash equilibrium. First, we explore whether the strategy profile $(1, s_2)$ is a Nash equilibrium for any $s_2 \in [0,1]$. Note that

$$\begin{aligned} u_1(1,s_2) &= 2 \text{ if } s_2 = 1 \\ &= 1 \text{ if } s_2 < 1 \\ u_2(1,s_2) &= 1 \text{ if } s_2 = 1 \\ &= 2s_2 \text{ if } s_2 < 1. \end{aligned}$$

It is easy to observe that $u_2(1,s_2)$ has a value $2s_2$ for $0 \leq s_2 < 1$. Therefore $u_2(1,s_2)$ increases when $s_2$ increases from 0, until $s_2$ reaches 1 when it suddenly drops to 1. Thus it is clear that a profile of the form $(1,s_2)$ cannot be a Nash equilibrium for any $s_2 \in [0,1]$. Similarly, no profile of the form $(s_1,1)$ can be a Nash equilibrium for any $s_1 \in [0,1]$.

Let us now explore if there exists any Nash equilibrium $(s_1^*, s_2^*)$, with $s_1^*, s_2^* \in [0,1)$. There are two cases here.

- **Case 1**: If $s_1^* \leq \frac{1}{2}$, then the best response for player 2 would be to bid $s_2 = 1$ since that would fetch him the maximum payoff. However bidding $s_2 = 1$ is not an option here since the range of values for $s_2$ is $[0,1)$.
- **Case 2**: If $s_1^* > \frac{1}{2}$, there are two cases: (1) $s_1^* \leq s_2^*$ (2) $s_1^* > s_2^*$. Suppose $s_1^* \leq s_2^*$. Then

$$\begin{aligned} u_1(s_1^*, s_2^*) &= 2s_1^* \\ u_2(s_1^*, s_2^*) &= s_2^*. \end{aligned}$$

Choose $s_2$ such that $\frac{1}{2} < s_2 < s_1^*$. Then

$$\begin{aligned} u_2(s_1^*, s_2) &= 2s_2 \\ &> s_2^* \text{ since } 2s_2 > 1 \text{ and } s_2^* < 1 \\ &= u_2(s_1^*, s_2^*). \end{aligned}$$

Thus we are able to improve upon $(s_1^*, s_2^*)$ and hence $(s_1^*, s_2^*)$ is not a Nash equilibrium.

Now, suppose, $s_1^* > s_2^*$. Then

$$\begin{aligned} u_1(s_1^*, s_2^*) &= s_1^* \\ u_2(s_1^*, s_2^*) &= 2s_2^*. \end{aligned}$$

Now let us choose $s_1$ such that $1 > s_1 > s_1^*$. Then

$$u_1(s_1, s_2^*) = s_1 > s_1^* = u_1(s_1^*, s_2^*).$$

Thus we can always improve upon $(s_1^*, s_2^*)$. Therefore this game does not have a pure strategy Nash equilibrium. We wish to remark here that this game does not even have a mixed strategy Nash equilibrium.

### *2.3.2 Interpretations of Nash Equilibrium*

Nash equilibrium is one of the most extensively discussed and debated topics in game theory. Many interpretations have been provided. Note that a Nash equilibrium is a profile of strategies, one for each of the $n$ players, that has the property that each player's choice is his best response given that the rest of the players play their Nash equilibrium strategies. By deviating from a Nash equilibrium strategy, a player will not be better off given that the other players stick to their Nash equilibrium strategies. The following discussion provides several interpretations put forward by game theorists.

A popular interpretation views a Nash equilibrium as a *prescription*. An adviser or a consultant to the $n$ players would essentially prescribe a Nash equilibrium strategy profile to the players. If the adviser recommends strategies that do not constitute a Nash equilibrium, then some players would find that it would be better for them to do differently than advised. If the adviser prescribes strategies that do constitute a Nash equilibrium, then the players are not unhappy because playing the equilibrium strategy is best under the assumption that the other players will play their equilibrium strategies.

Thus a logical, rational, adviser would recommend a Nash equilibrium profile to the players. The immediate caution, however, is that Nash equilibrium is an insurance against only unilateral deviations (that is, only one player at a time deviating from the equilibrium strategy). Two or more players deviating might result in players improving their payoffs compared to their equilibrium payoffs. For example, in the prisoner's dilemma problem, $(C,C)$ is a Nash equilibrium. If both the players decide to deviate, then the resulting profile is $(NC,NC)$, which is better for both the players. Note that $(NC,NC)$ is not a Nash equilibrium.

Another popular interpretation of Nash equilibrium is that of *prediction*. If the players are rational and intelligent, then a Nash equilibrium is a good prediction for the game. For example, a systematic elimination of strongly dominated strategies will lead to a reduced form that will include a Nash equilibrium. Often, iterated elimination of strongly dominated strategies leads to a unique prediction which would be invariably a Nash equilibrium.

An appealing interpretation of Nash equilibrium is that of *self-enforcing agreement*. A Nash equilibrium can be viewed as an implicit or explicit agreement between the players. Once this agreement is reached, it does not need any external means of enforcement because it is in the self-interest of each player to follow this agreement if the others do. In a noncooperative game, agreements cannot be enforced, hence, Nash equilibrium agreements are the only ones sustainable.

A natural, easily understood interpretation for Nash equilibrium has to do with *Evolution and Steady-State*. A Nash equilibrium is a potential stable point of a dynamic adjustment process in which players adjust their behavior to that of other players in the game, constantly searching for strategy choices that will give them the best results. This interpretation has been used to explain biological evolution. In this interpretation, Nash equilibrium is the outcome that results over time when a game is played repeatedly. A Nash equilibrium is like a stable social convention that people are happy to maintain forever.

Common knowledge was usually a standard assumption in determining conditions leading to a Nash equilibrium. More recently, it has been shown that the common knowledge assumption may not be required; instead, mutual knowledge is adequate. Suppose that each player is rational, knows his own payoff function, and knows the strategy choices of the others; then the strategy choices of the players will constitute a Nash equilibrium.

### 2.3.3 Existence of a Pure Strategy Nash Equilibrium

Consider a strategic form game $\Gamma = \langle N, (S_i), (u_i) \rangle$. We have seen examples (1) where only a unique pure strategy Nash equilibrium exists (2) where multiple pure strategy Nash equilibria exist, and (3) where a pure strategy Nash equilibrium does not exist. We now present a proposition that provides sufficient conditions under which a strategic form game is guaranteed to have at least one pure strategy Nash equilibrium. We do not delve into the technical aspects of this proposition; we refer the reader to consult [6] for more details.

**Proposition 2.1.** *Given a strategic form game $\Gamma = \langle N, (S_i), (u_i) \rangle$, a pure strategy Nash equilibrium exists if for all $i \in N$,*

- *$S_i$ is a non-empty, convex, compact subset of $\mathbb{R}^n$*
- *$u_i(s_1, \ldots, s_n)$ is continuous in $(s_1, \ldots, s_n)$ and quasi-concave in $s_i$*

In the next section where we consider mixed strategy Nash equilibria, we will state the celebrated Nash's Theorem, which guarantees the existence of a mixed strategy Nash equilibrium for a finite strategic form game.

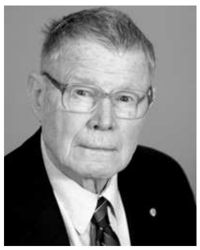
### 2.3.4 Existence of Multiple Nash Equilibria

We have seen a few examples of strategic form games where multiple Nash equilibria exist. If a game has multiple Nash equilibria, then a fundamental question to ask is which of these would be implemented by the players? This question has been addressed by numerous game theorists, in particular, Thomas Schelling, who proposed the *focal point effect*. According to Schelling, anything that tends to focus the player's attention on one equilibrium may make them all expect it and hence fulfill it, like a self-fulfilling prophecy. Such a Nash equilibrium, which has some property that distinguishes it from all other equilibria is called a *focal equilibrium* or a *Schelling Point*.

As an example, consider the battle of companies game that we discussed in Example 2.2. Recall the payoff matrix of this game:

| | 2 | |
|---|---|---|
| 1 | A | B |
| A | 2,1 | 0,0 |
| B | 0,0 | 1,2 |

Here $(A,A)$ and $(B,B)$ are both Nash equilibria. If there is a special interest (or hype) created about product A, then $(A,A)$ may become the focal equilibrium. On the other hand, if there is a marketing blitz on product B, then $(B,B)$ may become the focal equilibrium.



**Thomas Schelling** received, jointly with Robert Aumann, the 2005 Nobel Prize in Economic Sciences for contributions towards a clear understanding of conflict and cooperation through game theory analysis. Schelling's stellar contributions are best captured by several books that he has authored. The book *The Strategy of Conflict* that he wrote in 1960 is a classic work that has pioneered the study of bargaining and strategic behavior. It has been voted as one of the 100 most influential books since 1945. The notion of *focal point*, which is now called the *Schelling point* is introduced in this work to explain strategic behavior in the presence of multiple equilibria. Another book entitled *Arms and Influence* is also a popularly cited work. A highlight of Schelling's work has been to use simple game theoretic models in an imaginative way to obtain deep insights into global problems such as the cold war, nuclear arms race, war and peace, etc.

Schelling was born on April 14, 1921. He received his Doctorate in Economics from Harvard University in 1951. During 1950-53, Schelling was in the team of foreign policy advisers to the US President, and ever since, he has held many policy making positions in public service. He has played an influential role in the global warming debate also. He was at Harvard University from 1958 to 1990. Since 1990, he has been a Distinguished University Professor at the University of Maryland, in the Department of Economics and the School of Public policy.

## 2.4 Mixed Strategy Nash Equilibrium

### 2.4.1 Randomized Strategies or Mixed Strategies

Consider a strategic form game: $\Gamma = \langle N, (S_i), (u_i) \rangle$. The elements of $S_i$ are called the *pure strategies* of player $i$ ($i = 1, \ldots, n$). If player $i$ randomly chooses one element of the set $S_i$, we have a mixed strategy or a randomized strategy. In the discussion that follows, we assume that $S_i$ is a finite for each $i = 1, 2, \ldots, n$.

**Definition 2.11 (Mixed Strategies).** Given a player $i$ with $S_i$ as the set of pure strategies, a mixed strategy $\sigma_i$ for player $i$ is a probability distribution over $S_i$. That is, $\sigma_i : S_i \to [0,1]$ assigns to each pure strategy $s_i \in S_i$, a probability $\sigma_i(s_i)$ such that

$$\sum_{s_i \in S_i} \sigma_i(s_i) = 1.$$

A pure strategy of a player, say $s_i \in S_i$, can be considered as a mixed strategy that assigns probability 1 to $s_i$ and probability 0 to all other strategies of player $i$. Such a mixed strategy is said to be a *degenerate mixed strategy* and is denoted by $e(s_i)$ or simply by $s_i$.

If $S_i = \{s_{i1}, s_{i2}, \ldots, s_{im}\}$, then clearly, the set of all mixed strategies of player $i$ is the set of all probability distributions on the set $S_i$. In other words, it is the simplex:

$$\Delta(S_i) = \left\{ (\sigma_{i1}, \ldots, \sigma_{im}) \in \mathbb{R}^m : \sigma_{ij} \geq 0 \text{ for } j = 1, \ldots, m \text{ and } \sum_{j=1}^{m} \sigma_{ij} = 1 \right\}.$$

The above simplex is called the *mixed extension* of $S_i$. Using the mixed extensions of strategy sets, we would like to define a mixed extension of the pure strategy game $\Gamma = \langle N, (S_i), (u_i) \rangle$. Let us denote the mixed extension of $\Gamma$ by

$$\Gamma_{ME} = \langle N, (\Delta(S_i)), (U_i) \rangle.$$

Note that, for $i = 1, 2, \ldots, n$,

$$U_i : \times_{i \in N} \Delta(S_i) \to \mathbb{R}.$$

Given $\sigma_i \in \Delta(S_i)$ for $i = 1, \ldots, n$, we compute $U_i(\sigma_1, \ldots, \sigma_n)$ as follows. First, we make the standard assumption that the randomizations of individual players are mutually independent. This implies that given a profile $(\sigma_1, \ldots, \sigma_n)$, the random variables $\sigma_1, \ldots, \sigma_n$ are mutually independent. Therefore the probability of a pure strategy profile $(s_1, \ldots, s_n)$ is given by

$$\sigma(s_1, \ldots, s_n) = \prod_{i \in N} \sigma_i(s_i).$$

The payoff functions $U_i$ are defined in a natural way as

$$U_i(\sigma_1, \ldots, \sigma_n) = \sum_{(s_1, \ldots, s_n) \in S} \sigma(s_1, \ldots, s_n)\, u_i(s_1, \ldots, s_n).$$

In the sequel, when there is no confusion, we will write $u_i$ instead of $U_i$. For example, instead of writing $U_i(\sigma_1, \ldots, \sigma_n)$, we will simply write $u_i(\sigma_1, \ldots, \sigma_n)$.

*Example 2.18 (Mixed Extension of the Battle of Companies Game).* Recall the game discussed in Example 2.2, having the following payoff matrix:

| | 2 | |
|---|---|---|
| 1 | A | B |
| A | 2,1 | 0,0 |
| B | 0,0 | 1,2 |

Suppose $(\sigma_1, \sigma_2)$ is a mixed strategy profile. This means that $\sigma_1$ is a probability distribution on $S_1 = \{A, B\}$, and $\sigma_2$ is a probability distribution on $S_2 = \{A, B\}$. Let us represent

$$\sigma_1 = (\sigma_1(A), \sigma_1(B))$$
$$\sigma_2 = (\sigma_2(A), \sigma_2(B)).$$

We have

$$S = S_1 \times S_2 = \{(A,A),(A,B),(B,A),(B,B)\}.$$

We will now compute the payoff functions $u_1$ and $u_2$. Note that, for $i = 1,2$,

$$u_i(\sigma_1,\sigma_2) = \sum_{(s_1,s_2)\in S} \sigma(s_1,s_2)\, u_i(s_1,s_2).$$

The function $u_1$ can be computed as

$$\begin{aligned}
u_1(\sigma_1,\sigma_2) &= \sigma_1(A)\sigma_2(A)u_1(A,A)\\
&\quad +\sigma_1(A)\sigma_2(B)u_1(A,B)\\
&\quad +\sigma_1(B)\sigma_2(A)u_1(B,A)\\
&\quad +\sigma_1(B)\sigma_2(B)u_1(B,B)\\
&= 2\sigma_1(A)\sigma_2(A)+\sigma_1(B)\sigma_2(B)\\
&= 2\sigma_1(A)\sigma_2(A)+(1-\sigma_1(A))(1-\sigma_2(A))\\
&= 1+3\sigma_1(A)\sigma_2(A)-\sigma_1(A)-\sigma_2(A).
\end{aligned}$$

Similarly, we can show that

$$u_2(\sigma_1,\sigma_2) = 2+3\sigma_1(A)\sigma_2(A)-2\sigma_1(A)-2\sigma_2(A).$$

Suppose $\sigma_1 = \left(\frac{2}{3},\frac{1}{3}\right)$ and $\sigma_2 = \left(\frac{1}{3},\frac{2}{3}\right)$. Then it is easy to see that

$$u_1(\sigma_1,\sigma_2) = \frac{2}{3};\quad u_2(\sigma_1,\sigma_2) = \frac{2}{3}.$$

### *2.4.2 Mixed Strategy Nash Equilibrium*

We now define the notion of a mixed strategy Nash equilibrium.

**Definition 2.12 (Mixed Strategy Nash Equilibrium).** A mixed strategy profile $(\sigma_1^*,\ldots,\sigma_n^*)$ is called a Nash equilibrium if $\forall i \in N$,

$$u_i(\sigma_i^*,\sigma_{-i}^*) \geq u_i(\sigma_i,\sigma_{-i}^*) \quad \forall \sigma_i \in \Delta(S_i).$$

Define the best response functions $B_i(.)$ as follows.

$$B_i(\sigma_{-i}) = \{\sigma_i \in \Delta(S_i) : u_i(\sigma_i,\sigma_{-i}) \geq u_i(\sigma_i',\sigma_{-i})\ \ \forall \sigma_i' \in \Delta(S_i)\}.$$

Then, clearly, a mixed strategy profile $(\sigma_1^*,\ldots,\sigma_n^*)$ is a Nash equilibrium iff

$$\sigma_i^* \in B_i(\sigma_{-i}^*) \quad \forall i = 1, 2, \ldots, n.$$

*Example 2.19 (Mixed Strategy Nash Equilibria for the Battle of Companies Game).* Suppose $(\sigma_1, \sigma_2)$ is a mixed strategy profile. We have already shown that

$$u_1(\sigma_1, \sigma_2) = 1 + 3\sigma_1(A)\sigma_2(A) - \sigma_1(A) - \sigma_2(A)$$
$$u_2(\sigma_1, \sigma_2) = 2 + 3\sigma_1(A)\sigma_2(A) - 2\sigma_1(A) - 2\sigma_2(A).$$

Let $(\sigma_1^*, \sigma_2^*)$ be a mixed strategy equilibrium. Then

$$u_1(\sigma_1^*, \sigma_2^*) \geq u_1(\sigma_1, \sigma_2^*) \quad \forall \sigma_1 \in \Delta(S_1)$$

$$u_2(\sigma_1^*, \sigma_2^*) \geq u_2(\sigma_1^*, \sigma_2) \quad \forall \sigma_2 \in \Delta(S_2).$$

The above two equations are equivalent to:

$$3\sigma_1^*(A)\sigma_2^*(A) - \sigma_1^*(A) \geq 3\sigma_1(A)\sigma_2^*(A) - \sigma_1(A)$$

$$3\sigma_1^*(A)\sigma_2^*(A) - 2\sigma_2^*(A) \geq 3\sigma_1^*(A)\sigma_2(A) - 2\sigma_2(A).$$

The last two equations are equivalent to:

$$\sigma_1^*(A)\{3\sigma_2^*(A) - 1\} \geq \sigma_1(A)\{3\sigma_2^*(A) - 1\} \quad \forall \sigma_1 \in \Delta(S_1) \tag{2.1}$$

$$\sigma_2^*(A)\{3\sigma_1^*(A) - 2\} \geq \sigma_2(A)\{3\sigma_1^*(A) - 2\} \quad \forall \sigma_2 \in \Delta(S_2). \tag{2.2}$$

There are three possible cases.

- Case 1: $3\sigma_2^*(A) > 1$. This leads to the pure strategy Nash equilibrium $(A, A)$.
- Case 2: $3\sigma_2^*(A) < 1$. This leads to the pure strategy Nash equilibrium $(B, B)$.
- Case 3: $3\sigma_2^*(A) = 1$. This leads to the mixed strategy profile:

$$\sigma_1^*(A) = \frac{2}{3};\ \ \sigma_1^*(B) = \frac{1}{3};\ \ \sigma_2^*(A) = \frac{1}{3};\ \ \sigma_2^*(B) = \frac{2}{3}.$$

  This is a candidate mixed strategy Nash equilibrium. We will later on show that this is indeed a mixed strategy Nash equilibrium using a necessary and sufficient condition for a mixed strategy profile to be a Nash equilibrium, which we present next.

**Definition 2.13 (Support of a Mixed Strategy).** Let $\sigma_i$ be any mixed strategy of a player $i$. The support of $\sigma_i$, denoted by $\delta(\sigma_i)$, is the set of all pure strategies which have non-zero probabilities under $\sigma_i$. That is,

$$\delta(\sigma_i) = \{s_i \in S_i : \sigma_i(s_i) > 0\}.$$

**Theorem 2.1 (A Necessary and Sufficient Condition for Nash Equilibrium).** *The mixed strategy profile $(\sigma_1^*, \ldots, \sigma_n^*)$ is a mixed strategy Nash equilibrium iff for every player $i \in N$, we have*

1. *$u_i(s_i, \sigma_{-i}^*)$ is the same $\forall s_i \in \delta(\sigma_i^*)$.*

*2.* $u_i(s_i, \sigma^*_{-i}) \geq u_i(s'_i, \sigma^*_{-i}) \;\; \forall s_i \in \delta(\sigma^*_i) \;\; \forall s'_i \notin \delta(\sigma^*_i)$

*(that is, the payoff of the player i for each pure strategy having non-zero probability is the same and is greater than or equal to the payoff for each pure strategy having zero probability).*

This theorem has much significance for computing Nash equilibria. Any standard textbook may be looked up (for example [1]) for a proof of this theorem. The theorem has the following implications.

1. In a mixed strategy Nash equilibrium, each player gets the same payoff (as in Nash equilibrium) by playing *any pure strategy* having positive probability in his Nash equilibrium strategy.
2. The above implies that the player can be indifferent about which of the pure strategies (with positive probability) he/she will play.
3. To verify whether or not a mixed strategy profile is a Nash equilibrium, it is enough to consider the effects of only pure strategy deviations.

*Example 2.20 (Mixed Strategy Nash Equilibrium for the Battle of Companies Game).* We now show, using the above theorem, that the strategy profile

$$\sigma^*_1(A) = \frac{2}{3};\;\; \sigma^*_1(B) = \frac{1}{3};\;\; \sigma^*_2(A) = \frac{1}{3};\;\; \sigma^*_2(B) = \frac{2}{3}$$

is a mixed strategy Nash equilibrium. Note that $\delta(\sigma^*_1) = \{A, B\}$ and $\delta(\sigma^*_2) = \{A, B\}$. Note immediately that condition (2) of the above theorem is trivially satisfied for both the players. We will now investigate condition (1). Note that

$$u_1(A, \sigma^*_2) = u_1(B, \sigma^*_2) = 2 \times \frac{1}{3} + 0 \times \frac{2}{3} = \frac{2}{3}$$

$$u_2(\sigma^*_1, A) = u_2(\sigma^*_1, B) = 1 \times \frac{2}{3} + 0 \times \frac{1}{3} == \frac{2}{3}.$$

Therefore, condition (2) is also satisfied for both the players. Thus the above profile is a mixed strategy Nash equilibrium.

*Example 2.21 (Mixed Strategy Nash Equilibrium for Matching Companies).* For this game (Example 2.1), we have seen that there does not exist a pure strategy Nash equilibrium. The mixed strategy profile $(\sigma^*_1, \sigma^*_2)$ defined by

$$\sigma^*_1(A) = \frac{1}{2};\;\; \sigma^*_1(B) = \frac{1}{2};\;\; \sigma^*_2(A) = \frac{1}{2};\;\; \sigma^*_2(B) = \frac{1}{2}$$

can be easily shown to be a mixed strategy Nash equilibrium.

### 2.4.3 Existence of Mixed Strategy Nash Equilibrium

This is an important question that confronted game theorists for a long time. The first important result on this question was resolved by Von Neumann in 1928, who showed that a two person zero sum game is guaranteed to have a randomized saddle point (which is the same as a mixed strategy Nash equilibrium for two person zerosum games). In 1950, Nash introduced his notion of equilibrium for multiperson games and established a significant result that is stated in the following proposition.

**Theorem 2.2 (Nash's Theorem).** *Let $\Gamma = \langle N, (S_i), (u_i) \rangle$ be a finite strategic form game (that is, $N$ is finite and $S_i$ is finite for each $i \in N$). Then $\Gamma$ has at least one mixed strategy Nash equilibrium.*

The proof of Nash's theorem is based on Kakutani's fixed point theorem. The books [6, 1] may be consulted for a proof of this theorem. Nash's theorem, however, does not guarantee the existence of a Nash equilibrium for an infinite game. For example, the pricing game in a procurement exchange (Example 2.8), which is an infinite game, does not have a mixed strategy Nash equilibrium.

### 2.4.4 Computation of Nash Equilibria

The problem of computing Nash equilibria is one of the fundamental computational problems in game theory. In fact, this problem has been listed as one of the important current challenges in the area of algorithms and complexity. The problem has generated intense interest since the 1950s. A famous algorithm for computing a Nash equilibrium in bimatrix games (two player non-zero sum games) is due to Lemke and Howson (1964). Another famous algorithm for bimatrix games is due to Mangasarian (1964). Rosenmuller (1971) extended the Lemke-Howson algorithm to $n$-person finite games. Wilson (1971) and Scarf (1973) developed efficient algorithms for $n$-person games. McKelvey and McLennan [7] have provided an excellent survey of various algorithms for computing Nash equilibria.

In the case of two person zero-sum games, computing a mixed strategy Nash equilibrium can be accomplished by solving a linear program. This is a consequence of the famous minimax theorem and the linear programming duality based formulation provided by Von Neumann. The running time, therefore, is worst case polynomial time for two person zerosum games.

In the case of two person non-zero sum games, the complexity is unknown in the general case. There are no known complexity results even for symmetric two player games or even for pure strategy Bayesian Nash equilibria. According to Papadimitriou [8], proof of NP-completeness seems impossible. Papadimitriou has defined a complexity class called TFNP and shows that the following problems belong to TFNP: (a) computing NE, (b) factoring, (c) fixed point problems, (d) local optimization. Computing equilibria with certain special properties (such as maximal payoff, maximal support) have been shown to be NP-hard.

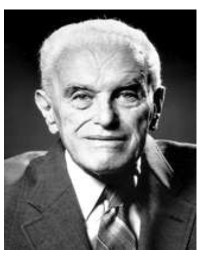
Complexity results for *n*-person games are mostly open and this is currently an active area of research. The articles by Papadimitriou [8] and Stengel [9] present a survey of the current art in this area.

## 2.5 Bayesian Games

We have so far studied strategic form games with complete information. We will now study games with incomplete information, which are crucial to the theory of mechanism design. A game with *incomplete information* is one in which, at the first point in time when the players can begin to plan their moves in the game, some players have *private information* about the game that other players do not know. In contrast, in *complete information* games, there is no such private information, and all information is publicly known to everybody. Clearly, incomplete information games are more realistic, more practical.

The initial private information that a player has, just before making a move in the game, is called the *type* of the player. For example, in an auction involving a single indivisible item, each player would have a valuation for the item, and typically the player himself would know this valuation deterministically while the other players may only have a guess about how much this player values the item.

John Harsanyi (Joint Nobel Prize winner in Economic Sciences in 1994 with John Nash and Reinhard Selten) proposed in 1968, *Bayesian form* games to represent games with incomplete information.



In 1994, **John Charles Harsanyi** was awarded the Nobel Prize in Economic Sciences, jointly with Professor John Nash and Professor Reinhard Selten, for their pioneering analysis of equilibria in the theory of non-cooperative games. Harsanyi is best known for his work on games with incomplete information and in particular Bayesian games, which he published as a series of three celebrated papers titled *Games with incomplete information played by Bayesian players* in the Management Science journal in 1967 and 1968. His work on analysis of Bayesian games is of foundational value to mechanism design since mechanisms crucially use the framework of games with incomplete information. Harsanyi is also acclaimed for his intriguing work on *utilitarian ethics*, where he applied game theory and economic reasoning in political and moral philosophy. Harsanyi's collaboration with Reinhard Selten on the topic of equilibrium analysis resulted in a celebrated book entitled *A General Theory of Equilibrium Selection in Games* (MIT Press, 1988).

John Harsanyi was born in Budapest, Hungary, on May 29, 1920. He got two doctoral degrees — the first one in philosophy from the University of Budapest in 1947 and the second one in economics from Stanford University in 1959. His adviser at Stanford University was Professor Kenneth Arrow, who got the Economics Nobel Prize in 1972. Harsanyi worked at the University of California, Berkeley, from 1964 to 1990 when he retired. He died on August 9, 2000, in Berkeley, California.

**Reinhard Selten** was jointly awarded the Nobel Prize in Economic Sciences in 1994, with Professor John Nash and Professor John Harsanyi. Harsanyi's analysis of Bayesian games was helped by the suggestions of Selten, and in fact Harsanyi refers to the type agent representation of Bayesian games as the *Selten game*. Selten is best known for his fundamental work on extensive form games and their transformation to strategic form through a representation called the agent normal form. Selten is also widely known for his deep work on bounded rationality. Furthermore, he is regarded as a pioneer of experimental economics. Harsanyi and Selten, in their remarkable book *A General Theory of Equilibrium Selection in Games* (MIT Press, 1988), develop a general framework to identify a unique equilibrium as the solution of a given finite strategic form game. Their solution can be thought of as a limit of an evolutionary process.

Selten was born in Breslau (currently in Poland but formerly in Germany) on October 5, 1930. He earned a doctorate in Mathematics from Frankfurt University, working with Professor Ewald Burger and Wolfgang Franz. He is currently a Professor Emeritus at the University of Bonn, Germany.

**Definition 2.14 (Bayesian Game).** A Bayesian game $\Gamma$ is defined as a tuple

$$\Gamma = \langle N, (\Theta_i), (S_i), (p_i), (u_i) \rangle$$

where

- $N = \{1, 2, \ldots, n\}$ is a set of players.
- $\Theta_i$ is the set of types of player $i$ where $i = 1, 2, \ldots, n$.
- $S_i$ is the set of actions or pure strategies of player $i$ where $i = 1, 2, \ldots, n$.
- The probability function $p_i$ is a function from $\Theta_i$ into $\Delta(\Theta_{-i})$, the set of probability distributions over $\Theta_{-i}$. That is, for any possible type $\theta_i \in \Theta_i$, $p_i$ specifies a probability distribution $p_i(.|\theta_i)$ over the set $\Theta_{-i}$ representing what player $i$ would believe about the types of the other players if his own type were $\theta_i$;
- The payoff function $u_i : \Theta \times S \rightarrow \mathbb{R}$ is such that, for any profile of actions and any profile of types $(\theta, s) \in \Theta \times S$, $u_i(\theta, s)$ specifies the payoff that player $i$ would get, in some *von Neumann – Morgenstern utility scale*, if the players' actual types were all as in $\theta$ and the players all chose their actions according to $s$.

The notation for Bayesian games is described in Table 2.7.

*Note 2.5.* When we study a Bayesian game, we assume that

1. Each player $i$ knows the entire structure of the game as defined above.
2. Each player knows his own type $\theta_i \in \Theta_i$.
3. The above facts are common knowledge among all the players in $N$.
4. The exact type of a player is not known deterministically to the other players who however have a probabilistic guess of what this type is. The belief functions $p_i$ describe these conditional probabilities. Note that the belief functions $p_i$ are also common knowledge among the players.

*Note 2.6.* The phrases *actions* and *strategies* are used differently in the Bayesian game context. A strategy for a player $i$ in Bayesian games is defined as a mapping

| | |
|---|---|
| $N$ | A set of players, $\{1, 2, \ldots, n\}$ |
| $\Theta_i$ | Set of types of player $i$ |
| $S_i$ | Set of actions or pure strategies of player $i$ |
| $\Theta$ | Set of all type profiles $= \Theta_1 \times \Theta_2 \times \ldots \times \Theta_n$ |
| $\theta$ | $\theta = (\theta_1, \ldots, \theta_n) \in \Theta$; a type profile |
| $\Theta_{-i}$ | Set of type profiles of agents except $i = \Theta_1 \times \ldots \Theta_{i-1} \times \Theta_{i+1} \times \ldots \times \Theta_n$ |
| $\theta_{-i}$ | $\theta_{-i} \in \Theta_{-i}$; a profile of types of agents except $i$ |
| $S$ | Set of all action profiles $= S_1 \times S_2 \times \ldots \times S_n$ |
| $p_i$ | A probability (belief) function of player $i$<br>A function from $\Theta_i$ into $\Delta(\Theta_{-i})$ |
| $u_i$ | Utility function of player $i$; $u_i : \Theta \times S \to \mathbb{R}$ |

**Table 2.7** Notation for a Bayesian game

from $\Theta_i$ to $S_i$. A strategy $s_i$ of a player $i$, therefore, specifies a pure action for each type of player $i$; $s_i(\theta_i)$ for a given $\theta_i \in \Theta_i$ would specify the pure action that player $i$ would play if his type were $\theta_i$. The notation $s_i(.)$ is used to refer to the pure action of player $i$ corresponding to an arbitrary type from his type set .

**Definition 2.15 (Consistency of Beliefs).** We say beliefs $(p_i)_{i \in N}$ in a Bayesian game are *consistent* if there is some common prior distribution over the set of type profiles $\Theta$ such that each player's beliefs given his type are just the conditional probability distributions that can be computed from the prior distribution by the Bayes' formula.

*Note 2.7.* If the game is finite, beliefs are consistent if there exists some probability distribution $P \in \Delta(\Theta)$ such that

$$p_i(\theta_{-i}|\theta_i) = \frac{P(\theta_i, \theta_{-i})}{\sum_{t_{-i} \in \Theta_{-i}} P(\theta_i, t_{-i})}$$
$$\forall \theta_i \in \Theta_i;\ \ \forall \theta_{-i} \in \Theta_{-i};\ \ \forall i \in N.$$

Consistency simplifies the definition of the model. The common prior on $\Theta$ determines all the probability functions.

### 2.5.1 Examples of Bayesian Games

*Example 2.22 (A Two Player Bargaining Game).* This example is taken from the book by Myerson [1]. There are two players, player 1 and player 2. Player 1 is the seller of some object, and player 2 is a potential buyer. Each player knows what the object is worth to himself but thinks that its value to the other player may be any integer from 1 to 100 with probability $\frac{1}{100}$. Assume that each player will simultaneously announce a bid between 0 and 100 for trading the object. If the buyer's bid is greater than or equal to the seller's bid they will trade the object at a price equal to the average of their bids; otherwise no trade occurs. For this game:

$$
\begin{aligned}
N &= \{1,2\} \\
\Theta_1 &= \Theta_2 = \{1,2, \ldots ,100\} \\
S_1 &= S_2 = \{0,1,2, \ldots ,100\} \\
p_i(\theta_{-i}|\theta_i) &= \frac{1}{100} \ \forall i \in N \quad \forall(\theta_i, \theta_{-i}) \in \Theta \\
u_1(\theta_1, \theta_2, s_1, s_2) &= \frac{s_1 + s_2}{2} - \theta_1 \quad \text{if } s_2 \geq s_1 \\
&= 0 \qquad\qquad\quad \text{if } s_2 < s_1 \\
u_2(\theta_1, \theta_2, s_1, s_2) &= \theta_2 - \frac{s_1 + s_2}{2} \quad \text{if } s_2 \geq s_1 \\
&= 0 \qquad\qquad\quad \text{if } s_2 < s_1.
\end{aligned}
$$

Note that the type of the seller indicates the willingness to sell (minimum price at which the seller is prepared to sell the item), and the type of the buyer indicates the willingness to pay (maximum price the buyer is prepared to pay for the item).

Also, note that the beliefs are consistent with the prior:

$$P(\theta_1, \theta_2) = \frac{1}{10000} \ \forall \theta_1 \in \Theta_1 \ \forall \theta_2 \in \Theta_2$$

where

$$\Theta_1 \times \Theta_2 = \{1, \ldots ,100\} \times \{1, \ldots ,100\}.$$

*Example 2.23 (Sealed Bid Auction).* Consider a seller who wishes to sell an indivisible item through an auction. Let there be two prospective buyers who bid for this item. The buyers have their individual valuations for this item. These valuations could be considered as the types of the buyers. Here the game consists of the two bidders, namely the buyers, so $N = \{1,2\}$. The two bidders submit bids, say $s_1$ and $s_2$ for the item. Let us say that the one who bids higher is awarded the item with a tie resolved in favor of bidder 1. The winner determination function therefore is:

$$
\begin{aligned}
f_1(s_1, s_2) &= 1 \ \text{ if } \ s_1 \geq s_2 \\
&= 0 \ \text{ if } \ s_1 < s_2
\end{aligned}
$$

$$
\begin{aligned}
f_2(s_1, s_2) &= 1 \ \text{ if } \ s_1 < s_2 \\
&= 0 \ \text{ if } \ s_1 \geq s_2.
\end{aligned}
$$

Assume that the valuation set for each buyer is the real interval $[0,1]$ and also that the strategy set for each buyer is again $[0,1]$. This means $\Theta_1 = \Theta_2 = [0,1]$ and $S_1 = S_2 = [0,1]$. If we assume that each player believes that the other player's valuation is chosen according to an independent uniform distribution, then note that

$$p_i([x,y]|\theta_i) = y - x \quad \forall \ 0 \leq x \leq y \leq 1; \quad i = 1,2.$$

In a first price auction, the winner will pay what is bid by her, and therefore the utility function of the players is given by

$$u_i(\theta_1, \theta_2, s_1, s_2) = f_i(s_1, s_2)(\theta_i - s_i); \;\; i = 1, 2.$$

This completes the definition of the Bayesian game underlying a first price auction involving two bidders. One can similarly develop the Bayesian game for the second price sealed bid auction.

### 2.5.2 Type Agent Representation and the Selten Game

This is a representation of Bayesian games that enables a Bayesian game to be transformed to a strategic form game (with complete information). Given a Bayesian game

$$\Gamma = \langle N, (\Theta_i), (S_i), (p_i), (u_i) \rangle$$

the Selten game is an equivalent strategic form game

$$\Gamma^s = \langle N^s, (S^s_j), (U_j) \rangle.$$

The idea used in formulating a Selten game is to have *type agents*. Each player in the original Bayesian game is now replaced with a number of type agents; in fact, a player is replaced by exactly as many type agents as the number of types in the type set of that player. We can safely assume that the type sets of the players are mutually disjoint. The set of players in the Selten game is given by:

$$N^s = \bigcup_{i \in N} \Theta_i.$$

Note that each type agent of a particular player can play precisely the same actions as the player himself. This means that for every $\theta_i \in \Theta_i$,

$$S^s_{\theta_i} = S_i.$$

From now on, we will use $S^s$ and $S$ interchangeably whenever there is no confusion.

The payoff function $U_{\theta_i}$ for each $\theta_i \in \Theta_i$ is the conditionally expected utility to player $i$ in the Bayesian game given that $\theta_i$ is his actual type. It is a mapping with the following domain and co-domain:

$$U_{\theta_i} : \underset{i \in N}{\times} \; \underset{\theta_i \in \Theta_i}{\times} S_i \rightarrow \mathbb{R}.$$

We will explain the way $U_{\theta_i}$ is derived using an example. This example is developed, based on the illustration in the book by Myerson [1].

*Example 2.24 (Selten Game for a Bayesian Pricing Game).* Consider two firms, company 1 and company 2. Company 1 produces a product $x_1$ whereas company 2 produces either product $x_2$ or product $y_2$. The product $x_2$ is somewhat similar to product $x_1$ while the product $y_2$ is a different line of product. The product to be produced by company 2 is a closely guarded secret, so it can be taken as private information of company 2. We thus have $N = \{1,2\}$, $\Theta_1 = \{x_1\}$, and $\Theta_2 = \{x_2, y_2\}$. Each firm has to choose a price for the product it produces, and this is the strategic decision to be taken by the company. Company 1 has the choice of choosing a low price $a_1$ or a high price $b_1$ whereas company 2 has the choice of choosing a low price $a_2$ or a high price $b_2$. We therefore have $S_1 = \{a_1, b_1\}$ and $S_2 = \{a_2, b_2\}$. The type of company 1 is common knowledge since $\Theta_1$ is a singleton. Therefore, the belief probabilities of company 2 about company 1 are given by $p_2(x_1|x_2) = 1$ and $p_2(x_1|y_2) = 1$. Let us assume the belief probabilities of company 1 about company 2 to be $p_1(x_2|x_1) = 0.6$ and $p_1(y_2|x_1) = 0.4$. To complete the definition of the Bayesian game, we now have to specify the utility functions. Let the utility functions for the two possible type profiles $\theta_1 = x_1,\ \theta_2 = x_2$ and $\theta_1 = x_1,\ \theta_2 = y_2$ be given as in Tables 2.8 and 2.9.

| | 2 | |
|---|---|---|
| 1 | $a_2$ | $b_2$ |
| $a_1$ | 1,2 | 0, 1 |
| $b_1$ | 0,4 | 1, 3 |

**Table 2.8** $u_1$ and $u_2$ for $\theta_1 = x_1; \theta_2 = x_2$

| | 2 | |
|---|---|---|
| 1 | $a_2$ | $b_2$ |
| $a_1$ | 1,3 | 0, 4 |
| $b_1$ | 0,1 | 1, 2 |

**Table 2.9** $u_1$ and $u_2$ for $\theta_1 = x_1; \theta_2 = y_2$

This completes the description of the Bayesian game. We now derive the equivalent Selten game:

$$\langle N^s, (S_{\theta_i})_{\substack{\theta_i \in \Theta_i \\ i \in N}}, (U_{\theta_i})_{\substack{\theta_i \in \Theta_i \\ i \in N}} \rangle.$$

We have

$$N^s = \Theta_1 \cup \Theta_2 = \{x_1, x_2, y_2\}$$
$$S_{x_1} = S_1 = \{a_1, b_1\}$$
$$S_{x_2} = S_{y_2} = S_2 = \{a_2, b_2\}.$$

Note that

$$U_{\theta_i} : S_1 \times S_2 \times S_2 \to \mathbb{R} \;\; \forall \theta_i \in \Theta_i, \forall i \in N$$

$$S_1 \times S_2 \times S_2 = \{(a_1,a_2,a_2),(a_1,a_2,b_2),(a_1,b_2,a_2),(a_1,b_2,b_2),(b_1,a_2,a_2), (b_1,a_2,b_2),(b_1,b_2,a_2),(b_1,b_2,b_2)\}.$$

The above set gives the set of all strategy profiles of all the type agents. A typical strategy profile can be represented as $(s_{x_1}, s_{x_2}, s_{y_2})$. This could also be represented as $(s_1(.), s_2(.))$ where the strategy $s_1$ is a mapping from $\Theta_1$ to $S_1$, and the strategy $s_2$ is a mapping from $\Theta_2$ to $S_2$. In general, for an $n$ player Bayesian game, a pure strategy profile is of the form

$$((s_{\theta_1})_{\theta_1 \in \Theta_1}, (s_{\theta_2})_{\theta_2 \in \Theta_2}, \ldots, (s_{\theta_n})_{\theta_n \in \Theta_n}).$$

Another way to write this would be $(s_1(.), s_2(.), \ldots, s_n(.))$, where $s_i$ is a mapping from $\Theta_i$ to $S_i$ for $i = 1, 2, \ldots, n$.

The payoffs for type agents (in the Selten game) are obtained as conditional expectations over the type profiles of the rest of the agents. For example, let us compute the payoff $U_{x_1}(a_1, a_2, a_2)$, which is the expected payoff obtained by type agent $x_1$ (belonging to player 1 ) when this type agent plays action $a_1$ and the type agents $x_2$ and $y_2$ of player 2 play the actions $a_2$ and $a_2$ respectively. In this case, the type of player 1 is known, but the type of player could be $x_2$ or $y_2$ with probabilities given by the belief function $p_1(.|x_1)$. The following conditional expectation gives the required payoff.

$$\begin{aligned} U_{x_1}(a_1,a_2,a_2) &= p_1(x_2|x_1)u_1(x_1,x_2,a_1,a_2) \\ &\quad + p_1(y_2|x_1)u_1(x_1,y_2,a_1,a_2) \\ &= (0.6)(1) + (0.4)(1) \\ &= 0.6 + 0.4 \\ &= 1. \end{aligned}$$

Similarly, the payoff $U_{x_1}(a_1, a_2, b_2)$ can be computed as follows.

$$\begin{aligned} U_{x_1}(a_1,a_2,b_2) &= p_1(x_2|x_1)u_1(x_1,x_2,a_1,a_2) \\ &\quad + p_1(y_2|x_1)u_1(x_1,y_2,a_1,b_2) \\ &= (0.6)(1) + (0.4)(0) \\ &= 0.6. \end{aligned}$$

It can be similarly shown that

$$
\begin{aligned}
U_{x_1}(b_1, a_2, a_2) &= 0 \\
U_{x_1}(b_1, a_2, b_2) &= 0.4 \\
U_{x_2}(a_1, a_2, b_2) &= 2 \\
U_{x_2}(a_1, b_2, b_2) &= 1 \\
U_{y_2}(a_1, a_2, b_2) &= 4 \\
U_{y_2}(a_1, a_2, a_2) &= 3.
\end{aligned} \tag{2.3}
$$

From the above, we see that

$$
\begin{aligned}
U_{x_1}(a_1, a_2, b_2) &> U_{x_1}(b_1, a_2, b_2) \\
U_{x_2}(a_1, a_2, b_2) &> U_{x_2}(a_1, b_2, b_2) \\
U_{y_2}(a_1, a_2, b_2) &> U_{y_2}(a_1, a_2, a_2).
\end{aligned} \tag{2.4}
$$

From this, we can conclude that the action profile $(a_1, a_2, b_2)$ is a Nash equilibrium of the type agent representation.

#### 2.5.2.1 Payoff Computation in Selten Game

From now on, when there is no confusion, we will use $u$ instead of $U$. In general, given: **(1)** a Bayesian game $\Gamma = \langle N, (\Theta_i), (S_i), (p_i), (u_i) \rangle$, **(2)** its equivalent Selten game $\Gamma^s = \langle N^s, (S_{\theta_i}), (U_{\theta_i}) \rangle$, and **(3)** an action profile in the type agent representation of the form

$$((s_{\theta_1})_{\theta_1 \in \Theta_1}, (s_{\theta_2})_{\theta_2 \in \Theta_2}, \ldots, (s_{\theta_n})_{\theta_n \in \Theta_n}),$$

the payoffs $u_{\theta_i}$ for $\theta_i \in \Theta_i$ ($i \in N$) are computed as follows.

$$u_{\theta_i}(s_{\theta_i}, s_{-\theta_i}) = \sum_{t_{-i} \in \Theta_{-i}} p_i(t_{-i}|\theta_i) u_i(\theta_i, t_{-i}, s_{\theta_i}, s_{t_{-i}})$$

where $s_{t_{-i}}$ is the strategy profile corresponding to the type agents in $t_{-i}$. A concise way of writing the above would be:

$$u_{\theta_i}(s_{\theta_i}, s_{-\theta_i}) = E_{\theta_{-i}}[u_i(\theta_i, \theta_{-i}, s_{\theta_i}, s_{\theta_{-i}})].$$

The notation $u_{\theta_i}$ refers to the utility of player $i$ conditioned on the type being equal to $\theta_i$. We will be using this notation frequently in this section. With this setup, we now look into the notion of an equilibrium in Bayesian games.

### 2.5.3 Equilibria in Bayesian Games

**Definition 2.16 (Pure Strategy Bayesian Nash Equilibrium).** A pure strategy Bayesian Nash equilibrium in a Bayesian game

$$\Gamma = \langle N, (\Theta_i), (S_i), (p_i), (u_i) \rangle$$

can be defined in a natural way as a pure strategy Nash equilibrium of the equivalent Selten game. That is, a profile of type agent strategies

$$s^* = ((s^*_{\theta_1})_{\theta_1 \in \Theta_1}, (s^*_{\theta_2})_{\theta_2 \in \Theta_2}, \ldots, (s^*_{\theta_n})_{\theta_n \in \Theta_n})$$

is said to be a pure strategy Bayesian Nash equilibrium of $\Gamma$ if $\forall i \in N$, $\forall \theta_i \in \Theta_i$,

$$u_{\theta_i}(s^*_{\theta_i}, s^*_{-\theta_i}) \geq u_{\theta_i}(s_i, s^*_{-\theta_i}) \;\; \forall s_i \in S_i.$$

Alternatively, a strategy profile $(s^*_1(.), s^*_2(.), \ldots, s^*_n(.))$ is said to be a Bayesian Nash equilibrium if

$$u_{\theta_i}(s^*_i(\theta_i), s^*_{-i}(\theta_{-i})) \geq u_{\theta_i}(s_i, s^*_{-i}(\theta_{-i})) \;\; \forall s_i \in S_i \;\; \forall \theta_i \in \Theta_i \;\; \forall \theta_{-i} \in \Theta_{-i} \;\; \forall i \in N.$$

*Example 2.25 (Pure Strategy Bayesian Nash Equilibrium).* Consider the example being discussed. We make the following observations.

- When $\theta_2 = x_2$, the strategy $b_2$ is strongly dominated by $a_2$. Thus player 2 chooses $a_2$ when $\theta_2 = x_2$.
- When $\theta_2 = y_2$, the strategy $a_2$ is strongly dominated by $b_2$ and therefore player 2 chooses $b_2$ when $\theta_2 = y_2$.
- When the action profiles are $(a_1, a_2)$ or $(b_1, b_2)$, player 1 has payoff 1 regardless of the type of player 2. In all other profiles, payoff of player 1 is zero.
- Since $p_1(x_2|x_1) = 0.6$ and $p_1(y_2|x_1) = 0.4$, player 1 thinks that the type $x_2$ of player 2 is more likely than type $y_2$.

The above arguments show that the unique pure strategy Bayesian Nash equilibrium in the above example is given by:

$$(s^*_{x_1} = a_1, s^*_{x_2} = a_2, s^*_{y_2} = b_2)$$

thus validating what we have already shown. Note that the equilibrium strategy for company 1 is always to price the product low whereas for company 2, the equilibrium strategy is to price it low if it produces $x_2$ and to price it high if it produces $y_2$.

The above example also illustrates the danger of analyzing each matrix separately. If it is common knowledge that player 2's type is $x_2$, then the unique Nash equilibrium is $(a_1, a_2)$. If it is common knowledge that player 2 has type $y_2$, then we get $(b_1, b_2)$ as the unique Nash equilibrium. However, in a Bayesian game, the type of player 2 is not common knowledge, and hence the above prediction based on analyzing the matrices separately would be wrong.

*Example 2.26 (Bayesian Nash Equilibrium of First Price Sealed Bid Auction).* Consider an auctioneer or a seller and two potential buyers as in Example 2.23. Here each buyer submits a sealed bid, $s_i \geq 0 \ \ (i = 1,2)$. The sealed bids are looked at, and the buyer with the higher bid is declared the winner. If there is a tie, buyer 1 is declared the winner. The winning buyer pays to the seller an amount equal to his bid. The losing bidder does not pay anything.

Let us make the following assumptions:

1. $\theta_1, \theta_2$ are independently drawn from the uniform distribution on $[0,1]$.
2. The sealed bid of buyer $i$ takes the form $s_i(\theta_i) = \alpha_i \theta_i$, where $\alpha_i \in [0,1]$. This assumption implies that player $i$ bids a fraction $\alpha_i$ of his value; this is a reasonable assumption that implies a linear relationship between the bid and the value.

Buyer 1's problem is now to bid in a way to maximize his expected payoff:

$$\max_{s_1 \geq 0} (\theta_1 - s_1) P\{s_2(\theta_2) \leq s_1\}.$$

Since the bid of player 2 is $s_2(\theta_2) = \alpha_2 \theta_2$ and $\theta_2 \in [0,1]$, the maximum bid of buyer 2 is $\alpha_2$. Buyer 1 knows this and therefore $s_1 \in [0, \alpha_2]$. Also,

$$\begin{aligned} P\{s_2(\theta_2) \leq s_1\} &= P\{\alpha_2 \theta_2 \leq s_1\} \\ &= P\{\theta_2 \leq \frac{s_1}{\alpha_2}\} \\ &= \frac{s_1}{\alpha_2} \text{ (since } \theta_2 \text{ is uniform over } [0,1]). \end{aligned}$$

Thus buyer 1's problem is:

$$\max_{s_1 \in [0,\alpha_2]} (\theta_1 - s_1) \frac{s_1}{\alpha_2}.$$

The solution to this problem is

$$s_1(\theta_1) = \begin{cases} \frac{1}{2}\theta_1 & \text{if } \frac{1}{2}\theta_1 \leq \alpha_2 \\ \alpha_2 & \text{if } \frac{1}{2}\theta_1 > \alpha_2. \end{cases}$$

We can show on similar lines that

$$s_2(\theta_2) = \begin{cases} \frac{1}{2}\theta_2 & \text{if } \frac{1}{2}\theta_2 \leq \alpha_1 \\ \alpha_1 & \text{if } \frac{1}{2}\theta_2 > \alpha_1. \end{cases}$$

Let $\alpha_1 = \alpha_2 = \frac{1}{2}$. Then we get

$$\begin{aligned} s_1(\theta_1) &= \frac{\theta_1}{2} \qquad \forall \ \theta_1 \in \Theta_1 = [0,1] \\ s_2(\theta_2) &= \frac{\theta_2}{2} \qquad \forall \ \theta_2 \in \Theta_2 = [0,1]. \end{aligned}$$

Note that if $s_2(\theta_2) = \frac{\theta_2}{2}$, the best response of buyer 1 is $s_1(\theta_1) = \frac{\theta_1}{2}$ and vice-versa. Hence the profile $\left(\frac{\theta_1}{2}, \frac{\theta_2}{2}\right)$ is a Bayesian Nash equilibrium.

### *2.5.4 Dominant Strategy Equilibria*

The dominant strategy equilibria of Bayesian games can again be defined using the Selten game representation.

**Definition 2.17 (Strongly Dominant Strategy Equilibrium).** Given a Bayesian game

$$\Gamma = \langle N, (\Theta_i), (S_i), (p_i), (u_i) \rangle$$

a profile of type agent strategies $(s_1^*(.), s_2^*(.), \ldots, s_n^*(.))$ is said to be a strongly dominant strategy equilibrium if

$$u_{\theta_i}(s_i^*(\theta_i), s_{-i}(\theta_{-i})) > u_{\theta_i}(s_i, s_{-i}(\theta_{-i}))$$

$$\forall s_i \in S_i \setminus \{s_i^*(\theta_i)\},\ \forall s_{-i}(\theta_{-i}) \in S_{-i},\ \forall \theta_i \in \Theta_i,\ \forall \theta_{-i} \in \Theta_{-i},\ \forall i \in N.$$

**Definition 2.18 (Weakly Dominant Strategy Equilibrium).** A profile of type agent strategies $(s_1^*(.), s_2^*(.), \ldots, s_n^*(.))$ is said to be a weakly dominant strategy equilibrium if

$$u_{\theta_i}(s_i^*(\theta_i), s_{-i}(\theta_{-i})) \geq u_{\theta_i}(s_i, s_{-i}(\theta_{-i}))$$

$$\forall s_i \in S_i,\ \forall s_{-i}(\theta_{-i}) \in S_{-i},\ \forall \theta_i \in \Theta_i,\ \forall \theta_{-i} \in \Theta_{-i},\ \forall i \in N$$

and strict inequality satisfied for at least one $s_i \in S_i$.

*Note 2.8.* The notion of dominant strategy equilibrium is independent of the belief functions, and this is what makes it a very powerful notion and a very strong property. The notion of a weakly dominant strategy equilibrium is used extensively in mechanism design theory to define *dominant strategy implementation* of mechanisms.

*Example 2.27 (Weakly Dominant Strategy Equilibrium of Second Price Auction).* We have shown above that the first price sealed bid auction has a Bayesian Nash equilibrium. Now we consider the second price sealed bid auction with two bidders and show that it has a weakly dominant strategy equilibrium. Let us say buyer 2 announces his bid as $\hat{\theta}_2$. There are two cases.

1. $\theta_1 \geq \hat{\theta}_2$.
2. $\theta_1 < \hat{\theta}_2$.

**Case 1:** $\theta_1 \geq \hat{\theta}_2$

Let $\hat{\theta}_1$ be the announcement of buyer 1. Here there are two cases.

- If $\hat{\theta}_1 \geq \hat{\theta}_2$, then the payoff for buyer 1 is $\theta_1 - \hat{\theta}_2 \geq 0$.
- If $\hat{\theta}_1 < \hat{\theta}_2$, then the payoff for buyer 1 is 0.
- Thus in this case, the maximum payoff possible is $\theta_1 - \hat{\theta}_2 \geq 0$.

If $\hat{\theta}_1 = \theta_1$ (that is, buyer 1 announces his true valuation), then payoff for buyer 1 is $\theta_1 - \hat{\theta}_2$, which happens to be the maximum possible payoff as shown above. Thus announcing $\theta_1$ is a best response to buyer 1 whatever the announcement of buyer 2.

**Case 2:** $\theta_1 < \hat{\theta}_2$

Here again there are two cases: $\hat{\theta}_1 \geq \hat{\theta}_2$ and $\hat{\theta}_1 < \hat{\theta}_2$.

- If $\hat{\theta}_1 > \hat{\theta}_2$, then the payoff for buyer 1 is $\theta_1 - \hat{\theta}_2$, which is negative.
- If $\hat{\theta}_1 < \hat{\theta}_2$, then buyer 1 does not win and payoff for him is zero.
- Thus in this case, the maximum payoff possible is 0.

If $\hat{\theta}_1 = \theta_1$, payoff for buyer 1 is 0. By announcing $\hat{\theta}_1 = \theta_1$, his true valuation, buyer 1 gets zero payoff, which in this case is a best response.

We can now make the following observations about this example.

- Bidding his true valuation is optimal for buyer 1 regardless of what buyer 2 announces.
- Similarly bidding his true valuation is optimal for buyer 2 whatever the announcement of buyer 1.
- This means truth revelation is a weakly dominant strategy for each player, and $(\theta_1, \theta_2)$ is a weakly dominant strategy equilibrium.

## 2.6 The Mechanism Design Environment

Mechanism design is concerned with how to implement *system-wide solutions* to problems that involve *multiple self-interested agents*, each with *private information* about their preferences. A mechanism could be viewed as an institution or a framework of protocols that would prescribe particular ways of interaction among the agents so as to ensure a socially desirable outcome from this interaction. Without the mechanism, the interaction among the agents may lead to an outcome that is far from socially optimal. One can view mechanism design as an approach to solving a well-formulated but *incompletely specified optimization problem* where *some of the inputs to the problem are held by the individual agents*. So in order to solve the problem, the *social planner* needs to elicit these private values from the individual agents.

The following provides a general setting for formulating, analyzing, and solving mechanism design problems.

- There are $n$ agents, $1, 2, \ldots, n$, with $N = \{1, 2, \ldots, n\}$. The agents are rational and intelligent.

- $X$ is a set of *alternatives* or *outcomes*. The agents are required to make a collective choice from the set $X$.
- Prior to making the collective choice, each agent privately observes his preferences over the alternatives in $X$. This is modeled by supposing that agent $i$ privately observes a parameter or signal $\theta_i$ that determines his preferences. The value of $\theta_i$ is known to agent $i$ and is not known to the other agents. $\theta_i$ is called a private value or type of agent $i$.
- We denote by $\Theta_i$ the set of private values of agent $i$, $i = 1, 2, \ldots, n$. The set of all type profiles is given by $\Theta = \Theta_1 \times \ldots \times \Theta_n$. A typical type profile is represented as $\theta = (\theta_1, \ldots, \theta_n)$.
- It is assumed that there is a common prior distribution $\Phi \in \Delta(\Theta)$. To maintain consistency of beliefs, individual belief functions $p_i$ that describe the beliefs that player $i$ has about the type profiles of the rest of the players can all be derived from the common prior.
- Individual agents have preferences over outcomes that are represented by a utility function $u_i : X \times \Theta_i \to \mathbb{R}$. Given $x \in X$ and $\theta_i \in \Theta_i$, the value $u_i(x, \theta_i)$ denotes the payoff that agent $i$ having type $\theta_i \in \Theta_i$ receives from a decision $x \in X$. In the more general case, $u_i$ depends not only on the outcome and the type of player $i$, but could depend on the types of the other players also, so $u_i : X \times \Theta \to \mathbb{R}$. We restrict our attention to the former case in this monograph since most real-world situations fall into the former category.
- The set of outcomes $X$, the set of players $N$, the type sets $\Theta_i$ $(i = 1, \cdots, n)$, the common prior distribution $\Phi \in \Delta(\Theta)$, and the payoff functions $u_i$ $(i = 1, \cdots, n)$ are assumed to be *common knowledge* among all the players. The specific value $\theta_i$ observed by agent $i$ is private information of agent $i$.

### Social Choice Functions

Since the agents' preferences depend on the realization of their types $\theta = (\theta_1, \cdots, \theta_n)$, it is natural to make the collective decision to depend on $\theta$. This leads to the definition of a social choice function.

**Definition 2.19 (Social Choice Function).** Given a set of agents $N = \{1, 2, \ldots, n\}$, their type sets $\Theta_1, \Theta_2, \ldots, \Theta_n$, and a set of outcomes $X$, a social choice function is a mapping

$$f : \Theta_1 \times \cdots \times \Theta_n \to X$$

that assigns to each possible type profile $(\theta_1, \theta_2, \ldots, \theta_n)$ a collective choice from the set of alternatives.

*Example 2.28 (Shortest Path Problem with Incomplete Information).* Consider a connected directed graph with a source vertex and destination vertex identified. Let the graph have $n$ edges, each owned by a rational and intelligent agent. Let the set of agents be denoted by $N = \{1, 2, \ldots, n\}$. Assume that the cost of the edge is private information of the agent owning the edge and let $\theta_i$ be this private information for agent $i$ ($i = 1,2,\ldots,n$). Let us say that a social planner is interested in finding a shortest path from the source vertex to the destination vertex. The social planner knows everything about the graph except the costs of the edges. So, the social planner first needs to extract this information from each agent and then find a shortest path from the source vertex to the destination vertex. Thus there are two problems facing the social planner, which are described below.

### Preference Elicitation Problem

Consider a social choice function $f : \Theta_1 \times \ldots \times \Theta_n \to X$. The types $\theta_1, \cdots, \theta_n$ of the individual agents are private information of the agents. Hence for the social choice $f(\theta_1, \cdots, \theta_n)$ to be chosen when the individual types are $\theta_1, \cdots, \theta_n$, each agent must disclose its true type to the social planner. However, given a social choice function $f$, a given agent may not find it in its best interest to reveal this information truthfully. This is called the *preference elicitation* problem or the *information revelation* problem. In the shortest path problem with incomplete information, the preference elicitation problem is to elicit the true values of the costs of the edges from the respective edge owners.

### Preference Aggregation Problem

Once all the agents report their types, the profile of reported types has to be transformed to an outcome, based on the social choice function. Let $\theta_i$ be the true type and $\hat{\theta}_i$ the reported type of agent $i$ ($i = 1,\ldots,n$). The process of computing $f(\hat{\theta}_1,\ldots,\hat{\theta}_n)$ is called the *preference aggregation* problem. In the shortest path problem with incomplete information, the preference aggregation problem is to compute a shortest path from the source vertex to the destination vertex, given the structure of the graph and the (reported) costs of the edges. The preference aggregation problem is usually an optimization problem. Figure 2.1 provides a pictorial representation of all the elements making up the mechanism design environment.

### Direct and Indirect Mechanisms

One can view mechanism design as the process of solving an incompletely specified optimization problem where the specification is first elicited and then the underlying optimization problem is solved. Specification elicitation is basically the preference elicitation or type elicitation problem. To elicit the type information from the agents in a truthful way, there are broadly two kinds of mechanisms, which are aptly called *indirect mechanisms* and *direct mechanisms*. We define these below. In these definitions, we assume that the set of agents $N$, the set of outcomes $X$, the sets of types $\Theta_1,\ldots,\Theta_n$, a common prior $\Phi \in \Delta(\Theta)$, and the utility functions $u_i : X \times \Theta_i \to \mathbb{R}$ are given and are common knowledge.

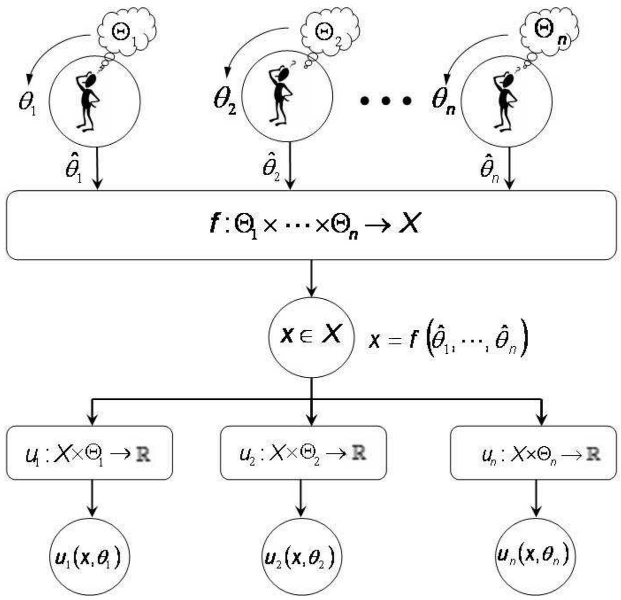



**Fig. 2.1** Mechanism design environment

**Definition 2.20 (Direct Mechanism).** Given a social choice function $f : \Theta_1 \times \Theta_2 \times \ldots \times \Theta_n \rightarrow X$, a direct (revelation) mechanism consists of the tuple $(\Theta_1, \Theta_2, \ldots, \Theta_n, f(.))$.

The idea of a direct mechanism is to *directly* seek the type information from the agents by asking them to reveal their true types.

**Definition 2.21 (Indirect Mechanism).** An indirect (revelation) mechanism consists of a tuple $(S_1, S_2, \ldots, S_n, g(.))$ where $S_i$ is a set of possible actions for agent $i$ $(i = 1, 2, \ldots, n)$ and $g : S_1 \times S_2 \times \ldots \times S_n \rightarrow X$ is a function that maps each action profile to an outcome.

The idea of an indirect mechanism is to provide a choice of actions to each agent and specify an outcome for each action profile. This induces a game among the players and the strategies played by the agents in an equilibrium of this game will indirectly reflect their original types.

In the next four sections of this chapter, we will understand the process of mechanism design in the following way. First, we provide an array of examples to understand social choice functions and to appreciate the need for mechanisms. Next, we understand the process of implementing social choice functions through mechanisms. Following this, we will introduce the important notion of incentive compatibility and present a fundamental result in mechanism design, the *revelation theorem*. Then we will look into different properties that we would like a social choice function to satisfy.

## 2.7 Examples of Social Choice Functions

*Example 2.29 (Technology Driven Supplier Selection).* Suppose there is a buyer who wishes to procure a certain volume of an item that is produced by two suppliers, call them 1 and 2. We have $N = \{1,2\}$. Supplier 1 is known to use technology $a_1$ to produce these items, while supplier 2 uses one of two possible technologies, a high end technology $a_2$ and a low end technology $b_2$. The technology $a_2$ is known to be superior to $a_1$ also. The technology elements could be taken as the types of the suppliers, so we have $\Theta_1 = \{a_1\};\ \Theta_2 = \{a_2, b_2\}$. See Figure 2.2.

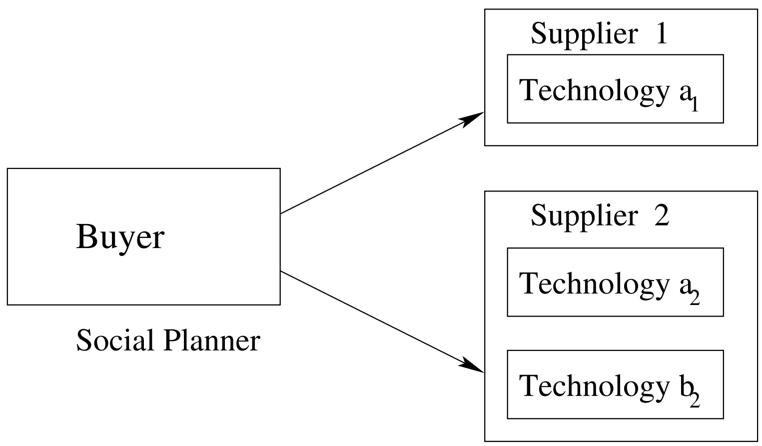



**Fig. 2.2** A sourcing scenario with one buyer and two suppliers

Let us define three outcomes (alternatives) $x, y, z$ for this situation. The alternative $x$ means that the entire volume required is sourced from supplier 1 while the alternative $z$ means that the entire volume required is sourced from supplier 2. The alternative $y$ indicates that 50% of the requirement is sourced from supplier 1 and the rest is sourced from supplier 2.

Since the buyer already has a long-standing relationship with supplier 1, this supplier is the preferred one. However, because of the superiority of technology $a_2$ over $a_1$ and $b_2$, the buyer would like to certainly source some quantity from supplier 2 if it is known that supplier 2 is guaranteed to use technology $a_2$. To reflect these facts, we assume the payoff functions to be given by:

$$u_1(x, a_1) = 100;\ u_1(y, a_1) = 50;\ u_1(z, a_1) = 0$$

$$u_2(x, a_2) = 0;\ u_2(y, a_2) = 50;\ u_2(z, a_2) = 100$$

$$u_2(x, b_2) = 0;\ u_2(y, b_2) = 50;\ u_2(z, b_2) = 25.$$

Note that $\Theta = \{(a_1, a_2), (a_1, b_2)\}$. Consider the social choice function $f(a_1, a_2) = y$ and $f(a_1, b_2) = x$. This means that when it is guaranteed that supplier 2 will use technology $a_2$, the buyer would like to procure from both the suppliers, whereas if it is known that supplier 2 uses technology $b_2$, the buyer would rather source the entire requirement from supplier 1.

Likewise, there are eight other social choice functions that one can define here. These are:

$$f(a_1,a_2) = x; \quad f(a_1,b_2) = x$$
$$f(a_1,a_2) = x; \quad f(a_1,b_2) = y$$
$$f(a_1,a_2) = x; \quad f(a_1,b_2) = z$$
$$f(a_1,a_2) = y; \quad f(a_1,b_2) = y$$
$$f(a_1,a_2) = y; \quad f(a_1,b_2) = z$$
$$f(a_1,a_2) = z; \quad f(a_1,b_2) = x$$
$$f(a_1,a_2) = z; \quad f(a_1,b_2) = y$$
$$f(a_1,a_2) = z; \quad f(a_1,b_2) = z.$$

It would be interesting to look at the implications of these social choice functions. We will return to this example later on, in many different contexts.

*Example 2.30 (Procurement of a Single Indivisible Resource).* Procurement is a ubiquitous activity in any organization. Every organization procures a variety of direct and indirect materials. For example, a factory procures raw material or sub-assemblies from a pool of suppliers. In a computational grid, a grid user procures computational or storage resources from the grid. A network user procures network resources. A dynamic supply chain planner procures supply chain service providers. Every organization procures indirect materials such as office supplies and services.

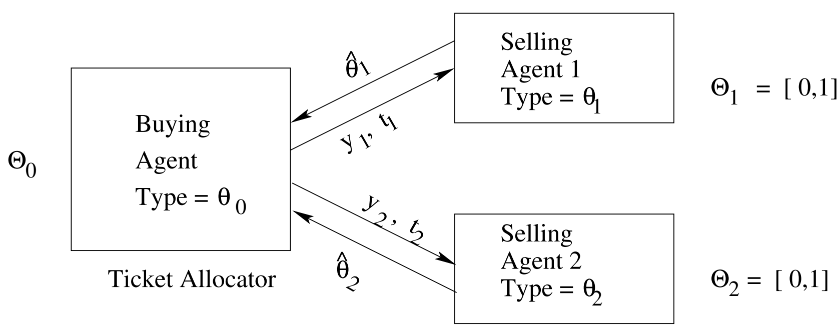



**Fig. 2.3** Procurement with two suppliers

The generic procurement situation involves a buyer, a pool of suppliers, and items to be procured. We consider a simple abstraction of the problem by considering a buying agent (call the agent 0) and two selling agents (call them 1 and 2), so we have $N = \{0,1,2\}$. See Figure 2.3. An indivisible item or resource is to be procured from one of the sellers in return for a monetary consideration. An outcome here can be represented by $x = (y_0, y_1, y_2, t_0, t_1, t_2)$. For $i = 0$, we have

$$\begin{aligned} y_0 &= 0 \quad \text{if the buyer buys the good} \\ &= 1 \quad \text{otherwise} \\ t_0 &= \text{monetary transfer received by the buyer.} \end{aligned}$$

For $i = 1,2$, we have

$$\begin{aligned} y_i &= 1 \quad \text{if agent } i \text{ supplies the good to the buyer} \\ &= 0 \quad \text{if agent } i \text{ does not supply the good} \\ t_i &= \text{monetary transfer received by the agent } i. \end{aligned}$$

The set $X$ of all feasible outcomes is given by

$$X = \{(y_0, y_1, y_2, t_0, t_1, t_2) : y_i \in \{0,1\}, \sum_{i=0}^{2} y_i = 1, t_i \in \mathbb{R}, \sum_{i=0}^{2} t_i \leq 0\}.$$

The constraint $\sum_i t_i \leq 0$ implies that the total money received by all the agents is less than or equal to zero. That is, total money paid by all the agents is greater than or equal to zero (that is, buyer pays at least as much as the sellers receive. The excess between the payment and receipts is the surplus). For $x = (y_0, y_1, y_2, t_0, t_1, t_2)$, define the utilities to be of the form:

$$u_i(x, \theta_i) = u_i((y_0, y_1, y_2, t_0, t_1, t_2), \theta_i) = -y_i \theta_i + t_i \; ; \; i = 1,2$$

where $\theta_i \in \mathbb{R}$ can be viewed as seller $i$'s valuation of the good. Such utility functions are said to be of *quasilinear* form (because it is linear in some of the variables and possibly non-linear in the other variables). We will be studying such utility forms quite extensively in this chapter.

We make the following assumptions regarding valuations.

- The buyer has a *known* value $\underline{\theta_0}$ for the good. This valuation does not depend on the choice of the seller from whom the item is purchased.
- Let $\Theta_i$ be the real interval $[\underline{\theta_i}, \overline{\theta_i}]$. The types $\theta_1$ and $\theta_2$ of the sellers are drawn independently from the interval $[\theta_i, \bar{\theta}_i]$ and this fact is *common knowledge* among all the players. The type of a seller is to be viewed as the *willingness to sell* (minimum price below which the seller is not interested in selling the item).

Consider the following social choice function.

- The buyer buys the good from the seller with the lowest willingness to sell. If both the sellers have the same type, the buyer will buy the object from seller 1.

- The buyer pays the selected selling agent his willingness to sell.

The above social choice function $f(\theta) = (y_0(\theta), y_1(\theta), y_2(\theta), t_0(\theta), t_1(\theta), t_2(\theta))$ can be precisely written as

$$\begin{aligned} y_0(\theta) &= 0 \quad \forall \theta \\ y_1(\theta) &= 1 \quad \text{if } \theta_1 \leq \theta_2 \\ &= 0 \quad \text{if } \theta_1 > \theta_2 \\ y_2(\theta) &= 1 \quad \text{if } \theta_1 > \theta_2 \\ &= 0 \quad \text{if } \theta_1 \leq \theta_2 \end{aligned}$$

$$\begin{aligned} t_1(\theta) &= \theta_1 \, y_1(\theta) \\ t_2(\theta) &= \theta_2 \, y_2(\theta) \\ t_0(\theta) &= -(t_1(\theta) + t_2(\theta)). \end{aligned}$$

We will refer to the above SCF as SCF-PROC1 in the sequel.

Suppose we consider another social choice function, which has the same allocation rule as the one we have just studied but has a different payment rule: The buyer now pays the winning seller a payment equal to the second lowest willingness to sell (as usual, the losing seller does not receive any payment). The new social choice function, which we will call SCF-PROC2, will be the following.

$$\begin{aligned} y_0(\theta) &= 0 \qquad \forall \, \theta \\ y_1(\theta) &= 1 \qquad \text{if } \theta_1 \leq \theta_2 \\ &= 0 \qquad \text{otherwise} \\ y_2(\theta) &= 1 \qquad \text{if } \theta_1 > \theta_2 \\ &= 0 \qquad \text{otherwise} \\ t_1(\theta) &= \theta_2 \, y_1(\theta) \\ t_2(\theta) &= \theta_1 \, y_2(\theta) \\ t_0(\theta) &= -(t_1(\theta) + t_2(\theta)). \end{aligned}$$

Let us define one more SCF, which we call SCF-PROC3, in the following way. SCF-PROC3 has the allocation rule as SCF-PROC1 and SCF-PROC2, but the payments are defined as:

$$\begin{aligned} t_1(\theta) &= \frac{(1+\theta_1)}{2} \; y_1(\theta) \\ t_2(\theta) &= \frac{(1+\theta_2)}{2} \; y_2(\theta) \\ t_0(\theta) &= -(t_1(\theta) + t_2(\theta)). \end{aligned}$$

The reason for defining the payment rule in the above way will become clear in the next section, where we will discuss the implementability of SCF-PROC1, SCF-PROC2, and SCF-PROC3.

*Example 2.31 (Funding a Public Project).* There is a set of agents $N = \{1, 2, \ldots, n\}$ who have a stake in a common infrastructure, for example, a bridge, community building, Internet infrastructure, etc. For example, the agents could be firms forming a business cluster and interested in creating a shared infrastructure. The cost of the project is to be shared by the agents themselves since there is no source of external funding. Let $k = 1$ indicate that the project is taken up, with $k = 0$ indicating that the project is dropped. Let $t_i \in \mathbb{R}$ denote the payment received by agent $i$ (which means $-t_i$ is the payment made by agent $i$) for each $i \in N$. Let the cost of the project be $C$. Since the agents have to fund the project themselves,

$$C \leq -\sum_{i \in N} t_i \quad \text{if } k = 1$$

If $k = 0$, we have

$$0 \leq -\sum_{i \in N} t_i$$

Combining the above two possibilities, we get the condition

$$\sum_{i \in N} t_i \leq -kC \; ; \; k \in \{0, 1\}.$$

Thus, a natural set of outcomes for this problem is:

$$X = \{(k, t_1, \ldots, t_n) : k \in \{0, 1\}, \; t_i \in \mathbb{R} \; \forall i \in N, \; \sum_{i \in N} t_i \leq -kC.\}$$

We assume the utility of agent $i$, when its type is $\theta_i$ corresponding to an outcome $(k, t_1, t_2, \ldots, t_n)$ to be given by

$$u_i((k, t_1, \ldots, t_n), \theta_i) = k\theta_i + t_i.$$

The type $\theta_i$ of agent $i$ has the natural interpretation of being the willingness to pay of agent $i$ (maximum amount that agent $i$ is prepared to pay) towards the project. A social choice function in this context is $f(\theta) = (k(\theta), t_1(\theta), \ldots, t_n(\theta))$ given by

$$k(\theta) = \begin{cases} 1 \text{ if } \sum_{i \in N} \theta_i \geq C \\ 0 \text{ otherwise} \end{cases}$$

$$t_i(\theta) = -\left( \frac{k(\theta) C}{n} \right).$$

The way $k(\theta)$ is defined ensures that the project is taken up only if the combined willingness to pay of all the agents is at least the cost of the project. The definition of $t_i(\theta)$ above follows the egalitarian principle, namely that the agents share the cost of the project equally among themselves.

*Example 2.32 (Bilateral Trade).* Consider two agents 1 and 2 where agent 1 is the seller of an indivisible private good and agent 2 is a prospective buyer of the good. See Figure 2.4. An outcome here is of the form $x = (y_1, y_2, t_1, t_2)$ where $y_i = 1$ if

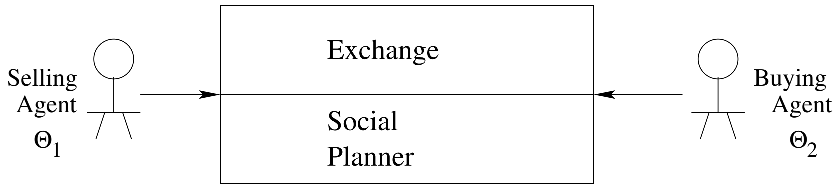



**Fig. 2.4** A bilateral trade environment

agent $i$ gets the good and $t_i$ denotes the payment received by agent $i$ $(i = 1,2)$. A natural set of outcomes here is

$$X = \{(y_1, y_2, t_1, t_2) : \; y_1 + y_2 = 1; \; y_1, y_2 \in \{0,1\}, \; t_1 + t_2 \leq 0\}.$$

The condition $t_1 + t_2 \leq 0$ indicates that the the amount paid by the buyer should be at least equal to the amount received by the seller (the surplus could perhaps be retained by a market maker or mediator). The utility of the agent $i$ $(i = 1,2)$ would be of the form

$$u_i((y_1, y_2, t_1, t_2), \theta_i) = y_i \theta_i + t_i.$$

The type $\theta_1$ of agent 1 (seller) can be interpreted as the willingness to sell of the agent (minimum price at which agent 1 is willing to sell). The type $\theta_2$ of agent 2 (buyer) has the natural interpretation of willingness to pay (maximum price the buyer is willing to pay). A social choice function here would be $f(\theta) = (y_1(\theta), y_2(\theta), t_1(\theta), t_2(\theta))$ defined as

$$\begin{aligned}
y_1(\theta_1, \theta_2) &= 1 \quad \theta_1 > \theta_2 \\
&= 0 \quad \theta_1 \leq \theta_2 \\
y_2(\theta_1, \theta_2) &= 1 \quad \theta_1 \leq \theta_2 \\
&= 0 \quad \theta_1 > \theta_2 \\
t_1(\theta_1, \theta_2) &= y_2(\theta_1, \theta_2) \frac{\theta_1 + \theta_2}{2} \\
t_2(\theta_1, \theta_2) &= -y_2(\theta_1, \theta_2) \frac{\theta_1 + \theta_2}{2}.
\end{aligned}$$

*Example 2.33 (Network Formation Problem).* Consider a supply chain network scenario where a supply chain planner (SCP) is interested in forming an optimal network for delivering products/services. Multiple service providers or supply chain partners are needed for executing the end-to-end process. The directed graph shown

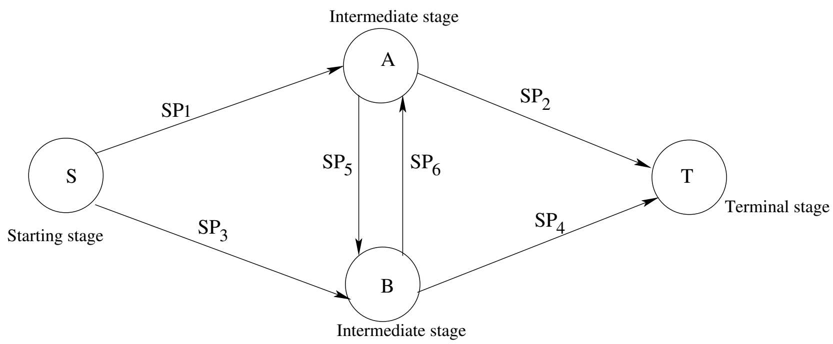



**Fig. 2.5** A graph representation of network formation

in Figure 2.5 describes different possible ways in which the supply chain can be formed. The node $S$ denotes the starting stage, and the node $T$ denotes the terminal stage in the supply chain. $SP_1, SP_2, \ldots, SP_6$ denote the service providers. They own the respective edges of the directed graph. The nodes $A$ and $B$ are two intermediate stages in the supply chain process. Each path in the network from $S$ to $T$ corresponds to a particular way in which the supply chain network can be formed. For example, $(SP_1, SP_5, SP_4)$ is a path from $S$ to $T$ and this is one possible solution to the supply chain formation problem. Each service provider has a service cost that is private information of the provider. The problem facing the supply chain planner here is to determine the network of service providers who will enable forming the supply chain at the least cost. Since the private values of the service providers are not known, the supply chain planner is faced with a mechanism design problem.

A situation such as described above occurs in many other settings. The first example is a logistics network scenario where $S$ denotes the source, $T$ denotes the destination, and the nodes $A$ and $B$ denote two separate logistics hubs. The edges denote the connectivity between logistics hubs, the source, and the destination. The service providers here are logistics providers who own the transportation service on the edges. Another example is a procurement network scenario where the service providers correspond to suppliers. One more example is that of a telecom network where the service providers could correspond to Internet/bandwidth service providers, and the nodes correspond to cities.

For this example, we have $N = \{SCP, SP_1, \ldots, SP_6\}$. For brevity, we will call this set $\{0,1,2,3,4,5,6\}$, where 0 is the supply chain planner and $1,2,\ldots,6$ are the service providers. Let us assume the type sets to be:

$$\Theta_0 = \{\theta_0\}; \quad \Theta_i = \left[\underline{\theta}_i, \bar{\theta}_i\right] \subset \mathbb{R} \;\; \forall\, i \in \{1,2,\ldots,6\}.$$

An outcome is of the form $(y_0, y_1, \ldots, y_6, t_0, t_1, \ldots, t_6)$ where

$y_0$ = 1 if the supply chain planner buys a path from $S$ to $T$
= 0 otherwise.

For $i = 1, \ldots, 6$,

$$\begin{aligned} y_i &= 1 \text{ if } SP_i \text{ is part of the supply chain formed} \\ &= 0 \text{ otherwise} \\ t_i &= \text{ payment received by service provider } SP_i. \end{aligned}$$

Note that $-t_0$, the payment required to be made by the supply chain planner is just the sum of $t_1, \ldots, t_6$. The set of outcomes in this example is given by

$$\begin{aligned} X = \{&(y_0, y_1, ..., y_6, t_0, t_1, ..., t_6) : y_i \in \{0,1\}\ \forall i \in N; \\ &\{SP_i \in N \setminus \{0\} \ :\ y_i = 1\} \text{ forms a path from S to T;} \\ &t_i \in \mathbb{R}\ \forall i \in N;\ t_0 = -\textstyle\sum_{i=1}^{6} t_i\}. \end{aligned}$$

A social choice function in this problem is given by

$$f(\theta) = (y_0(\theta), y_1(\theta), \ldots, y_6(\theta), t_0(\theta), t_1(\theta), \ldots, t_6(\theta))$$

$$y_0(\theta) = 1\ \ \forall \theta \in \Theta.$$

For $i = 1, 2, \ldots, 6$, $\forall \theta \in \Theta$,

$$\begin{aligned} y_i(\theta) &= 1 \text{ if } SP_i \text{ is part of a shortest cost path} \\ &\quad\ \text{from } S \text{ to } T \text{ selected by the supply chain planner} \\ &= 0 \text{ otherwise} \\ t_i(\theta) &= \text{ payment received by } SP_i;\ i = 1, 2, ..., 6 \\ t_0(\theta) &= -\textstyle\sum_{i=1}^{6} t_i(\theta) \\ u_i(f(\theta), \theta_i) &= u_i(y_0(\theta), \ldots, y_6(\theta), t_0(\theta), \ldots, t_6(\theta); \theta_i) \\ &= t_i(\theta) - y_i(\theta)\theta_i. \end{aligned}$$

$\theta_i$ is to be viewed as the willingness to sell of the agent $i$ where $i = 1, 2, \ldots, 6$.

## 2.8 Implementation of Social Choice Functions

In the preceding section, we have seen a series of examples of social choice functions. In this section, we motivate, through illustrative examples, the concept of implementation of social choice functions. We then formally define the notion of implementation through direct mechanisms and indirect mechanisms.

### *2.8.1 Implementation Through Direct Mechanisms*

We first provide examples to motivate implementation by direct mechanisms.

*Example 2.34 (Technology driven Supplier Selection).* Recall Example 2.29 where $N = \{1, 2\}$; $\Theta_1 = \{a_1\}$; , $\Theta_2 = \{a_2, b_2\}$ and $\Theta = \{(a_1, a_2), (a_1, b_2)\}$. Suppose the social planner (in this case, the buyer) wishes to implement the social choice function $f$ with $f(a_1, a_2) = y$ and $f(a_1, b_2) = x$. Announcing this as the social choice function, let us say the social planner asks the agents to reveal their types. Agent 1 has nothing to reveal since his type is common knowledge (as his type set is a singleton). We will now check whether agent 2 would be willing to truthfully reveal its type.

- If $\theta_2 = a_2$, then, because $f(a_1,a_2) = y$ and $f(a_1,b_2) = x$ and $u_2(y,a_2) > u_2(x,a_2)$, agent 2 is happy to reveal $a_2$ as its type.
- However if $\theta_2 = b_2$, then because $u_2(y,b_2) > u_2(x,b_2)$ and $f(a_1,b_2) = x$, agent 2 would wish to lie and claim that its type is $a_2$ and not $b_2$.

Thus though the social planner would like to implement an SCF $f(\cdot)$, the social planner would be unable to implement the above SCF since one of the agents (in this case agent 2) does not find it in his best interest to reveal his true type.

On the other hand, let us say the social planner wishes to implement the social choice function $f$ given by $f(a_1,a_2) = z$ and $f(a_1,b_2) = y$. One can show in this case that the SCF can be implemented. Table 2.10 lists all the nine SCFs and their implementability.

| SCF | | Implementable |
|---|---|---|
| $f(a_1,a_2)$ | $f(a_1,b_2)$ | |
| $x$ | $x$ | $\surd$ |
| $x$ | $y$ | $\times$ |
| $x$ | $z$ | $\times$ |
| $y$ | $x$ | $\times$ |
| $y$ | $y$ | $\surd$ |
| $y$ | $z$ | $\times$ |
| $z$ | $x$ | $\times$ |
| $z$ | $y$ | $\surd$ |
| $z$ | $z$ | $\surd$ |

**Table 2.10** Social choice functions and their implementability

*Example 2.35 (Implementability of SCF-PROC1).* Recall the social choice function SCF-PROC1 that we introduced in the context of procurement of a single indivisible resource (Example 2.30). Recall the definition of SCF-PROC1:

$$\begin{aligned} y_0(\theta) &= 0 \quad \forall\theta \\ y_1(\theta) &= 1 \quad \text{if } \theta_1 \leq \theta_2 \\ &= 0 \quad \text{if } \theta_1 > \theta_2 \\ y_2(\theta) &= 1 \quad \text{if } \theta_1 > \theta_2 \\ &= 0 \quad \text{if } \theta_1 \leq \theta_2 \end{aligned}$$

$$\begin{aligned} t_1(\theta) &= \theta_1\, y_1(\theta) \\ t_2(\theta) &= \theta_2\, y_2(\theta) \\ t_0(\theta) &= -(t_1(\theta) + t_2(\theta)). \end{aligned}$$

We note that the social choice function is very attractive to the buyer since the buyer will capture all of the consumption benefits that are generated by the good. We assume that $\theta_1$ and $\theta_2$ are drawn independently from a uniform distribution over $[0,1]$. Now we ask the question: Can we implement this social choice function? The answer for this question is *no*. The following analysis shows why.

Let us say seller 2 announces his true value $\theta_2$. Suppose the valuation of seller 1 is $\theta_1$, and he announces $\hat{\theta}_1$. If $\theta_2 \geq \hat{\theta}_1$, then seller 1 is the winner, and his utility will be $\hat{\theta}_1 - \theta_1$. If $\theta_2 < \hat{\theta}_1$, then seller 2 is the winner, and seller 1's utility is zero. Since seller 1 wishes to maximize his expected utility he solves the problem

$$\max_{\hat{\theta}_1} (\hat{\theta}_1 - \theta_1)\, P\{\theta_2 \geq \hat{\theta}_1\}.$$

Since $\theta_2$ is uniformly distributed on [0,1],

$$P\{\theta_2 \geq \hat{\theta}_1\} = 1 - P\{\theta_2 < \hat{\theta}_1\} = 1 - \hat{\theta}_1.$$

Thus seller 1 tries to solve the problem:

$$\max_{\hat{\theta}_1} (\hat{\theta}_1 - \theta_1)(1 - \hat{\theta}_1).$$

This problem has the solution

$$\hat{\theta}_1 = \frac{1+\theta_1}{2}.$$

Thus if seller 2 announces his true valuation, then the best response for seller 1 is to announce $\frac{1+\theta_1}{2}$.

Similarly if seller 1 announces his true valuation $\theta_1$, then the best response of seller 2 is to announce $\frac{1+\theta_2}{2}$. Thus there is no incentive for the sellers to announce their true valuations. So, a social planner who wishes to realize the above social choice function finds the rational players will not report their true private values. Thus the social choice function cannot be implemented through a direct mechanism.

*Example 2.36 (Implementability of SCF-PROC2).* Recall the social choice function SCF-PROC2, again in the context of procurement of a single indivisible resource (Example 2.30):

$$\begin{aligned}
y_0(\theta) &= 0 \quad \forall\theta \\
y_1(\theta) &= 1 \quad \text{if } \theta_1 \leq \theta_2 \\
&= 0 \quad \text{if } \theta_1 > \theta_2 \\
y_2(\theta) &= 1 \quad \text{if } \theta_1 > \theta_2 \\
&= 0 \quad \text{if } \theta_1 \leq \theta_2
\end{aligned}$$

$$\begin{aligned}
t_1(\theta) &= \theta_2\, y_1(\theta) \\
t_2(\theta) &= \theta_1\, y_2(\theta) \\
t_0(\theta) &= -(t_1(\theta) + t_2(\theta)).
\end{aligned}$$

We now show that the function SCF-PROC2 can be implemented. Let us say seller 2 announces his valuation as $\hat{\theta}_2$. There are two cases.

1. $\theta_1 \leq \hat{\theta}_2$
2. $\theta_1 > \hat{\theta}_2$.

**Case 1:** $\theta_1 \leq \hat{\theta}_2$

Let $\hat{\theta}_1$ be the announcement of seller 1. Here there are two cases.

- If $\hat{\theta}_1 \leq \hat{\theta}_2$, then the payoff for seller 1 is $\hat{\theta}_2 - \theta_1 \geq 0$.
- If $\hat{\theta}_1 > \hat{\theta}_2$, then the payoff for seller 1 is 0.
- Thus in this case, the maximum payoff possible is $\hat{\theta}_2 - \theta_1 \geq 0$.

If $\hat{\theta}_1 = \theta_1$ (that is, seller 1 announces his true valuation), then payoff for seller 1 is $\hat{\theta}_2 - \theta_1$, which happens to be the maximum possible payoff as shown above. Thus announcing $\theta_1$ is a best response to seller 1 whatever the announcement of seller 2.

**Case 2:** $\theta_1 > \hat{\theta}_2$

Here again there are two cases: $\hat{\theta}_1 \leq \hat{\theta}_2$ and $\hat{\theta}_1 > \hat{\theta}_2$.

- If $\hat{\theta}_1 \leq \hat{\theta}_2$, then the payoff for seller 1 is $\hat{\theta}_2 - \theta_1$, which is negative.
- If $\hat{\theta}_1 > \hat{\theta}_2$, then seller 1 does not win, and payoff for him is zero.
- Thus in this case, the maximum payoff possible is 0.

If $\hat{\theta}_1 = \theta_1$, payoff for seller 1 is 0. By announcing $\hat{\theta}_1 = \theta_1$, his true valuation, seller 1 gets zero payoff, which in this case is a best response.
We can now make the following observations about this example.

- Revealing his true valuation is optimal for seller 1 regardless of what seller 2 announces.
- Similarly, announcing his true valuation is optimal for seller 2 whatever the announcement of seller 1.
- More formally, truth revelation is a weakly dominant strategy for each player.
- Thus this social choice function can be implemented even though the valuations are private values. We simply ask each seller to report his type and then we choose $f(\theta)$.

### *2.8.2 Implementation Through Indirect Mechanisms*

The examples above have shown us a possible way in which to try to implement a social choice function. The protocol we followed for implementing the social choice functions was:

- Ask each agent to reveal his or her types $\theta_i$;
- Given the announcements $(\hat{\theta}_1, \ldots, \hat{\theta}_n)$, choose the outcome $x = f(\hat{\theta}_1, \ldots, \hat{\theta}_n) \in X$.

Such a method of trying to implement an SCF is referred to as a *direct revelation mechanism*. Another approach to implementing a social choice function is the *indirect way*. Here the mechanism makes the agents interact through an institutional framework in which there are rules governing the actions the agents would be allowed to play and in which there is a way of transforming these actions into a social outcome. The actions the agents choose will depend on their private values and become the strategies of the players. Auctions provide an example of *indirect mechanisms*. We provide an example right away.

*Example 2.37 (First Price Procurement Auction).* Consider an auctioneer or a buyer and two potential sellers as before. Here each seller submits a sealed bid, $b_i \geq 0$ $(i = 1, 2)$. The sealed bids are examined and the seller with the lower bid is declared the winner. If there is a tie, seller 1 is declared the winner. The winning seller receives an amount equal to his bid from the buyer. The losing seller does not receive anything.

Note that there is a subtle difference between the situations in Example 2.35 and Example 2.37. In Example 2.35 (direct mechanism), each seller is asked to announce his type, whereas in Example 2.37 (indirect mechanism), each seller is asked to submit a bid. The bid submitted may (and will) of course depend on the type. Based on the type, the seller has a strategy for bidding. So it becomes a game.

Let us make the following assumptions:

1. $\theta_1, \theta_2$ are independently drawn from the uniform distribution on $[0,1]$.
2. The sealed bid of seller $i$ takes the form $b_i(\theta_i) = \alpha_i \theta_i + \beta_i$, where $\alpha_i \in [0,1], \beta_i \in [0, 1-\alpha_i]$. He has to make sure that $b_i \in [0,1]$. The term $\beta_i$ is like a fixed cost whereas $\alpha_i \theta_i$ indicates a fraction of the true cost.

Seller 1's problem is now to bid in a way to maximize his payoff:

$$\max_{1 \geq b_1 \geq 0} (b_1 - \theta_1) P\{b_2(\theta_2) \geq b_1\}$$

$$\begin{aligned}
P\{b_2(\theta_2) \geq b_1\} &= 1 - P\{b_2(\theta_2) < b_1\} \\
&= 1 - P\{\alpha_2 \theta_2 + \beta_2 < b_1\} \\
&= 1 - \frac{b_1 - \beta_2}{\alpha_2} \text{ if } b_1 \geq \beta_2 \qquad (2.5) \\
&\quad \text{since } \theta_2 \text{ is uniform over } [0,1] \\
&= 1 \text{ if } b_1 < \beta_2. \qquad (2.6)
\end{aligned}$$

Thus seller 1's problem is:

$$\max_{b_1 \geq \beta_2} (b_1 - \theta_1)\left(1 - \frac{b_1 - \beta_2}{\alpha_2}\right).$$

The solution to this problem is

$$b_1(\theta_1) = \frac{\alpha_2 + \beta_2}{2} + \frac{\theta_1}{2}. \tag{2.7}$$

We can show on similar lines that

$$b_2(\theta_2) = \frac{\alpha_1 + \beta_1}{2} + \frac{\theta_2}{2}. \tag{2.8}$$

As the bid of seller $i$ takes the form $b_i(\theta_i) = \alpha_i\theta_i + \beta_i$, where $\alpha_i \in [0,1], \beta_i \in [0, 1 - \alpha_i]$, from the equations (2.7) and (2.8), we obtain $\alpha_1 = \alpha_2 = \frac{1}{2}$. As the goal of each seller is to maximize the profit and $\beta_i \in [0, 1-\alpha_i]$, $\beta_1 = \beta_2 = \frac{1}{2}$. Then we get

$$\begin{aligned} b_1(\theta_1) &= \frac{1+\theta_1}{2} \qquad \forall\ \theta_1 \in \Theta_1 = [0,1] \\ b_2(\theta_2) &= \frac{1+\theta_2}{2} \qquad \forall\ \theta_2 \in \Theta_2 = [0,1]. \end{aligned}$$

Note that if $b_2(\theta_2) = \frac{1+\theta_2}{2}$, the best response of seller 1 is $b_1(\theta_1) = \frac{1+\theta_1}{2}$ and vice-versa. Hence the profile $\left(\frac{1+\theta_1}{2}, \frac{1+\theta_2}{2}\right)$ is a Bayesian Nash equilibrium of an underlying Bayesian game. In other words, there is a Bayesian Nash equilibrium of an underlying game (induced by the indirect mechanism called the first price procurement auction) that (indirectly) yields the outcome

$$f(\theta) = (y_0(\theta), y_1(\theta), y_2(\theta), t_0(\theta), t_1(\theta), t_2(\theta))$$

such that

$$\begin{aligned} y_0(\theta) &= 0 \qquad \forall\ \theta \in \Theta \\ y_1(\theta) &= 1 \qquad \text{if } \theta_1 \le \theta_2 \\ &= 0 \qquad \text{else} \\ y_2(\theta) &= 1 \qquad \text{if } \theta_1 > \theta_2 \\ &= 0 \qquad \text{else} \\ t_1(\theta) &= \frac{1+\theta_1}{2} y_1(\theta) \\ t_2(\theta) &= \frac{1+\theta_2}{2} y_2(\theta) \\ t_0(\theta) &= -(t_1(\theta) + t_2(\theta)). \end{aligned}$$

The above SCF is precisely SCF-PROC3 that we had introduced in Example 2.30.

*Example 2.38 (Second Price Procurement Auction).* Here, each seller is asked to submit a sealed bid $b_i \geq 0$. The bids are examined, and the seller with the lower bid is declared the winner. In case there is a tie, seller 1 is declared the winner. The winning seller receives as payment from the buyer an amount equal to the second lowest bid. The losing bidder does not receive anything. In this case, we can show that $b_i(\theta_i) = \theta_i$ for $i = 1,2$ constitutes a weakly dominant strategy for each player. The arguments are identical to those in Example 2.36.

Thus the game induced by the indirect mechanism second price procurement auction has a weakly dominant strategy in which the social choice function SCF-PROC2 is implemented.

We can summarize the findings of the current section so far in the following way.

- The function SCF-PROC1 cannot be implemented.
- The function SCF-PROC2 can be implemented in dominant strategies by a direct mechanism. Also, the indirect mechanism, namely second price procurement auction, implements SCF-PROC2 in dominant strategies.
- The function SCF-PROC3 is implemented in Bayesian Nash equilibrium by an indirect mechanism, the first price procurement auction.

### *2.8.3 Bayesian Game Induced by a Mechanism*

Recall that a mechanism is an institution or a framework with a set of rules that prescribe the actions available to players and specify how these action profiles are transformed into outcomes. A mechanism specifies an action set for each player. The outcome function gives the rule for obtaining outcomes from action profiles. Given:

1. a set of agents $\{1,2, \ldots, n\}$,
2. type sets $\Theta_1, \ldots, \Theta_n$,
3. a common prior $\phi \in \Delta(\Theta)$,
4. a set of outcomes $X$,
5. utility functions $u_1, \ldots, u_n$, with $u_i : X \times \Theta_i \to \mathbb{R}$,

a mechanism $M = (S_1, \ldots, S_n, g(.))$ induces a Bayesian game

$$(N, (\Theta_i), (S_i), (p_i), (U_i))$$

among the players where

$$U_i(\theta_1, \ldots, \theta_n, s_1, \ldots, s_n) = u_i(g(s_1, \ldots, s_n), \theta_i).$$

#### Strategies in the Induced Bayesian Game

A strategy $s_i$ for an agent $i$ in the induced Bayesian game is a function $s_i : \Theta_i \to S_i$. Thus, given a private value $\theta_i \in \Theta_i$, $s_i(\theta_i)$ will give the action of player $i$. The strategy $s_i(.)$ will specify actions corresponding to private values. In the auction scenario, the bid $b_i$ of player $i$ is a function of his valuation $\theta_i$. For example, $b_i(\theta_i) = \alpha_i\theta_i + \beta_i$ is a particular strategy for player $i$.

Figure 2.6 captures the idea behind an indirect mechanism and the Bayesian game that is induced by an indirect mechanism.

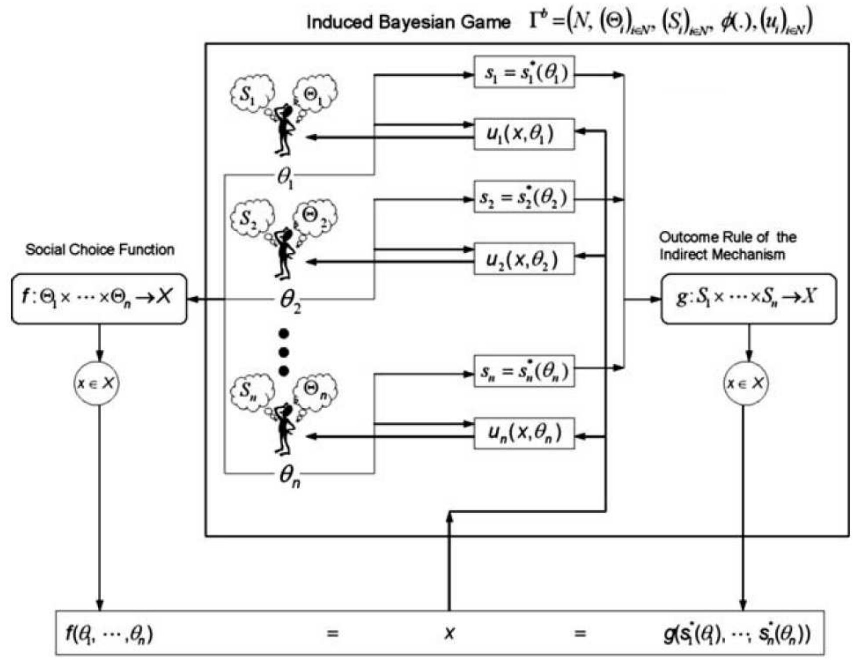



**Fig. 2.6** The idea behind implementation by an indirect mechanism

*Example 2.39 (Bayesian Game Induced by First Price Procurement Auction).* First, note that $N = \{0,1,2\}$. The type sets are $\Theta_0, \Theta_1, \Theta_2$, and the common prior is $\phi \in \Delta(\Theta)$. The set of outcomes is

$$X = \{(y_0, y_1, y_2, t_0, t_1, t_2) : y_i \in \{0,1\}, y_0 + y_1 + y_2 = 1, t_i \in \mathbb{R}, t_0 + t_1 + t_2 \leq 0\}$$

$$u_i((y_0, y_1, y_2, t_0, t_1, t_2), \theta_i) = -\theta_i y_i + t_i \quad i = 1,2$$

$$u_0((y_0, y_1, y_2, t_0, t_1, t_2), \theta_0) = \theta_0 y_0 + t_0$$

$$S_1 = \mathbb{R}_+ \; ; \; S_2 = \mathbb{R}_+$$

$$g(b_0, b_1, b_2) = (y_0(b_0,b_1,b_2), y_1(b_0,b_1,b_2), y_2(b_0,b_1,b_2), \\ t_0(b_0,b_1,b_2), t_1(b_0,b_1,b_2), t_2(b_0,b_1,b_2))$$

such that

$$
\begin{aligned}
y_0(b_0,b_1,b_2) &= 0 \qquad \forall\, b_0,b_1,b_2\\
y_1(b_0,b_1,b_2) &= 1 \qquad \text{if } b_1 \le b_2\\
&= 0 \qquad \text{if } b_1 > b_2\\
y_2(b_0,b_1,b_2) &= 1 \qquad \text{if } b_1 > b_2\\
&= 0 \qquad \text{if } b_1 \le b_2\\
t_1(b_0,b_1,b_2) &= b_1 y_1(b_0,b_1,b_2)\\
t_2(b_0,b_1,b_2) &= b_2 y_2(b_0,b_1,b_2)\\
t_0(b_0,b_1,b_2) &= -(t_1(b_0,b_1,b_2)+t_2(b_0,b_1,b_2)).
\end{aligned}
$$

The game induced by the second price procurement auction will be similar except for appropriate changes in $t_1$ and $t_2$.

### 2.8.4 Implementation of a Social Choice Function by a Mechanism

We now formalize the notion of implementation of a social choice function by a mechanism.

**Definition 2.22 (Implementation of an SCF).** We say that a mechanism $\mathscr{M} = ((S_i)_{i\in N}, g(\cdot))$ implements the social choice function $f(\cdot)$ if there is a pure strategy equilibrium $s^*(\cdot) = (s_1^*(\cdot),\ldots,s_n^*(\cdot))$ of the Bayesian game $\Gamma^b$ induced by $\mathscr{M}$ such that $g\,(s_1^*(\theta_1),\ldots,s_n^*(\theta_n)) = f\,(\theta_1,\ldots,\theta_n)\ \forall\,(\theta_1,\ldots,\theta_n)\in\Theta$.

Figure 2.6 explains the idea behind a mechanism implementing a social choice function. Depending on the nature of the underlying equilibrium, two ways of implementing an SCF $f(\cdot)$ are standard in the literature.

**Definition 2.23 (Implementation in Dominant Strategies).** We say that a mechanism $\mathscr{M} = ((S_i)_{i\in N}, g(\cdot))$ implements the social choice function $f(\cdot)$ in dominant strategy equilibrium if there is a weakly dominant strategy equilibrium $s^*(\cdot) = (s_1^*(\cdot),\ldots,s_n^*(\cdot))$ of the game $\Gamma^b$ induced by $\mathscr{M}$ such that

$$g\,(s_1^*(\theta_1),\ldots,s_n^*(\theta_n)) = f\,(\theta_1,\ldots,\theta_n)\ \forall\,(\theta_1,\ldots,\theta_n)\in\Theta.$$

*Note 2.9.* Since a strongly dominant strategy equilibrium is automatically a weakly dominant strategy equilibrium, the above definition applies to the strongly dominant case also. In the latter case, we could say the implementation is in strongly dominant strategy equilibrium. It is worth recalling that there could exist multiple weakly dominant strategy equilibria whereas a strongly dominant strategy equilibrium, if it exists, is unique.

**Definition 2.24 (Implementation in Bayesian Nash Equilibrium).** We say that a mechanism $\mathscr{M} = ((S_i)_{i\in N}, g(\cdot))$ implements the social choice function $f(\cdot)$ in Bayesian Nash equilibrium if there is a pure strategy Bayesian Nash equilibrium $s^*(\cdot) = (s_1^*(\cdot),\ldots,s_n^*(\cdot))$ of the game $\Gamma^b$ induced by $\mathscr{M}$ such that

$$g\left(s_1^*(\theta_1),\ldots,s_n^*(\theta_n)\right) = f\left(\theta_1,\ldots,\theta_n\right)\ \forall\left(\theta_1,\ldots,\theta_n\right) \in \Theta.$$

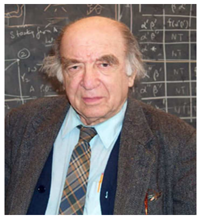
*Note 2.10.* In the definition, what is implicitly implied is a pure strategy Bayesian Nash equilibrium. Such an equilibrium may or may not exist, but we implicitly assume that such an equilibrium exists.

*Note 2.11.* The game $\Gamma^b$ induced by the mechanism $\mathscr{M}$ may have more than one equilibrium, but the above definition requires only that *one of them* induces outcomes in accordance with the SCF $f(\cdot)$. Implicitly, then, the above definition assumes that, if multiple equilibria exist, the agents will play the equilibrium that the mechanism designer (social planner) wants. This is an extremely important problem that is addressed by a theory called *implementation theory*. A brief idea about implementation theory will be provided in Section 2.21.

*Note 2.12.* Another implicit assumption of the above definition is that the game induced by the mechanism is a simultaneous move game, that is all the agents, after learning their types, choose their actions simultaneously.

## 2.9 Incentive Compatibility and the Revelation Theorem

The notion of incentive compatibility is perhaps the most fundamental concept in mechanism design, and the revelation theorem is perhaps the most fundamental result in mechanism design. We have already seen that mechanism design involves the preference revelation (or elicitation) problem and the preference aggregation problem. The preference revelation problem involves eliciting truthful information from the agents about their types. In order to elicit truthful information, there is a need to somehow make truth revelation a best response for the agents, consistent with rationality and intelligence assumptions. Offering incentives is a way of doing this; incentive compatibility essentially refers to offering the right amount of incentive to induce truth revelation by the agents. There are broadly two types of incentive compatibility: (1) Truth revelation is a best response for each agent irrespective of what is reported by the other agents; (2) Truth revelation is a best response for each agent whenever the other agents also reveal their true types. The first one is called dominant strategy incentive compatibility (DSIC), and the second one is called Bayesian Nash incentive compatibility (BIC). Since truth revelation is always with respect to types, only direct revelation mechanisms are relevant when formalizing the notion of incentive compatibility. The notion of incentive compatibility was first introduced by Leonid Hurwicz [10].



**Leonid Hurwicz** , Eric Maskin, and Roger Myerson were jointly awarded the Nobel prize in economic sciences in 2007 for having laid the foundations of mechanism design theory. Hurwicz, born in 1917, has become the oldest winner of the Nobel prize. It was Hurwicz who first introduced the notion of mechanisms with his work in 1960 [11]. He defined a mechanism as a communication system in which participants send messages to each other and perhaps to a *message center* and a prespecified rule assigns an outcome (such as allocation of goods and payments to be made) for every collection of received messages. Hurwicz [10] introduced the key notion of incentive compatibility in 1972. This notion allowed mechanism design to incorporate the incentives of rational players and opened up the area of mechanism design. The notion of incentive compatibility plays a central role in the revelation theorem, which is a fundamental result in mechanism design theory. Hurwicz is also credited with many important possibility and impossibility results in mechanism design. For example, he showed that, in a standard exchange economy, no incentive compatible mechanism that satisfies individual rationality can produce Pareto optimal outcomes. Hurwicz's work in game theory and mechanism design demonstrated, beyond doubt, the value of using analytical methods in modeling economic institutions.

Hurwicz was, until his demise on June 24, 2008, Regents Professor Emeritus in the Department of Economics at the University of Minnesota. He taught there in the areas of welfare economics, public economics, mechanisms and institutions, and mathematical economics.

**Eric Maskin** is a joint winner with Leonid Hurwicz and Roger Myerson, of the Nobel prize in Economic Sciences in 2007. One of his most creative contributions was his work on implementation theory, which addresses the following problem: Given a social goal, can we characterize when we can design a mechanism whose equilibrium outcomes coincide with the outcomes that are desirable according to that goal? Maskin [34] gave a general solution to this problem. He brilliantly showed that if social goals are to be implementable, then they must satisfy a certain kind of *monotonicity* which is now famously called *Maskin Monotonicity*. He also showed that monotonicity guarantees implementation under certain mild conditions (at least three players and no veto power). He has also made major contributions to dynamic games. One of his early contributions was to formalize the Revelation Theorem to the setting of Bayesian incentive compatible mechanisms.

Maskin was born on December 12, 1950, in New York city. He earned an A.B. in Mathematics and a Ph. D. in Applied Mathematics from Harvard University in 1976. He taught at the Massachusetts Institute of Technology during 1977–1984 and at Harvard University during 1985–2000. Since 2000, he is the Albert O. Hirschman Professor of Social Science at the Institute for Advanced Study in Princeton, NJ, USA.

**Roger Bruce Myerson** jointly received the Nobel Prize in Economic Sciences in 2007, with Leonid Hurwicz and Eric Maskin, for having laid the foundations of mechanism design theory. Myerson has straddled several subareas in game theory and mechanism design, and his contributions have left a deep impact in the area. He was instrumental in conceptualizing and proving the revelation theorem in mechanism design for Bayesian implementations in its most generality. His work on optimal auctions in 1981 is a landmark result and has led to a phenomenal amount of further work in the area of optimal auctions. He has also made major contributions in bargaining with incomplete information and cooperative games with incomplete information. His textbook *Game Theory: Analysis of Conflict* is a scholarly and comprehensive reference text that embodies all important results in game theory in a rigorous, yet insightful way. Myerson has also worked on economic analysis of political institutions and written several influential papers in this area including recently on democratization and the Iraq war.

Myerson was born on March 29, 1951. He received his A.B., S.M., and Ph.D., all in Applied Mathematics from Harvard University. He completed his Ph.D. in 1976, working with the legendary Kenneth Arrow. He was a Professor of Economics at the Kellogg School of Management in Northwestern University during 1976-2001. Since 2001, he has been the Glen A. Lloyd Distinguished Service Professor of Economics at the University of Chicago.

### *2.9.1 Incentive Compatibility (IC)*

**Definition 2.25 (Incentive Compatibility).** A social choice function $f : \Theta_1 \times \ldots \times \Theta_n \to X$ is said to be incentive compatible (or truthfully implementable) if the Bayesian game induced by the direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ has a pure strategy equilibrium $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ in which $s_i^*(\theta_i) = \theta_i, \forall \theta_i \in \Theta_i, \forall i \in N$.

That is, truth revelation by each agent constitutes an equilibrium of the game induced by $\mathscr{D}$. It is easy to infer that if an SCF $f(\cdot)$ is incentive compatible then the direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ can implement it. That is, directly asking the agents to report their types and using this information in $f(\cdot)$ to get the social outcome will solve both the problems, namely, preference elicitation and preference aggregation.

Based on the type of equilibrium concept used, two types of incentive compatibility are defined.

**Definition 2.26 (Dominant Strategy Incentive Compatibility (DSIC)).** A social choice function $f : \Theta_1 \times \ldots \times \Theta_n \to X$ is said to be dominant strategy incentive compatible (or truthfully implementable in dominant strategies) if the direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ has a *weakly dominant strategy equilibrium* $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ in which $s_i^*(\theta_i) = \theta_i, \forall \theta_i \in \Theta_i, \forall i \in N$.

That is, truth revelation by each agent constitutes a dominant strategy equilibrium of the game induced by $\mathscr{D}$. Strategy-proof, cheat-proof, straightforward are the alternative phrases used for this property.

*Example 2.40 (Dominant Strategy Incentive Compatibility of Second Price Procurement Auction).* It is easy to see that the social choice function implemented by the second price auction is dominant strategy incentive compatible.

Using the definition of a dominant strategy equilibrium in Bayesian games (Section 2.5), the following necessary and sufficient condition for an SCF $f(\cdot)$ to be dominant strategy incentive compatible can be easily derived:

$$u_i\left(f\left(\theta_i,\theta_{-i}\right),\theta_i\right) \geq u_i(f(\hat{\theta}_i,\theta_{-i}),\theta_i), \forall i \in N, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}, \forall \hat{\theta}_i \in \Theta_i. \quad (2.9)$$

The above condition says that if the SCF $f(.)$ is DSIC, then, irrespective of what the other agents report, it is always a best response for agent $i$ to report his true type $\theta_i$.

**Definition 2.27 (Bayesian Incentive Compatibility (BIC)).** A social choice function $f : \Theta_1 \times \ldots \times \Theta_n \to X$ is said to be Bayesian incentive compatible (or truthfully implementable in Bayesian Nash equilibrium) if the direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i\in N}, f(\cdot))$ has a *Bayesian Nash equilibrium* $s^*(\cdot) = (s_1^*(\cdot),\ldots,s_n^*(\cdot))$ in which $s_i^*(\theta_i) = \theta_i, \forall \theta_i \in \Theta_i, \forall i \in N$.

That is, truth revelation by each agent constitutes a Bayesian Nash equilibrium of the game induced by $\mathscr{D}$.

*Example 2.41 (Bayesian Incentive Compatibility of First Price Procurement Auction).* We have seen that the first price procurement auction for a single indivisible item implements the following social choice function:

$$f(\theta) = (y_0(\theta), y_1(\theta), y_2(\theta), t_0(\theta), t_1(\theta), t_2(\theta))$$

with

$$\begin{aligned}
y_0(\theta) &= 0 \quad \forall \theta \in \Theta \\
y_1(\theta) &= 1 \quad \text{if } \theta_1 \leq \theta_2 \\
&= 0 \quad \text{otherwise} \\
y_2(\theta) &= 1 \quad \text{if } \theta_1 > \theta_2 \\
&= 0 \quad \text{otherwise} \\
t_1(\theta) &= \frac{1+\theta_1}{2} y_1(\theta) \\
t_2(\theta) &= \frac{1+\theta_2}{2} y_2(\theta) \\
t_0(\theta) &= -(t_1(\theta) + t_2(\theta)).
\end{aligned}$$

If seller 1 has type $\theta_1$, then his optimal bid $\hat{\theta}_1$ is obtained by solving

$$\max_{\hat{\theta}_1} \left( \frac{1+\hat{\theta}_1}{2} - \theta_1 \right) P\{\theta_2 \geq \hat{\theta}_1\}.$$

This is the same as

$$\max_{\hat{\theta}_1} \left( \frac{1+\hat{\theta}_1}{2} - \theta_1 \right) (1-\hat{\theta}_1).$$

This yields $\hat{\theta}_1 = \theta_1$. Thus it is optimal for seller 1 to reveal his true private value if seller 2 reveals his true value. The same situation applies to seller 2. This implies that the social choice function is Bayesian Nash incentive compatible (since the equilibrium involved is a Bayesian Nash equilibrium).

Using the definition of a Bayesian Nash equilibrium in Bayesian games (Section 2.5), the following necessary and sufficient condition for an SCF $f(\cdot)$ to be Bayesian incentive compatible can be easily derived:

$$E_{\theta_{-i}}\left[u_i\left(f\left(\theta_i,\theta_{-i}\right),\theta_i\right)|\theta_i\right] \geq E_{\theta_{-i}}[u_i(f(\hat{\theta}_i,\theta_{-i}),\theta_i)|\theta_i], \forall i \in N, \forall \theta_i \in \Theta_i, \forall \hat{\theta}_i \in \Theta_i \quad (2.10)$$

where the expectation is taken over the type profiles of agents other than agent $i$.

*Note 2.13.* If a social choice function $f(\cdot)$ is dominant strategy incentive compatible then it is also Bayesian incentive compatible. The proof of this follows trivially from the fact that a weakly dominant strategy equilibrium is necessarily a Bayesian Nash equilibrium.

## *2.9.2 The Revelation Principle for Dominant Strategy Equilibrium*

The revelation principle basically illustrates the relationship between an indirect mechanism $\mathscr{M}$ and a direct revelation mechanism $\mathscr{D}$ with respect to a given SCF $f(\cdot)$. This result enables us to restrict our inquiry about truthful implementation of an SCF to the class of direct revelation mechanisms only.

**Theorem 2.3.** *Suppose that there exists a mechanism $\mathscr{M} = (S_1,\ldots,S_n,g(\cdot))$ that implements the social choice function $f(\cdot)$ in dominant strategy equilibrium. Then $f(\cdot)$ is dominant strategy incentive compatible.*

**Proof:** If $\mathscr{M} = (S_1,\ldots,S_n,g(\cdot))$ implements $f(\cdot)$ in dominant strategies, then there exists a profile of strategies $s^*(\cdot) = (s_1^*(\cdot),\ldots,s_n^*(\cdot))$ such that

$$g\left(s_1^*(\theta_1),\ldots,s_n^*(\theta_n)\right) = f\left(\theta_1,\ldots,\theta_n\right) \ \forall\left(\theta_1,\ldots,\theta_n\right) \in \Theta \quad (2.11)$$

and

$$u_i(g(s_i^*(\theta_i),s_{-i}(\theta_{-i})),\theta_i) \geq u_i(g(s_i^{'}(\theta_i),s_{-i}(\theta_{-i})),\theta_i)$$
$$\forall i \in N, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}, \forall s_i^{'}(\cdot) \in S_i, \forall s_{-i}(\cdot) \in S_{-i}. \quad (2.12)$$

Condition (2.12) implies, in particular, that

$$u_i(g(s_i^*(\theta_i),s_{-i}^*(\theta_{-i})),\theta_i) \geq u_i(g(s_i^*(\hat{\theta}_i),s_{-i}^*(\theta_{-i})),\theta_i)$$
$$\forall i \in N, \forall \theta_i \in \Theta_i, \forall \hat{\theta}_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}. \quad (2.13)$$

Conditions (2.11) and (2.13) together imply that

$$u_i\left(f\left(\theta_i,\theta_{-i}\right),\theta_i\right) \geq u_i(f(\hat{\theta}_i,\theta_{-i}),\theta_i), \forall i \in N, \forall \theta_i \in \Theta_i, \forall \theta_{-i} \in \Theta_{-i}, \forall \hat{\theta}_i \in \Theta_i.$$

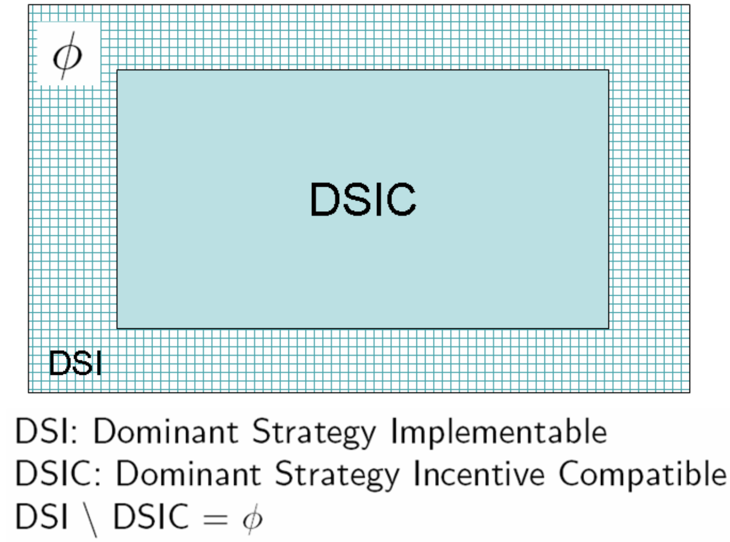



**Fig. 2.7** Revelation principle for dominant strategy equilibrium

But this is precisely condition (2.9), the condition for $f(\cdot)$ to be truthfully implementable in dominant strategies.

*Q.E.D.*

The idea behind the revelation principle can be understood with the help of Figure 2.7. In this picture, **DSI** represents the set of all social choice functions that are implementable in dominant strategies and **DSIC** is the set of all social choice functions that are dominant strategy incentive compatible. The picture depicts the obvious fact that **DSIC** is a subset of **DSI** and illustrates the revelation theorem by showing that the set difference between these two sets is the empty set, thus implying that **DSIC** is precisely the same as **DSI**.

### *2.9.3 The Revelation Principle for Bayesian Nash Equilibrium*

**Theorem 2.4.** *Suppose that there exists a mechanism $\mathscr{M} = (S_1, \ldots, S_n, g(\cdot))$ that implements the social choice function $f(\cdot)$ in Bayesian Nash equilibrium. Then $f(\cdot)$ is truthfully implementable in Bayesian Nash equilibrium (Bayesian incentive compatible).*

**Proof:** If $\mathscr{M} = (S_1, \ldots, S_n, g(\cdot))$ implements $f(\cdot)$ in Bayesian Nash equilibrium, then there exists a profile of strategies $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ such that

$$g\left(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)\right) = f\left(\theta_1, \ldots, \theta_n\right) \; \forall \left(\theta_1, \ldots, \theta_n\right) \in \Theta \tag{2.14}$$

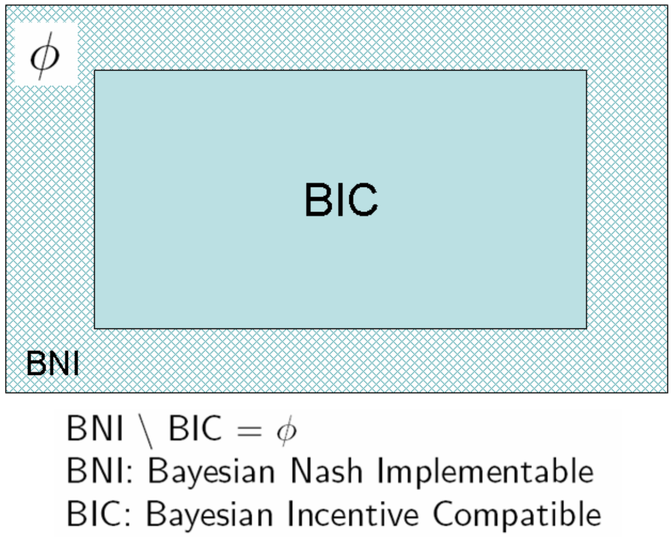



**Fig. 2.8** Revelation principle for Bayesian Nash equilibrium

and

$$E_{\theta_{-i}}\left[u_i(g(s_i^*(\theta_i),s_{-i}^*(\theta_{-i})),\theta_i)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(g(s_i'(\theta_i),s_{-i}^*(\theta_{-i})),\theta_i)|\theta_i\right]$$
$$\forall i \in N, \forall \theta_i \in \Theta_i, \forall s_i'(\cdot) \in S_i. \quad (2.15)$$

Condition (2.15) implies, in particular, that

$$E_{\theta_{-i}}\left[u_i(g(s_i^*(\theta_i),s_{-i}^*(\theta_{-i})),\theta_i)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(g(s_i^*(\hat{\theta}_i),s_{-i}^*(\theta_{-i})),\theta_i)|\theta_i\right]$$
$$\forall i \in N, \forall \theta_i \in \Theta_i, \forall \hat{\theta}_i \in \Theta_i. \quad (2.16)$$

Conditions (2.14) and (2.16) together imply that

$$E_{\theta_{-i}}\left[u_i\left(f\left(\theta_i,\theta_{-i}\right),\theta_i\right)|\theta_i\right] \geq E_{\theta_{-i}}\left[u_i(f(\hat{\theta}_i,\theta_{-i}),\theta_i)|\theta_i\right], \forall i \in N, \forall \theta_i \in \Theta_i, \forall \hat{\theta}_i \in \Theta_i.$$

But this is precisely condition (2.10), the condition for $f(\cdot)$ to be truthfully implementable in Bayesian Nash equilibrium.

*Q.E.D.*

In a way similar to the revelation principle for dominant strategy equilibrium, the revelation principle for Bayesian Nash equilibrium can be explained with the help of Figure 2.8. In this picture, **BNI** represents the set of all social choice functions which are implementable in Bayesian Nash equilibrium and **BIC** is the set of all social choice functions which are Bayesian incentive compatible. The picture depicts the fact that **BIC** is a subset of **BNI** and illustrates the revelation theorem by showing that the set difference between these two sets is the empty set, thus implying that **BIC** is precisely the same as **BNI**.

Figure 2.9 provides a combined view of both the revelation theorems that we have seen in this section.

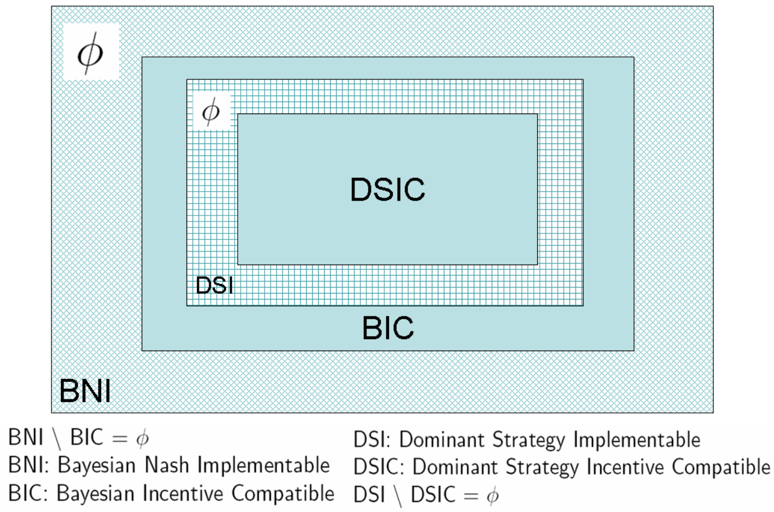



**Fig. 2.9** Combined view of revelation theorems for dominant strategy equilibrium and Bayesian Nash equilibrium

## 2.10 Properties of Social Choice Functions

We have seen that a mechanism provides a solution to both the preference elicitation problem and preference aggregation problem, if the mechanism can implement the desired social choice function $f(\cdot)$. It is obvious that some SCFs are implementable and some are not. Before we look into the question of characterizing the space of implementable social choice functions, it is important to know which social choice function ideally a social planner would wish to implement. In this section, we highlight a few properties of an SCF that ideally a social planner would wish the SCF to have.

### 2.10.1 Ex-Post Efficiency

**Definition 2.28 (Ex-Post Efficiency).** The SCF $f : \Theta \to X$ is said to be ex-post efficient (or Paretian) if for every profile of agents' types, $\theta \in \Theta$, the outcome $f(\theta)$ is a Pareto optimal outcome. The outcome $f(\theta_1, \ldots, \theta_n)$ is Pareto optimal if there does not exist any $x \in X$ such that:

$$u_i(x, \theta_i) \geq u_i(f(\theta), \theta_i) \;\forall\; i \in N \text{ and } u_i(x, \theta_i) > u_i(f(\theta), \theta_i) \text{ for some } i \in N.$$

*Example 2.42 (Supplier Selection Problem).* Consider the supplier selection problem (Example 2.29). Let the social choice function $f$ be given by

$$f(a_1, a_2) = x$$
$$f(a_1, b_2) = x.$$

The outcome $f(a_1, a_2) = x$ is Pareto optimal since the other outcomes $y$ and $z$ are such that

$$u_1(y, a_1) < u_1(x, a_1)$$
$$u_1(z, a_1) < u_1(x, a_1).$$

The outcome $f(a_1, b_2) = x$ is Pareto optimal since the other outcomes $y$ and $z$ are such that

$$u_1(y, a_1) < u_1(x, a_1)$$
$$u_1(z, a_1) < u_1(x, a_1).$$

Thus SCF 1 is ex-post efficient.

*Example 2.43 (Procurement of a Single Indivisible Item).* We have looked at three social choice functions, SCF-PROC1, SCF-PROC2, SCF-PROC3, in the previous section. One can show that all these SCFs are ex-post efficient.

### 2.10.2 Dictatorship in SCFs

We define this through a dictatorial social choice function.

**Definition 2.29 (Dictatorship).** A social choice function $f : \Theta \to X$ is said to be dictatorial if there exists an agent $d$ (called dictator) who satisfies the following property:

$$\forall\; \theta \in \Theta,\; f(\theta) \text{ is such that } u_d(f(\theta), \theta_d) \geq u_d(x, \theta_d) \;\forall\; x \in X.$$

A social choice function that is not dictatorial is said to be nondictatorial .

In a dictatorial SCF, every outcome that is picked by the SCF is such that it is a most favored outcome for the dictator.

*Example 2.44 (Supplier Selection Problem).* Let the social choice function $f$ be given by

$$f(a_1, a_2) = x; \;\; f(a_1, b_2) = x.$$

It is easy to see that agent 1 is a dictator and hence this is a dictatorial SCF. On the other hand, consider the following SCF:

$$f(a_1, a_2) = x; \;\; f(a_1, b_2) = y.$$

One can verify that this is not a dictatorial SCF.

## 2.10.3 Individual Rationality

Individual rationality is also often referred to as voluntary participation property. Individual rationality of a social choice function essentially means that each agent gains a nonnegative utility by participating in a mechanism that implements the social choice function. There are three stages at which individual rationality constraints (also called participation constraints) may be relevant in a mechanism design situation.

### 2.10.3.1 Ex-Post Individual Rationality

These constraints become relevant when any agent $i$ is given a choice to withdraw from the mechanism at the ex-post stage, that is, after all the agents have announced their types and an outcome in $X$ has been chosen. Let $\overline{u_i}(\theta_i)$ be the utility that agent $i$ receives by withdrawing from the mechanism when his type is $\theta_i$. Then, to ensure agent $i$'s participation, we must satisfy the following *ex-post participation (or individual rationality) constraints*

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq \overline{u_i}(\theta_i) \;\; \forall \, (\theta_i, \theta_{-i}) \in \Theta.$$

### 2.10.3.2 Interim Individual Rationality

Let the agent $i$ be allowed to withdraw from the mechanism only at an interim stage that arises after the agents have learned their type but before they have chosen their actions in the mechanism. In such a situation, the agent $i$ will participate in the mechanism only if his interim expected utility $U_i(\theta_i|f) = E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i]$ from social choice function $f(\cdot)$, when his type is $\theta_i$, is greater than $\overline{u_i}(\theta_i)$. Thus, *interim participation (or individual rationality) constraints* for agent $i$ require that

$$U_i(\theta_i|f) = E_{\theta_{-i}}[u_i(f(\theta_i,\theta_{-i}),\theta_i)|\theta_i] \geq \overline{u_i}(\theta_i) \;\; \forall\, \theta_i \in \Theta_i.$$

#### 2.10.3.3 Ex-Ante Individual Rationality

Let agent $i$ be allowed to refuse to participate in a mechanism only at ex-ante stage, that is, before the agents learn their type. In such a situation, the agent $i$ will participate in the mechanism only if his ex-ante expected utility $U_i(f) = E_\theta[u_i(f(\theta_i,\theta_{-i}),\theta_i)]$ from social choice function $f(\cdot)$ is at least $E_{\theta_i}[\overline{u_i}(\theta_i)]$. Thus, *ex-ante participation (or individual rationality) constraints* for agent $i$ require that

$$U_i(f) = E_\theta[u_i(f(\theta_i,\theta_{-i}),\theta_i)] \geq E_{\theta_i}[\overline{u_i}(\theta_i)].$$

The following proposition establishes a relationship among the three different participation constraints discussed above. The proof is left as an exercise.

**Proposition 2.2.** *For any social choice function $f(\cdot)$, we have*

$$f(\cdot) \text{ is ex-post IR} \Rightarrow f(\cdot) \text{ is interim IR} \Rightarrow f(\cdot) \text{ is ex-ante IR}.$$

### *2.10.4 Efficiency*

We have seen the notion of ex-post efficiency already. Depending on the epoch at which we look into the game, we have three notions of efficiency, on the lines of individual rationality. These notions were introduced by Holmstorm and Myerson [12]. Let $F$ be any collection of social choice functions that are of interest.

**Definition 2.30 (Ex-Ante Efficiency).** For any given set of social choice functions $F$, and any member $f(\cdot) \in F$, we say that $f(\cdot)$ is ex-ante efficient in $F$ if there is no other $\hat{f}(\cdot) \in F$ having the following two properties:

$$E_\theta[u_i(\hat{f}(\theta),\theta_i)] \geq E_\theta[u_i(f(\theta),\theta_i)] \;\; \forall\, i = 1,\ldots,n,$$
$$E_\theta[u_i(\hat{f}(\theta),\theta_i)] > E_\theta[u_i(f(\theta),\theta_i)] \text{ for some } i.$$

**Definition 2.31 (Interim Efficiency).** For any given set of social choice functions $F$, and any member $f(\cdot) \in F$, we say that $f(\cdot)$ is interim efficient in $F$ if there is no other $\hat{f}(\cdot) \in F$ having the following two properties:

$$E_{\theta_{-i}}[u_i(\hat{f}(\theta),\theta_i)|\theta_i] \geq E_{\theta_{-i}}[u_i(f(\theta),\theta_i)|\theta_i] \;\; \forall\, i = 1,\ldots,n,\ \forall\, \theta_i \in \Theta_i,$$
$$E_{\theta_{-i}}[u_i(\hat{f}(\theta),\theta_i)|\theta_i] > E_{\theta_{-i}}[u_i(f(\theta),\theta_i)|\theta_i] \text{ for some } i \text{ and some } \theta_i \in \Theta_i.$$

**Definition 2.32 (Ex-Post Efficiency).** For any given set of social choice functions $F$, and any member $f(\cdot) \in F$, we say that $f(\cdot)$ is ex-post efficient in $F$ if there is no other $\hat{f}(\cdot) \in F$ having the following two properties:

$$u_i(\hat{f}(\theta), \theta_i) \geq u_i(f(\theta), \theta_i) \;\; \forall\, i = 1, \ldots, n, \; \forall\, \theta \in \Theta,$$
$$u_i(\hat{f}(\theta), \theta_i) > u_i(f(\theta), \theta_i) \text{ for some } i \text{ and some } \theta \in \Theta.$$

Using the above definition of ex-post efficiency, we can say that a social choice function $f(\cdot)$ is ex-post efficient in the sense of Definition 2.28 if and only if it is ex-post efficient in the sense of Definition 2.32 when we take $F = \{f : f$ is a mapping from $\Theta$ to $X\}$.

The following proposition establishes a relationship among these three different notions of efficiency.

**Proposition 2.3.** *Given any set of feasible social choice functions $F$ and $f(\cdot) \in F$, we have*

$$f(\cdot) \textit{ is ex-ante efficient} \Rightarrow f(\cdot) \textit{ is interim efficient} \Rightarrow f(\cdot) \textit{ is ex-post efficient}.$$

For a proof of the above proposition, refer to Proposition 23.F.1 of [6]. Also, compare the above proposition with the Proposition 2.2.

## 2.11 The Gibbard–Satterthwaite Impossibility Theorem

We have seen in the last section that dominant strategy incentive compatibility is an extremely desirable property of social choice functions. However the DSIC property, being a strong one, precludes certain other desirable properties to be satisfied. In this section, we discuss the Gibbard–Satterthwaite impossibility theorem (G–S theorem, for short), which shows that the DSIC property will force an SCF to be dictatorial if the utility environment is an unrestricted one. In fact, in the process, even ex-post efficiency will have to be sacrificed. One can say that the G–S theorem has shaped the course of research in mechanism design during the 1970s and beyond, and is therefore a landmark result in mechanism design theory. The G–S theorem is credited independently to Gibbard in 1973 [13] and Satterthwaite in 1975 [14]. The G–S theorem is a brilliant reinterpretation of the famous Arrow's impossibility theorem (which we discuss in the next section). We start our discussion of the G–S theorem with a motivating example.

**Allan Gibbard** is currently Richard B. Brandt Distinguished University Professor of Philosophy at the University of Michigan. His classic paper *Manipulation of Voting Schemes: A General Result* published in Econometrica (Volume 41, Number 4) in 1973 presents the famous Gibbard Satterthwaite theorem, which was also independently proposed by Mark Satterthwaite. Professor Gibbard's current research interests are in ethical theory. He is the author of two widely popular books: *Thinking How to Live* (2003 - Harvard University Press) and *Wise Choices, Apt Feelings* (1990 - Harvard University Press and Oxford University Press).

**Mark Satterthwaite** is currently A.C. Buehler Professor in Hospital and Health Services Management and Professor of Strategic Management and Managerial Economics at the Kellogg School of Management, Northwestern university. He is a microeconomic theorist with keen interest in how health care markets work. His paper *Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions* published in the Journal of Economic Theory (Volume 10, April 1975) presented a brilliant reinterpretation of Arrow's impossibility theorem, which is now famously known as the Gibbard–Satterthwaite Theorem. He has authored a large number of scholarly papers in the areas of dynamic matching in markets, organizational dynamics, and mechanism design.

*Example 2.45 (Supplier Selection Problem).* We have seen this example earlier (Example 2.29). We have $N = \{1,2\}$, $X = \{x,y,z\}$, $\Theta_1 = \{a_1\}$, and $\Theta_2 = \{a_2,b_2\}$. Consider the following utility functions (note that these are different from the ones considered in Example 2.29):

$$u_1(x,a_1) = 100;\ u_1(y,a_1) = 50;\ u_1(z,a_1) = 0$$
$$u_2(x,a_2) = 0;\ u_2(y,a_2) = 50;\ u_2(z,a_2) = 100$$
$$u_2(x,b_2) = 30;\ u_2(y,b_2) = 60;\ u_2(z,b_2) = 20.$$

We observe for this example that the DSIC and BIC notions are identical since the type of player 1 is common knowledge and hence player 1 always reports the true type (since the type set is a singleton). Consider the social choice function $f$ given by $f(a_1,a_2) = x;\ f(a_1,b_2) = x$. It can be seen that this SCF is ex-post efficient.

To investigate DSIC, suppose the type of player 2 is $a_2$. If player 2 reports his true type, then the outcome is $x$. If he misreports his type as $b_2$, then also the outcome is $x$. Hence there is no incentive for player 2 to misreport. A similar situation presents itself when the type of player 2 is $b_2$. Thus $f$ is DSIC.

In both the type profiles, the outcome happens to be the most favorable one for player 1, that is, $x$. Therefore, player 1 is a dictator and $f$ is dictatorial. Thus the above function is ex-post efficient and DSIC but dictatorial.

Now, let us consider a different SCF $h$ defined by $h(a_1,a_2) = y; h(a_1,b_2) = x$. Following similar arguments as above, $h$ can be shown to be ex-post efficient and nondictatorial but not DSIC. Table 2.11 lists all the nine possible social choice functions in this scenario and the combination of properties each function satisfies.

Note that the situation is quite desirable with the following SCFs.

$$f_5(a_1,a_2) = y;\ f_5(a_1,b_2) = y$$
$$f_7(a_1,a_2) = z;\ f_7(a_1,b_2) = x.$$

The reason is these functions are ex-post efficient, DSIC, and also nondictatorial. Unfortunately however, such desirable situations do not occur in general. In the present case, the desirable situations do occur because of certain reasons that will

| $i$ | $f_i(a_1, a_2)$ | $f_i(a_1, b_2)$ | EPE | DSIC | NON-DICT |
|---|---|---|---|---|---|
| 1 | $x$ | $x$ | $\surd$ | $\surd$ | $\times$ |
| 2 | $x$ | $y$ | $\surd$ | $\times$ | $\surd$ |
| 3 | $x$ | $z$ | $\times$ | $\times$ | $\surd$ |
| 4 | $y$ | $x$ | $\surd$ | $\times$ | $\surd$ |
| 5 | $y$ | $y$ | $\surd$ | $\surd$ | $\surd$ |
| 6 | $y$ | $z$ | $\times$ | $\times$ | $\surd$ |
| 7 | $z$ | $x$ | $\surd$ | $\surd$ | $\surd$ |
| 8 | $z$ | $y$ | $\surd$ | $\surd$ | $\times$ |
| 9 | $z$ | $z$ | $\times$ | $\surd$ | $\surd$ |

**Table 2.11** Social choice functions and properties satisfied by them

become clear soon. In a general setting, ex-post efficiency, DSIC, and nondictatorial properties can never be satisfied simultaneously. In fact, even DSIC and nondictatorial properties cannot coexist. This is the implication of the powerful Gibbard–Satterthwaite theorem.

### *2.11.1 The G–S Theorem*

We will build up some notation before presenting the theorem. We have already seen that the preference of an agent $i$, over the outcome set $X$, when its type is $\theta_i$ can be described by means of a *utility function* $u_i(\cdot, \theta_i) : X \to \mathbb{R}$, which assigns a real number to each element in $X$. A utility function $u_i(\cdot, \theta_i)$ always induces a *unique* preference relation $\succsim$ on $X$ which can be described in the following manner

$$x \succsim y \Leftrightarrow u_i(x, \theta_i) \geq u_i(y, \theta_i).$$

The above preference relation is often called a rational preference relation and it is formally defined as follows.

**Definition 2.33 (Rational Preference Relation).** We say that a relation $\succsim$ on the set $X$ is called a rational preference relation if it possesses the following three properties:

1. Reflexivity: $\forall\ x \in X$, we have $x \succsim x$.
2. Completeness: $\forall\ x, y \in X$, we have that $x \succsim y$ or $y \succsim x$ (or both).
3. Transitivity: $\forall\ x, y, z \in X$, if $x \succsim y$ and $y \succsim z$, then $x \succsim z$.

The following proposition establishes the relationship between these two ways of expressing the preferences of an agent $i$ over the set $X$.

**Proposition 2.4.**

1. *If a preference relation $\succsim$ on X is induced by some utility function $u_i(\cdot, \theta_i)$, then it will be a rational preference relation.*

2. *For every preference relation $\succsim$ on $X$, there may not exist a utility function that induces it. However, when the set $X$ is finite, given any preference relation, there will exist a utility function that induces it.*
3. *For a given preference relation $\succsim$ on $X$, there might be several utility functions that induce it. Indeed, if the utility function $u_i(\cdot,\theta_i)$ induces $\succsim$, then $u_i'(x,\theta_i) = f(u_i(x,\theta_i))$ is another utility function that will also induce $\succsim$, where $f : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function.*

### Strict Total Preference Relations

We now define a special class of rational preference relations that satisfy the antisymmetry property also.

**Definition 2.34 (Strict-total Preference Relation).** We say that a rational preference relation $\succsim$ is strict-total if it possesses the antisymmetry property, in addition to reflexivity, completeness, and transitivity. By antisymmetry, we mean that, for any $x, y \in X$ such that $x \neq y$, we have either $x \succsim y$ or $y \succsim x$, but not both.

The strict-total preference relation is also known as a *linear order relation* because it satisfies the properties of the usual *greater than or equal to* relationship on the real line. Let us denote the set of all rational preference relations and strict-total preference relations on the set $X$ by $\mathscr{R}$ and $\mathscr{P}$, respectively. It is easy to see that $\mathscr{P} \subset \mathscr{R}$.

### Ordinal Preference Relations

In a mechanism design problem, for agent $i$, the preference over the set $X$ is described in the form of a utility function $u_i : X \times \Theta_i \to \mathbb{R}$. That is, for every possible type $\theta_i \in \Theta_i$ of agent $i$, we can define a utility function $u_i(\cdot,\theta_i)$ over the set $X$. Let this utility function induce a rational preference relation $\succsim_i(\theta_i)$ over $X$. The set $\mathscr{R}_i = \{\succsim \, : \, \succsim = \succsim_i(\theta_i) \text{ for some } \theta_i \in \Theta_i\}$ is known as the set of ordinal preference relations for agent $i$. It is easy to see that $\mathscr{R}_i \subset \mathscr{R} \ \ \forall\, i \in N$.

With all the above notions in place, we are now in a position to state the G–S theorem.

**Theorem 2.5 (Gibbard–Satterthwaite Impossibility Theorem).** *Consider a social choice function $f : \Theta \to X$. Suppose that*

1. *The outcome set $X$ is finite and contains at least three elements,*
2. *$\mathscr{R}_i = \mathscr{P} \ \ \forall\, i \in N$,*
3. *$f(\cdot)$ is an onto mapping, that is, the image of SCF $f(\cdot)$ is the set $X$.*

*Then the social choice function $f(\cdot)$ is dominant strategy incentive compatible iff it is dictatorial.*

For a proof of this theorem, the reader is referred to Proposition 23.C.3 of the book by Mas-Colell, Whinston, and Green [6]. We only provide a brief outline of the proof. To prove the necessity, we assume that the social choice function $f(\cdot)$ is dictatorial and it is shown that $f(\cdot)$ is DSIC. This can be shown in a fairly straightforward way. The proof of the sufficiency part of the theorem starts with the assumption that $f(\cdot)$ is DSIC and proceeds in three steps:

1. It is shown using the second condition of the theorem ($\mathscr{R}_i = \mathscr{P} \quad \forall \; i \in N$) that $f(\cdot)$ is monotonic.
2. Next using conditions (2) and (3) of the theorem, it is shown that monotonicity implies ex-post efficiency.
3. Finally, it is shown that a SCF $f(\cdot)$ that is monotonic and ex-post efficient is necessarily dictatorial.

Figure 2.10 shows a pictorial representation of the G–S theorem. The figure depicts two classes $F_1$ and $F_2$ of social choice functions. The class $F_1$ is the set of all SCFs that satisfy conditions (1) and (2) of the theorem while the class $F_2$ is the set of all SCFs that satisfy conditions (1) and (3) of the theorem. The class $GS$ is the set of all SCFs in the intersection of $F_1$ and $F_2$ which are DSIC. The functions in the class GS have to be necessarily dictatorial.

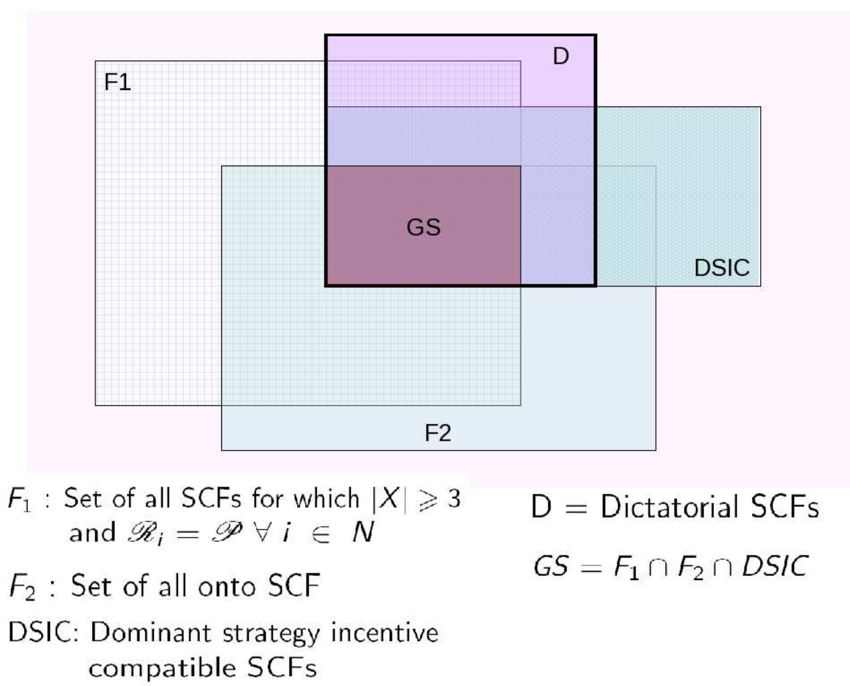



**Fig. 2.10** An illustration of the Gibbard–Satterthwaite Theorem

### *2.11.2 Implications of the G–S Theorem*

One way to get around the impossible situation described by the G–S Theorem is to hope that at least one of the conditions (1), (2), and (3) of the theorem does not hold. We discuss each one of these below.

- Condition (1) asserts that $|X| \geq 3$. This condition is violated only if $|X| = 1$ or $|X| = 2$. The case $|X| = 1$ corresponds to a trivial situation and is not of interest. The case $|X| = 2$ is more interesting but is of only limited interest. A public project problem where only a go or no-go decision is involved and no payments by agents are involved corresponds to this situation.
- Condition (2) asserts that $\mathscr{R}_i = \mathscr{P}\ \forall\ i \in N$. This means that the preferences of each agent cover the entire space of strict total preference relations on $X$. That is, each agent has an extremely rich set of preferences. If we are able to somehow restrict the preferences, we can hope to violate this condition. One can immediately note that this condition was violated in the motivating example (Example 2.45, the supplier selection problem). The celebrated class of VCG mechanisms has been derived by restricting the preferences to the quasilinear domain. This will be discussed in good detail in a later section.
- Condition (3) asserts that $f$ is an onto function. Note that this condition also was violated in Example 2.45. This provides one more route for getting around the G–S Theorem.

Another way of escaping from the jaws of the G–S Theorem is to settle for a weaker form of incentive compatibility than DSIC. We have already discussed Bayesian incentive compatibility (BIC) which only guarantees that reporting true types is a best response for each agent whenever all other agents also report their true types. Following this route leads us to Bayesian incentive compatible mechanisms. These are discussed in good detail in a future section.

The G–S Theorem is an influential result that defined the course of mechanism design research in the 1970s and 1980s. As already stated, the theorem happens to be an ingenious reinterpretation, in the context of mechanism design, of the celebrated Arrow's impossibility theorem, which is discussed next.

## 2.12 Arrow's Impossibility Theorem

This famous impossibility theorem is due to Kenneth Arrow (1951), Nobel laureate in Economic Sciences in 1972. This result has shaped the discipline of social choice theory in many significant ways.

**Kenneth Joseph Arrow** received the Nobel Prize in Economic Sciences in 1972, jointly with John R. Hicks, for their pioneering contributions to general economic equilibrium theory and welfare theory. Arrow is regarded as one of the most influential economists of all time. With his path-breaking contributions in social choice theory, general equilibrium analysis, endogenous growth theory, and economics of information behind him, Kenneth Arrow is truly a legend of economics. Three of his doctoral students, John Harsanyi, Michael Spencer, and Roger Myerson are also Economics Nobel laureates. The famous Arrow impossibility theorem was one of the outstanding results included in his classic book in 1951 *Social Choice and Individual Values*, which itself was inspired by his doctoral work. This theorem is perhaps the most important result in welfare economics and also has far-reaching ramifications for mechanism design theory (in fact, the Gibbard–Satterthwaite theorem is an ingenious reinterpretation of the Arrow Impossibility Theorem).

Kenneth Arrow was born in the New York City on August 23, 1921. He earned his doctorate from Columbia University in 1951, working with Professor Harold Hotelling. He is a recipient of the von Neumann Theory Prize in 1986, and he was awarded in 2004 the National Medal of Science, the highest scientific honor in the United States. His joint work with Gerard Debreu on general equilibrium theory is also a major landmark that was prominently noted in the Nobel Prize awarded to Gerard Debreu in 1983. Arrow is currently the Joan Kenney Professor of Economics and Professor of Operations Research, Emeritus, at Stanford University.

Before discussing this result, we first set up some relevant notation. Consider a set of agents $N = \{1,2,\ldots,n\}$ and a set of outcomes $X$. Let $\succsim_i$ be a rational preference relation of agent $i$ ($i \in N$). Subscript $i$ in $\succsim_i$ indicates that the relation corresponds to agent $i$. For example, $\succsim_i$ could be induced by $u_i(.,\theta_i)$ where $\theta_i$ is a certain type of agent $i$. Each agent is thus naturally associated with a set $\mathscr{R}_i$ of rational preference relations derived from the utility functions $u_i(.,\theta_i)$ where $\theta_i \in \Theta_i$.

Given a rational preference relation $\succsim_i$, let us denote by $\succ_i$ the relation defined by

$$(x,y) \in \succ_i \text{ iff } (x,y) \in \succsim_i \text{ and } (y,x) \notin \succsim_i .$$

The relation $\succ_i$ is said to be the *strict total preference relation* derived from $\succsim_i$. Note that $\succ_i = \succsim_i$ if $\succsim_i$ itself is a strict total preference relation. Given an outcome set $X$, a strict total preference relation can be simply represented as an ordered tuple of elements of $X$. Given $\succsim_i$, let us denote by $\sim_i$ the relation defined by

$$(x,y) \in \sim_i \text{ iff } (x,y) \in \succsim_i \text{ and } (y,x) \in \succsim_i .$$

The relation $\sim_i$ is said to be the *indifference relation* derived from $\succsim_i$.

As usual $\mathscr{R}$ and $\mathscr{P}$ denote, respectively, the set of all rational preference relations and strict total preference relations on the set $X$. Let $\mathscr{A}$ be any nonempty subset of $\mathscr{R}^n$. We define a social welfare functional as a mapping from $\mathscr{A}$ to $\mathscr{R}$.

**Definition 2.35 (Social Welfare Functional).** Given a set of agents $N = \{1,2,\ldots,n\}$, an outcome set $X$, and a set of profiles $\mathscr{A}$ of rational preference relations of the agents, $\mathscr{A} \subset \mathscr{R}^n$, a social welfare functional is a mapping $W : \mathscr{A} \longrightarrow \mathscr{R}$.

Note that a social welfare functional $W$ assigns a rational preference relation $W(\succsim_1, \ldots, \succsim_n)$ to a given profile of rational preference relations $(\succsim_1, \ldots, \succsim_n) \in \mathscr{A}$.

*Example 2.46 (Social Welfare Functional).* Consider the example of the supplier selection problem discussed in Example 2.45, where $N = \{1,2\}$, $X = \{x,y,z\}$, $\Theta_1 = \{a_1\}$, and $\Theta_2 = \{a_2, b_2\}$. Recall the utility functions:

$$u_1(x,a_1) = 100;\ u_1(y,a_1) = 50;\ u_1(z,a_1) = 0$$
$$u_2(x,a_2) = 0;\ u_2(y,a_2) = 50;\ u_2(z,a_2) = 100$$
$$u_2(x,b_2) = 30;\ u_2(y,b_2) = 60;\ u_2(z,b_2) = 20.$$

The utility function $u_1$ leads to the following strict preference relation:

$$\succsim_{a_1} = (x,y,z).$$

The utility function $u_2$ leads to the strict total preference relations:

$$\succsim_{a_2} = (z,y,x);\ \succsim_{b_2} = (y,x,z).$$

Let the set $\mathscr{A}$ be defined as

$$\mathscr{A} = \{(\succsim_{a_1}, \succsim_{a_2}), (\succsim_{a_1}, \succsim_{b_2})\}.$$

An example of a social welfare functional here would be the mapping $W_1$ given by

$$W_1(\succsim_{a_1}, \succsim_{a_2}) = (x,y,z);\quad W_1(\succsim_{a_1}, \succsim_{b_2}) = (y,x,z).$$

Another example would be the mapping $W_2$ given by

$$W_2(\succsim_{a_1}, \succsim_{a_2}) = (x,y,z);\quad W_2(\succsim_{a_1}, \succsim_{b_2}) = (z,y,x).$$

Note the difference between a social choice function and a social welfare functional. In the case of a social choice function, the preferences are summarized in terms of types and each type profile is mapped to a social outcome. On the other hand, a social welfare functional maps a profile of individual preferences to a social preference relation. Recall that the type of an agent determines a preference relation on the set $X$ through the utility function.

We now define three properties of a social welfare functional: *unanimity* (also called *Paretian property*); *pairwise independence* (also called *independence of irrelevant alternatives* (IIA)), and *dictatorship*.

**Definition 2.36 (Unanimity).** A social welfare functional $W : \mathscr{A} \longrightarrow \mathscr{R}$ is said to be unanimous if $\forall\ (\succsim_1, \ldots, \succsim_n) \in \mathscr{A}$ and $\forall x,y \in X$,

$$(x,y) \in\ \succsim_i\ \forall i \in N \Longrightarrow (x,y) \in W_p(\succsim_1 \ldots, \succsim_n)$$

where $W_p(\succsim_1 \ldots, \succsim_n)$ is the strict preference relation derived from $W(\succsim_1 \ldots, \succsim_n)$.

The above definition means that, for all pairs $x, y \in X$, whenever $x$ is preferred to $y$ for every agent, then $x$ is also socially preferred to $y$.

*Example 2.47 (Unanimity).* For the problem being discussed, let

$$W_1(\succsim_{a_1}, \succsim_{a_2}) = W_1((x,y,z),(z,y,x)) = (x,y,z)$$

$$W_1(\succsim_{a_1}, \succsim_{b_2}) = W_1((x,y,z),(y,x,z)) = (y,x,z).$$

This is unanimous because

- $(y,z) \in \succsim_{a_1}, (y,z) \in \succsim_{b_2}$, and $(y,z) \in W_1(\succsim_{a_1}, \succsim_{b_2})$; and
- $(x,z) \in \succsim_{a_1}, (x,z) \in \succsim_{b_2}$, and $(x,z) \in W_1(\succsim_{a_1}, \succsim_{b_2})$.

On the other hand, let

$$W_2((x,y,z),(z,y,x)) = (x,y,z); \;\; W_2((x,y,z),(y,x,z)) = (z,y,x)$$

Here $(y,z) \in \succsim_{a_1}$ and $(y,z) \in \succsim_{b_2}$ but $(y,z) \notin W_2(\succsim_{a_1}, \succsim_{b_2})$. So $W_2$ is not unanimous.

**Definition 2.37 (Pairwise Independence).** The social welfare functional $W : \mathscr{A} \longrightarrow \mathscr{R}$ is said to satisfy pairwise independence if $\forall x, y \in X$, the social preference between $x$ and $y$ will depend only on the individual preferences between $x$ and $y$. That is, $\forall x, y \in X$, $\forall(\succsim_1 \ldots, \succsim_n) \in \mathscr{A}$, $\forall(\succsim_1' \ldots, \succsim_n') \in \mathscr{A}$, with the property that

$$(x,y) \in \succsim_i \;\Leftrightarrow (x,y) \in \succsim_i' \;\text{ and }\; (y,x) \in \succsim_i \Leftrightarrow (y,x) \in \succsim_i' \;\; \forall i \in N,$$

we have that

$$(x,y) \in W(\succsim_1, \ldots, \succsim_n) \Leftrightarrow (x,y) \in W(\succsim_1', \ldots, \succsim_n'); \;\text{ and}$$

$$(y,x) \in W(\succsim_1, \ldots, \succsim_n) \Leftrightarrow (y,x) \in W(\succsim_1', \ldots, \succsim_n').$$

*Example 2.48 (Pairwise Independence).* Consider the example as before and let

$$W_3(\succsim_{a_1}, \succsim_{a_2}) = W_3((x,y,z),(z,y,x)) = (x,y,z)$$

$$W_3(\succsim_{a_1}, \succsim_{b_2}) = W_3((x,y,z),(y,x,z)) = (y,z,x).$$

Here agent 1 prefers $x$ to $y$ in both the profiles while agent 2 prefers $y$ to $x$ in both the profiles. However in the first case, $x$ is socially preferred to $y$ while in the second case $y$ is socially preferred to $x$. Thus the social preference between $x$ and $y$ is not exclusively dependent on the individual preferences between $x$ and $y$. This shows that $W_1$ is not pairwise independent. On the other hand, consider $W_3$ given by

$$W_4((x,y,z),(z,y,x)) = (x,y,z)$$

$$W_4((x,y,z),(y,x,z)) = (z,x,y).$$

Now this social welfare functional satisfies pairwise independence.

The pairwise independence property is a very appealing property since it ensures that the social ranking between any pair of alternatives $x$ and $y$ does not in any way depend on other alternatives or the relative positions of these other alternatives in the individual preferences. Secondly, the pairwise independence property has a close connection to a property called the weak preference reversal property, which is quite crucial for ensuring dominant strategy incentive compatibility of social choice functions. Further, this property leads to a nice decomposition of the problem of social ranking. For instance, if we wish to determine a social ranking on the outcomes of a subset $Y$ of $X$, we do not need to worry about individual preferences on the set $X \backslash Y$.

**Definition 2.38 (Dictatorship).** A social welfare functional $W : \mathscr{A} \longrightarrow \mathscr{R}$ is called a dictatorship if there exists an agent, $d \in N$, called the dictator such that $\forall x, y \in X$ and $\forall (\succsim_1, \ldots, \succsim_n) \in \mathscr{A}$, we have

$$(x,y) \in \succsim_d \Rightarrow (x,y) \in W_p(\succsim_1, \ldots, \succsim_n).$$

This means that whenever the dictator prefers $x$ to $y$, then $x$ is also socially preferred to $y$, irrespective of the preferences of the other agents. A social welfare functional that does not have a dictator is said to be nondictatorial.

*Example 2.49 (Dictatorship).* Consider the social welfare functional

$$W_5((x,y,z),(z,y,x)) = (x,y,z)$$

$$W_5((x,y,z),(y,x,z)) = (x,y,z).$$

It is clear that agent 1 is a dictator here. On the other hand, the social welfare functional

$$W_3((x,y,z),(z,y,x)) = (x,y,z)$$

$$W_3((x,y,z),(y,x,z)) = (y,z,x)$$

is not dictatorial.

Ideally, a social planner would like to implement a social welfare functional that is unanimous, satisfies the pairwise independence property, and is nondictatorial. Unfortunately, this belongs to the realm of impossible situations when the preference profiles of the agents are *rich*. This is the essence of the Arrow's Impossibility Theorem, which is stated next.

**Theorem 2.6 (Arrow's Impossibility Theorem).** *Suppose*

1. $|X| \geq 3$,
2. $\mathscr{A} = \mathscr{R}^n$ or $\mathscr{A} = \mathscr{P}^n$.

*Then every social welfare functional* $W : \mathscr{A} \longrightarrow \mathscr{R}$ *that is unanimous and satisfies pairwise independence is dictatorial.*

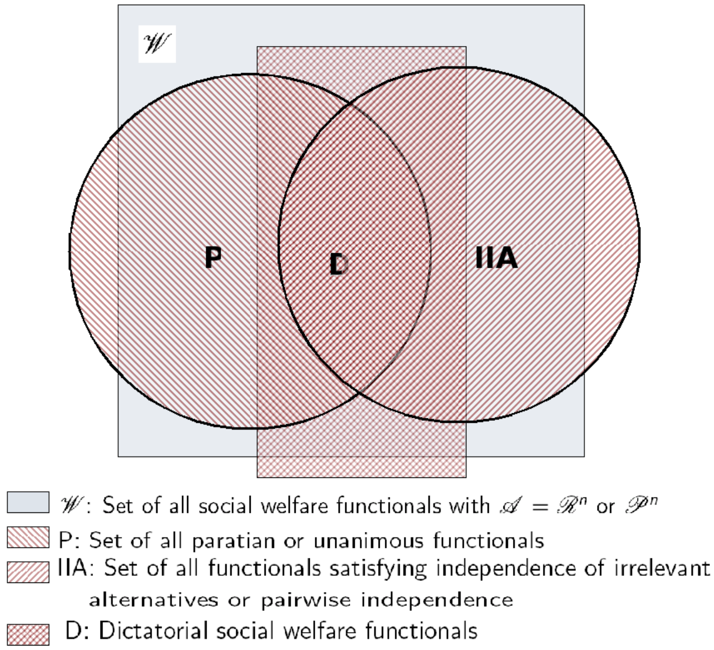



**Fig. 2.11** An illustration of the Arrow's impossibility theorem

For a proof of this theorem, we refer the reader to proposition 21.C.1 of Mas-Colell, Whinston, and Green [6]. Arrow's Impossibility Theorem is pictorially depicted in Figure 2.11. The set $P$ denotes the set of all Paretian or unanimous social welfare functionals. The set IIA denotes the set of all social welfare functionals that satisfy independence of irrelevant alternatives (or pairwise independence). The diagram shows that the intersection of $P$ and IIA is necessarily a subset of $D$, the class of all dictatorial social welfare functionals.

The Gibbard–Satterthwaite theorem has close connections to Arrow's Impossibility Theorem. The property of unanimity of social welfare functionals is related to ex-post efficiency of social choice functions. The notions of dictatorship of social welfare functionals and social choice functions are closely related. The pairwise independence property of social welfare functionals has intimate connections with the DSIC property of social choice functions through the weak preference reversal property and monotonicity. We do not delve deep into this here; interested readers are referred to the book of Mas-Colell, Whinston, and Green [6] (Chapters 21 and 23).

## 2.13 The Quasilinear Environment

This is the most extensively studied special class of environments where the Gibbard–Satterthwaite theorem does not hold. In fact, the rest of this chapter assumes this environment most of the time. In the quasilinear environment, an alternative $x \in X$ is a vector of the form $x = (k, t_1, \ldots, t_n)$, where $k$ is an element of a set $K$, which is called the set of project choices or set of allocations. The set $K$ is usually assumed to be finite. The term $t_i \in \mathbb{R}$ represents the monetary transfer to agent $i$. If $t_i > 0$ then agent $i$ will receive the money and if $t_i < 0$ then agent $i$ will pay the money. We assume that we are dealing with a system in which the $n$ agents have no external source of funding, i.e., $\sum_{i=1}^{n} t_i \leq 0$. This condition is known as the *weak budget balance* condition. The set of alternatives $X$ is therefore

$$X = \left\{ (k, t_1, \ldots, t_n) : k \in K;\ t_i \in \mathbb{R}\ \forall\, i \in N;\ \sum_i t_i \leq 0 \right\}.$$

A social choice function in this quasilinear environment takes the form $f(\theta) = (k(\theta), t_1(\theta), \ldots, t_n(\theta))$ where, for every $\theta \in \Theta$, we have $k(\theta) \in K$ and $\sum_i t_i(\theta) \leq 0$. Note that here we are using the symbol $k$ both as an element of the set $K$ and as a function going from $\Theta$ to $K$. It should be clear from the context as to which of these two we are referring. For a direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ in this environment, the agent $i$'s utility function takes the quasilinear form

$$u_i(x, \theta_i) = u_i((k, t_1, \ldots, t_n), \theta_i) = v_i(k, \theta_i) + m_i + t_i$$

where $m_i$ is agent $i$'s initial endowment of the money and the function $v_i(\cdot)$ is known as agent $i$'s valuation function. Recall from our discussion of mechanism design environment (Section 2.6) that the utility functions $u_i(\cdot)$ are common knowledge. In the context of a quasilinear environment, this implies that for any given type $\theta_i$ of any agent $i$, the social planner and every other agent $j$ have a way to know the function $v_i(., \theta_i)$. In many cases, the set $\Theta_i$ of the direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ is actually the set of all feasible valuation functions $v_i$ of agent $i$. That is, each possible function represents the possible types of agent $i$. Therefore, in such settings, reporting a type is the same as reporting a valuation function.

Immediate examples of quasilinear environment include many of the previously discussed examples, such as the first price and second price auctions (Example 2.30), the public project problem (Example 2.31), the network formation problem (Example 2.33), bilateral trade (Example 2.32), etc. In the quasilinear environment, we can define two important properties of a social choice function, namely, allocative efficiency and budget balance.

**Definition 2.39 (Allocative Efficiency (AE)).** We say that a social choice function $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ is allocatively efficient if for each $\theta \in \Theta$, $k(\theta)$ satisfies the following condition[1]

[1] *We will be using the symbol $k^*(\cdot)$ for a function $k(\cdot)$ that satisfies Equation (2.17).*

$$k(\theta) \in \underset{k \in K}{\arg\max} \sum_{i=1}^{n} v_i(k, \theta_i). \tag{2.17}$$

Equivalently,

$$\sum_{i=1}^{n} v_i(k(\theta), \theta_i) = \max_{k \in K} \sum_{i=1}^{n} v_i(k, \theta_i).$$

The above definition implies that for every $\theta \in \Theta$, the allocation $k(\theta)$ will maximize the sum of the values of the players. In other words, every allocation is a value maximizing allocation, or the objects are allocated to the players who value the objects most. This is an extremely desirable property to have for any social choice function. The above definition implicitly assumes that for any given $\theta$, the function $\sum_{i=1}^{n} v_i(., \theta_i) : K \to \mathbb{R}$ attains a maximum over the set $K$.

*Example 2.50 (Public Project Problem).* Consider the public project problem with two agents $N = \{1, 2\}$. Let the cost of the public project be 50 units of money. Let the type sets of the two players be given by

$$\Theta_1 = \Theta_2 = \{20, 60\}.$$

Each agent either has a low willingness to pay, 20, or a high willingness to pay, 60. Let the set of project choices be

$$K = \{0, 1\}$$

with 1 indicating that the project is taken up and 0 indicating that the project is dropped.

Assume that if $k = 1$, then the two agents will equally share the cost of the project by paying 25 each. If $k = 0$, the agents do not pay anything. A reasonable way of defining the valuation function would be

$$v_i(k, \theta_i) = k(\theta_i - 25).$$

This means, if $k = 0$, the agents derive zero value while if $k = 1$, the value derived is willingness to pay minus 25.

Define the following allocation function:

$$\begin{aligned} k(\theta_1, \theta_2) &= 0 \text{ if } \theta_1 = \theta_2 = 20 \\ &= 1 \text{ otherwise.} \end{aligned}$$

This means, the project is taken up only when at least one of the agents has a high willingness to pay. We can see that this function is allocatively efficient. This may be easily inferred from Table 2.12, which shows the values derived by the agents for different type profiles. The second column gives the actual value of $k$.

*Example 2.51 (A Non-Allocatively Efficient SCF).* Let the $v$ function be defined as under:

| $(\theta_1, \theta_2)$ | $k$ | $v_1(k, \theta_1)$ when $k = 0$ | $v_2(k, \theta_2)$ when $k = 0$ | $v_1(k, \theta_1)$ when $k = 1$ | $v_2(k, \theta_2)$ when $k = 1$ |
|---|---|---|---|---|---|
| $(20, 20)$ | 0 | 0 | 0 | -5 | -5 |
| $(20, 60)$ | 1 | 0 | 0 | -5 | 35 |
| $(60, 20)$ | 1 | 0 | 0 | 35 | -5 |
| $(60, 60)$ | 1 | 0 | 0 | 35 | 35 |

**Table 2.12** Values for different type profiles when $v_i(k, \theta_i) = k(\theta_i - 25)$

$$v_i(k, \theta_i) = k\theta_i \qquad i = 1, 2.$$

With respect to the above function, the allocation function $k$ defined in the previous example can be seen to be not allocatively efficient. The values for different type profiles are shown in Table 2.13. If the type profile is $(20, 20)$, the allocation is $k = 0$ and the total value of allocation is 0. However, the total value is 40 if the allocation were $k = 1$.

| $(\theta_1, \theta_2)$ | $k$ | $v_1(k, \theta_1)$ when $k = 0$ | $v_2(k, \theta_2)$ when $k = 0$ | $v_1(k, \theta_1)$ when $k = 1$ | $v_2(k, \theta_2)$ when $k = 1$ |
|---|---|---|---|---|---|
| $(20, 20)$ | 0 | 0 | 0 | 20 | 20 |
| $(20, 60)$ | 1 | 0 | 0 | 20 | 60 |
| $(60, 20)$ | 1 | 0 | 0 | 60 | 20 |
| $(60, 60)$ | 1 | 0 | 0 | 60 | 60 |

**Table 2.13** Values for different type profiles when $v_i(k, \theta_i) = k\theta_i$

**Definition 2.40 (Budget Balance (BB)).** We say that a social choice function $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ is budget balanced if for each $\theta \in \Theta$, $t_1(\theta), \ldots, t_n(\theta)$ satisfy the following condition:

$$\sum_{i=1}^{n} t_i(\theta) = 0. \tag{2.18}$$

Many authors prefer to call this property *strong budget balance*, and they refer to the property of having $\sum_{i=1}^{n} t_i(\theta) \leq 0$ as *weak budget balance*. In this monograph, we will use the term budget balance to refer to strong budget balance.

Budget balance ensures that the total receipts are equal to total payments. This means that the system is a closed one, with no surplus and no deficit. The weak budget balance property means that the total payments are greater than or equal to total receipts.

The following lemma establishes an important relationship of these two properties of an SCF with the ex-post efficiency of the SCF.

**Lemma 2.1.** *A social choice function $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ is ex-post efficient in quasilinear environment if and only if it is allocatively efficient and budget balanced.*

**Proof:** Let us assume that $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ is allocatively efficient and budget balanced. This implies that for any $\theta \in \Theta$, we have

$$\begin{aligned}\sum_{i=1}^{n} u_i(f(\theta), \theta_i) &= \sum_{i=1}^{n} v_i(k(\theta), \theta_i) + \sum_{i=1}^{n} t_i(\theta) \\ &= \sum_{i=1}^{n} v_i(k(\theta), \theta_i) + 0 \\ &\geq \sum_{i=1}^{n} v_i(k, \theta_i) + \sum_{i=1}^{n} t_i; \ \ \forall\, x = (k, t_1, \ldots, t_n) \\ &= \sum_{i=1}^{n} u_i(x, \theta_i); \ \ \forall\, (k, t_1, \ldots, t_n) \in X.\end{aligned}$$

That is if the SCF is allocatively efficient and budget balanced then for any type profile $\theta$ of the agent, the outcome chosen by the social choice function will be such that it maximizes the total utility derived by all the agents. This will automatically imply that the SCF is ex-post efficient.

To prove the other part, we will first show that if $f(\cdot)$ is not allocatively efficient, then, it cannot be ex-post efficient and next we will show that if $f(\cdot)$ is not budget balanced then it cannot be ex-post efficient. These two facts together will imply that if $f(\cdot)$ is ex-post efficient then it will have to be allocatively efficient and budget balanced, thus completing the proof of the lemma.

To start with, let us assume that $f(\cdot)$ is not allocatively efficient. This means that $\exists\, \theta \in \Theta$, and $k \in K$ such that

$$\sum_{i=1}^{n} v_i(k, \theta_i) > \sum_{i=1}^{n} v_i(k(\theta), \theta_i).$$

This implies that there exists at least one agent $j$ for whom $v_j(k, \theta_i) > v_j(k(\theta), \theta_i)$. Now consider the following alternative $x$

$$x = \Big(k, (t_i = t_i(\theta) + v_i(k(\theta), \theta_i) - v_i(k, \theta_i))_{i \neq j}, t_j = t_j(\theta)\Big).$$

It is easy to verify that $u_i(x, \theta_i) = u_i(f(\theta), \theta_i)\ \forall\, i \neq j$ and $u_j(x, \theta_i) > u_j(f(\theta), \theta_i)$, implying that $f(\cdot)$ is not ex-post efficient.

Next, we assume that $f(\cdot)$ is not budget balanced. This means that there exists at least one agent $j$ for whom $t_j(\theta) < 0$. Let us consider the following alternative $x$

$$x = \Big(k, (t_i = t_i(\theta))_{i \neq j}, t_j = 0\Big).$$

It is easy to verify that for the above alternative $x$, we have $u_i(x,\theta_i) = u_i(f(\theta),\theta_i)\ \forall\, i \neq j$ and $u_j(x,\theta_i) > u_j(f(\theta),\theta_i)$ implying that $f(\cdot)$ is not ex-post efficient.

*Q.E.D.*

The next lemma summarizes another fact about social choice functions in quasilinear environment.

**Lemma 2.2.** *All social choice functions in quasilinear environments are nondictatorial.*

**Proof:** If possible, assume that a social choice function, $f(\cdot)$, is dictatorial in the quasilinear environment. This means that there exists an agent called the dictator, say $d \in N$, such that for each $\theta \in \Theta$, we have

$$u_d(f(\theta),\theta_d) \geq u_d(x,\theta_d)\ \forall\, x \in X.$$

However, because of the environment being quasilinear, we have $u_d(f(\theta),\theta_d) = v_d(k(\theta),\theta_d) + t_d(\theta)$. Now consider the following alternative $x \in X$ :

$$x = \begin{cases} \left(k(\theta),(t_i = t_i(\theta))_{i \neq d}, t_d = t_d(\theta) - \sum_{i=1}^{n} t_i(\theta)\right) & : \quad \sum_{i=1}^{n} t_i(\theta) < 0 \\ \left(k(\theta),(t_i = t_i(\theta))_{i \neq d,j}, t_d = t_d(\theta) + \varepsilon, t_j = t_j(\theta) - \varepsilon\right) & : \quad \sum_{i=1}^{n} t_i(\theta) = 0 \end{cases}$$

where $\varepsilon > 0$ is any arbitrary number, and $j$ is any agent other than $d$. It is easy to verify, for the above outcome $x$, that we have $u_d(x,\theta_d) > u_d(f(\theta),\theta_d)$, which contradicts the fact that $d$ is a dictator.

*Q.E.D.*

In view of Lemma 2.2, the social planner need not have to worry about the nondictatorial property of the social choice function in quasilinear environments and he can simply look for whether there exists any SCF that is both ex-post efficient and dominant strategy incentive compatible. Furthermore, in the light of Lemma 2.1, we can say that the social planner can look for an SCF that is allocatively efficient, budget balanced, and dominant strategy incentive compatible. Once again the question arises whether there could exist social choice functions which satisfy all these three properties — AE, BB, and DSIC. We explore this and other questions in the forthcoming sections.

## 2.14 Groves Mechanisms

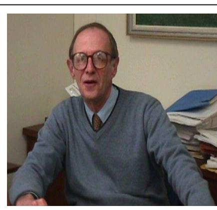
The main result in this section is that in the quasilinear environment, there exist social choice functions that are both allocatively efficient and dominant strategy incentive compatible. These are in general called the VCG (Vickrey–Clarke–Groves) mechanisms.

### *2.14.1 VCG Mechanisms*

The VCG mechanisms are named after their famous inventors William Vickrey, Edward Clarke, and Theodore Groves. It was Vickrey who introduced the famous Vickrey auction (second price sealed bid auction) in 1961 [15]. To this day, the Vickrey auction continues to enjoy a special place in the annals of mechanism design. Clarke [16] and Groves [17] came up with a generalization of the Vickrey mechanisms and helped define a broad class of dominant strategy incentive compatible mechanisms in the quasilinear environment. VCG mechanisms are by far the most extensively used among quasilinear mechanisms. They derive their popularity from their mathematical elegance and the strong properties they satisfy.

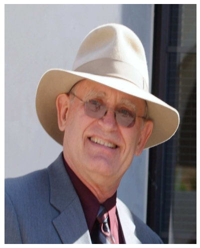
**William Vickrey** is the inventor of the famous *Vickrey Auction*, which is considered a major breakthrough in the design of auctions. He showed that the second price sealed bid auction enjoys the strong property of dominant strategy incentive compatibility, in his classic paper *Counterspeculation, Auctions, and Competitive Sealed Tenders* which appeared in the Journal of Finance in 1961. This work demonstrated for the first time the value of game theory in understanding auctions. Apart from this famous auction, Vickrey is known for an early version of revenue equivalence theorem, a key result in auction theory. He is also known for pioneering work in congestion pricing, where he introduced the idea of pricing roads and services as a natural means of regulating heavy demand. His ideas were subsequently put into practice in London city transportation. The Nobel prize in economic sciences in 1996 was jointly won by James A. Mirrlees and William Vickrey for their fundamental contributions to the economic theory of incentives under asymmetric information. However, just three days before the prize announcement, Vickrey passed away on October 11, 1996.

Vickrey was born on June 21, 1914 in Victoria, British Columbia. He earned a Ph.D. from Columbia University in 1948. His doctoral dissertation titled *Agenda for Progressive Taxation* is considered a pioneering piece of work. He taught at Columbia from 1946 until his retirement in 1982.



**Edward Clarke** distinguished himself as a senior economist with the Office of Management and Budget (Office of Information and Regulatory Affairs) involved in transportation regulatory affairs. He is a graduate of Princeton University and the University of Chicago, where he received an MBA and a Ph.D. (1978). He has worked in public policy at the city/regional (Chicago), state, federal, and international levels.

In public economics, he developed the demand revealing mechanism for public project selection, which was noted in the Nobel Committee's award of the 1996 Nobel Prize in Economics to William Vickrey. Clarke's 1971 paper *Multi-part Pricing of Public Goods* in the journal Public Choice in 1971 is a classic in mechanism design. Among VCG mechanisms, Clarke's mechanism is a natural and popular approach used in mechanism design problems. The website http://www.clarke.pair.com/clarke.html may be looked up for more details about Clarke's research and teaching.



**Theodore Groves** is credited with the most general among the celebrated class of VCG mechanisms. In a classic paper entitled *Incentives in Teams* published in Econometrica in 1973, Groves proposed a general class of allocatively efficient, dominant strategy incentive compatible mechanisms. The Groves mechanism generalizes the Clarke mechanism (proposed in 1971), which in turn generalizes the Vickrey auction proposed in 1961. Groves earned a doctorate in economics at the University of California, Berkeley, and he is currently a Professor of Economics at the University of California, San Diego. The website http://weber.ucsd.edu/ tgroves/ may be looked up for more details about Groves's research and teaching.

### *2.14.2 The Groves' Theorem*

The following theorem provides a sufficient condition for an allocatively efficient social function in quasilinear environment to be dominant strategy incentive compatible. We will refer to this theorem in the sequel as Groves theorem, rather than Groves' theorem.

**Theorem 2.7 (Groves Theorem).** *Let the SCF* $f(\cdot) = (k^*(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ *be allocatively efficient. Then* $f(\cdot)$ *is dominant strategy incentive compatible if it satisfies the following payment structure (popularly known as the Groves payment (incentive) scheme):*

$$t_i(\theta) = \left[\sum_{j \neq i} v_j(k^*(\theta), \theta_j)\right] + h_i(\theta_{-i}) \;\; \forall\, i = 1, \ldots, n \tag{2.19}$$

*where* $h_i : \Theta_{-i} \to \mathbb{R}$ *is any arbitrary function that honors the feasibility condition* $\sum_i t_i(\theta) \leq 0 \; \forall\, \theta \in \Theta$.

**Proof**: The proof is by contradiction. Suppose $f(\cdot)$ satisfies both allocative efficiency and the Groves payment structure but is not DSIC. This implies that $f(\cdot)$ does not satisfy the following necessary and sufficient condition for DSIC: $\forall i \in N \;\; \forall \theta \in \Theta$,

$$u_i(f(\theta_i, \theta_{-i}), \theta_i) \geq u_i(f(\theta_i^{'}, \theta_{-i}), \theta_i) \;\; \forall \theta_i^{'} \in \Theta_i \;\; \forall \theta_{-i} \in \Theta_{-i}.$$

This implies that there exists at least one agent $i$ for which the above is false. Let $i$ be one such agent. That is, for agent $i$,

$$u_i(f(\theta_i^{'}, \theta_{-i}), \theta_i) > u_i(f(\theta_i, \theta_{-i}), \theta_i)$$

for some $\theta_i \in \Theta_i$, for some $\theta_{-i} \in \Theta_{-i}$, and for some $\theta_i^{'} \in \Theta_i$. Thus, for agent $i$, there would exist $\theta_i \in \Theta_i, \theta_i^{'} \in \Theta_i, \theta_{-i} \in \Theta_{-i}$ such that

$$v_i(k^*(\theta_i^{'}, \theta_{-i}), \theta_i) + t_i(\theta_i^{'}, \theta_{-i}) + m_i > v_i(k^*(\theta_i, \theta_{-i}), \theta_i) + t_i(\theta_i, \theta_{-i}) + m_i.$$

Recall that

$$t_i(\theta_i, \theta_{-i}) = h_i(\theta_{-i}) + \sum_{j \neq i} (k^*(\theta_i, \theta_{-i}), \theta_j)$$

$$t_i(\theta_i^{'}, \theta_{-i}) = h_i(\theta_{-i}) + \sum_{j \neq i} (k^*(\theta_i^{'}, \theta_{-i}), \theta_j).$$

Substituting these, we get

$$v_i(k^*(\theta_i^{'}, \theta_{-i}), \theta_i) + \sum_{j \neq i} v_i(k^*(\theta_i^{'}, \theta_{-i}), \theta_j) > v_i(k^*(\theta_i, \theta_{-i}), \theta_i) + \sum_{j \neq i} v_i(k^*(\theta_i, \theta_{-i}), \theta_j),$$

which implies

$$\sum_{i=1}^{n} v_i(k^*(\theta_i^{'}, \theta_{-i}), \theta_i) > \sum_{i=1}^{n} v_i(k^*(\theta_i, \theta_{-i}), \theta_i).$$

The above contradicts the fact that $f(\cdot)$ is allocatively efficient. This completes the proof.

*Q.E.D.*

The following are a few interesting implications of the above theorem.

1. Given the announcements $\theta_{-i}$ of agents $j \neq i$, the monetary transfer to agent $i$ depends on his announced type only through effect of the announcement of agent $i$ on the project choice $k^*(\theta)$.
2. The change in the monetary transfer of agent $i$ when his type changes from $\theta_i$ to $\hat{\theta}_i$ is equal to the effect that the corresponding change in project choice has on total value of the rest of the agents. That is,

$$t_i(\theta_i, \theta_{-i}) - t_i(\hat{\theta}_i, \theta_{-i}) = \sum_{j \neq i} \left[ v_j(k^*(\theta_i, \theta_{-i}), \theta_j) - v_j(k^*(\hat{\theta}_i, \theta_{-i}), \theta_j) \right].$$

   Another way of describing this is to say that the change in monetary transfer to agent $i$ reflects exactly the externality he is imposing on the other agents.

After the famous result of Groves, a direct revelation mechanism in which the implemented SCF is allocatively efficient and satisfies the Groves payment scheme is called a *Groves Mechanism*.

**Definition 2.41 (Groves Mechanisms).** A direct mechanism, $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ in which $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ satisfies allocative efficiency (2.17) and Groves payment rule (2.19) is known as a Groves mechanism.

In mechanism design parlance, Groves mechanisms are popularly known as Vickrey–Clarke–Groves (VCG) mechanisms because the Clarke mechanism is a special case of Groves mechanism, and the Vickrey mechanism is a special case of Clarke mechanism. We will discuss this relationship later in this monograph.

The Groves theorem provides a sufficiency condition under which an allocatively efficient (AE) SCF will be DSIC. The following theorem due to Green and Laffont

[18] provides a set of conditions under which the condition of Groves Theorem also becomes a necessary condition for an AE SCF to be DSIC. In this theorem, we let $\mathscr{F}$ denote the set of all possible functions $f : K \rightarrow \mathbb{R}$.

**Theorem 2.8 (First Characterization Theorem of Green–Laffont).** *Suppose for each agent $i \in N$ that $\{v_i(., \theta_i) : \theta_i \in \Theta_i\} = \mathscr{F}$, that is, every possible valuation function from $K$ to $\mathbb{R}$ arises for some $\theta_i \in \Theta_i$. Then any allocatively efficient social choice function $f(\cdot)$ will be dominant strategy incentive compatible if and only if it satisfies the Groves payment scheme given by Equation (2.19).*

Note that in the above theorem, every possible valuation function from $K$ to $\mathbb{R}$ arises for any $\theta_i \in \Theta_i$. In the following characterization theorem, again due to Green and Laffont [18], $\mathscr{F}$ is replaced with with $\mathscr{F}_c$ where $\mathscr{F}_c$ denotes the set of all possible continuous functions $f : K \rightarrow \mathbb{R}$.

**Theorem 2.9 (Second Characterization Theorem of Green–Laffont).** *Suppose for each agent $i \in N$ that $\{v_i(., \theta_i) : \theta_i \in \Theta_i\} = \mathscr{F}_c$, that is, every possible continuous valuation function from $K$ to $\mathbb{R}$ arises for some $\theta_i \in \Theta_i$. Then any allocatively efficient social choice function $f(\cdot)$ will be dominant strategy incentive compatible if and only if it satisfies the Groves payment scheme given by Equation (2.19).*

## 2.14.3 Groves Mechanisms and Budget Balance

Note that a Groves mechanism always satisfies the properties of AE and DSIC. Therefore, if a Groves mechanism is budget balanced, then it will solve the problem of the social planner because it will then be ex-post efficient and dominant strategy incentive compatible. By looking at the definition of the Groves mechanism, one can conclude that it is the functions $h_i(\cdot)$ that decide whether or not the Groves mechanism is budget balanced. The natural question that arises now is whether there exists a way of defining functions $h_i(\cdot)$ such that the Groves mechanism is budget balanced. In what follows, we present one possibility result and one impossibility result in this regard.

### 2.14.3.1 Possibility and Impossibility Results for Quasilinear Environments

Green and Laffont [18] showed that in a quasilinear environment, if the set of possible types for each agent is sufficiently rich then ex-post efficiency and DSIC cannot be achieved together. The precise statement is given in the form of the following theorem.

**Theorem 2.10 (Green–Laffont Impossibility Theorem).** *Suppose for each agent $i \in N$ that $\mathscr{F} = \{v_i(., \theta_i) : \theta_i \in \Theta_i\}$, that is, every possible valuation function from $K$ to $\mathbb{R}$ arises for some $\theta_i \in \Theta_i$. Then there is no social choice function that is ex-post efficient and DSIC.*

Thus, the above theorem says that if the set of possible types for each agent is sufficiently rich then there is no hope of finding a way to define the functions $h_i(\cdot)$ in Groves payment scheme so that we have $\sum_{i=1}^{n} t_i(\theta) = 0$. However, one special case in which a positive result arises is summarized in the form of following possibility result.

**Theorem 2.11 (A Possibility Result for Budget Balance of Groves Mechanisms).** *If there is at least one agent whose preferences are known (that is, the type set is a singleton set) then it is possible to choose the functions $h_i(\cdot)$ so that $\sum_{i=1}^{n} t_i(\theta) = 0$.*

**Proof**: Let agent $i$ be such that his preferences are known, that is $\Theta_i = \{\theta_i\}$. In view of this condition, it is easy to see that for an allocatively efficient social choice function $f(\cdot) = (k^*(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$, the allocation $k^*(\cdot)$ depends only on the types of the agents other than $i$. That is, the allocation $k^*(\cdot)$ is a mapping from $\Theta_{-i}$ to $K$. Let us define the functions $h_j(\cdot)$ in the following manner:

$$h_j(\theta_{-j}) = \begin{cases} h_j(\theta_{-j}) & : \quad j \neq i \\ -\sum_{r \neq i} h_r(\theta_{-r}) - (n-1)\sum_{r=1}^{n} v_r(k^*(\theta), \theta_r) & : \quad j = i. \end{cases}$$

It is easy to see that under the above definition of the functions $h_i(\cdot)$, we will have $t_i(\theta) = -\sum_{j \neq i} t_j(\theta)$. *Q.E.D.*

Figure 2.12 summarizes the main results of this section by showing what the space of social choice functions looks like in the quasilinear environment. The exhibit brings out various possibilities and impossibilities in the quasilinear environment, based on the results that we have discussed so far.

## 2.15 Clarke (Pivotal) Mechanisms

A special case of Groves mechanism was developed independently by Clarke in 1971 [16] and is known as the *Clarke*, or the *pivotal* mechanism. It is a special case of Groves mechanisms in the sense of using a natural special form for the function $h_i(\cdot)$. In the Clarke mechanism, the function $h_i(\cdot)$ is given by the following relation:

$$h_i(\theta_{-i}) = -\sum_{j \neq i} v_j(k^*_{-i}(\theta_{-i}), \theta_j) \;\; \forall\, \theta_{-i} \in \Theta_{-i}, \forall\, i = 1, \ldots, n \tag{2.20}$$

where $k^*_{-i}(\theta_{-i}) \in K_{-i}$ is the choice of a project that is allocatively efficient if there were only the $n-1$ agents $j \neq i$. Formally, $k^*_{-i}(\theta_{-i})$ must satisfy the following condition.

$$\sum_{j \neq i} v_j(k^*_{-i}(\theta_{-i}), \theta_j) \geq \sum_{j \neq i} v_j(k, \theta_j) \; \forall\, k \in K_{-i} \tag{2.21}$$

where the set $K_{-i}$ is the set of project choices available when agent $i$ is absent. Substituting the value of $h_i(\cdot)$ from Equation (2.20) in Equation (2.19), we get the following expression for agent $i$'s transfer in the Clarke mechanism:

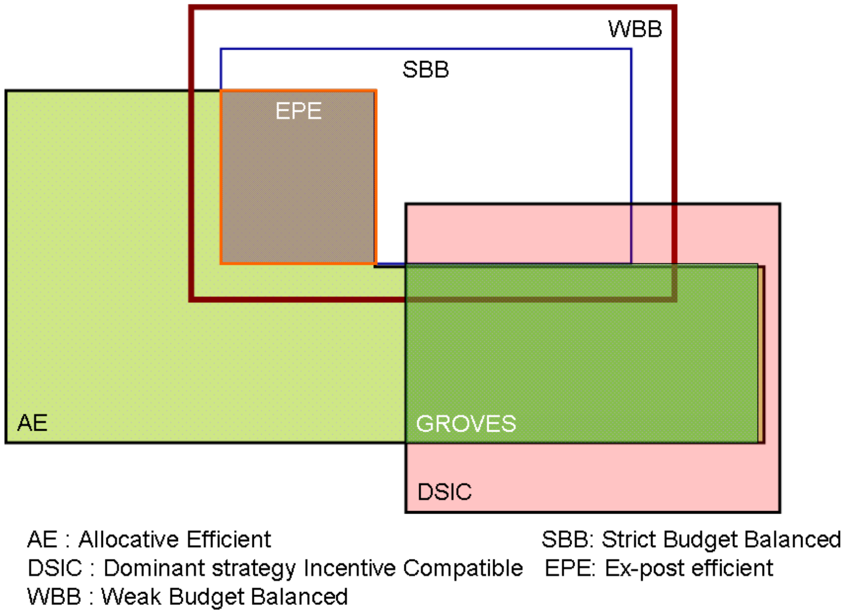



**Fig. 2.12** Space of social choice functions in quasilinear environment

$$t_i(\theta) = \left[\sum_{j \neq i} v_j(k^*(\theta), \theta_j)\right] - \left[\sum_{j \neq i} v_j(k^*_{-i}(\theta_{-i}), \theta_j)\right]. \qquad (2.22)$$

The above payment rule has an appealing interpretation: Given a type profile $\theta = (\theta_1, \ldots, \theta_n)$, the monetary transfer to agent $i$ is given by the total value of all agents other than $i$ under an efficient allocation when agent $i$ is present in the system minus the total value of all agents other than $i$ under an efficient allocation when agent $i$ is absent in the system.

### *2.15.1 Clarke Mechanisms and Weak Budget Balance*

Recall from the definition of Groves mechanisms that, for weak budget balance, we should choose the functions $h_i(\theta_{-i})$ in a way that the weak budget balance condition $\sum_{i=1}^{n} t_i(\theta) \leq 0$ is satisfied. In this sense, the Clarke mechanism is a useful special case because it achieves weak budget balance under fairly general settings. In order to understand these general sufficiency conditions, we define following quantities

$$B^*(\theta) = \left\{ k \in K : k \in \begin{matrix} \arg\max \\ k \in K \end{matrix} \sum_{j=1}^{n} v_j(k, \theta_j) \right\}$$

$$B^*(\theta_{-i}) = \left\{ k \in K_{-i} : k \in \begin{matrix} \arg\max \\ k \in K_{-i} \end{matrix} \sum_{j \neq i} v_j(k, \theta_j) \right\}$$

where $B^*(\theta)$ is the set of project choices that are allocatively efficient when all the agents are present in the system. Similarly, $B^*(\theta_{-i})$ is the set of project choices that are allocatively efficient if all agents except agent $i$ were present in the system. It is obvious that $k^*(\theta) \in B^*(\theta)$ and $k^*_{-i}(\theta_{-i}) \in B^*(\theta_{-i})$.

Using the above quantities, we define the following properties of a direct revelation mechanism in quasilinear environment.

**Definition 2.42 (No Single Agent Effect).** We say that mechanism $\mathscr{M}$ has no single agent effect if for each agent $i$, for each $\theta \in \Theta$, and for each $k^*(\theta) \in B^*(\theta)$, we have a $k \in K_{-i}$ such that

$$\sum_{j \neq i} v_j(k, \theta_j) \geq \sum_{j \neq i} v_j(k^*(\theta), \theta_j).$$

In view of the above properties, we have the following proposition that gives a sufficiency condition for Clarke mechanism to be weak budget balanced.

**Proposition 2.5.** *If the Clarke mechanism has no single agent effect, then the monetary transfer to each agent would be non-positive, that is,* $t_i(\theta_i) \leq 0 \; \forall \, \theta \in \Theta; \; \forall \, i = 1, \ldots, n$*. In such a situation, the Clarke mechanism would satisfy the weak budget balance property.*

**Proof:** Note that by virtue of no single agent effect, for each agent $i$, each $\theta \in \Theta$, and each $k^*(\theta) \in B^*(\theta)$, there exists a $k \in K_{-i}$ such that

$$\sum_{j \neq i} v_j(k, \theta_j) \geq \sum_{j \neq i} v_j(k^*(\theta), \theta_j).$$

However, by definition of $k^*_{-i}(\theta_{-i})$, given by Equation (2.21), we have

$$\sum_{j \neq i} v_j(k^*_{-i}(\theta_{-i}), \theta_j) \geq \sum_{j \neq i} v_j(k, \theta_j) \; \forall \, k \in K_{-i}.$$

Combining the above two facts, we get

$$\begin{aligned} \sum_{j \neq i} v_j(k^*_{-i}(\theta_{-i}), \theta_j) &\geq \sum_{j \neq i} v_j(k^*(\theta), \theta_j) \\ &\Rightarrow 0 \geq t_i(\theta) \\ &\Rightarrow 0 \geq \sum_{i=1}^{n} t_i(\theta). \end{aligned}$$

*Q.E.D.*

In what follows, we present an interesting corollary of the above proposition.

**Corollary 2.1.**

1. *$t_i(\theta) = 0$ iff $k^*(\theta) \in B^*(\theta_{-i})$. That is, agent i's monetary transfer is zero iff his announcement does not change the project decision relative to what would be allocatively efficient for agents $j \neq i$ in isolation.*
2. *$t_i(\theta) < 0$ iff $k^*(\theta) \notin B^*(\theta_{-i})$. That is, agent i's monetary transfer is negative iff his announcement changes the project decision relative to what would be allocatively efficient for agents $j \neq i$ in isolation. In such a situation, the agent i is known to be "pivotal" to the efficient project choice, and he pays a tax equal to his effect on the other agents.*

### *2.15.2 Clarke Mechanisms and Individual Rationality*

We have studied individual rationality (also called voluntary participation) property in Section 2.10.3. The following proposition investigates the individual rationality of the Clarke mechanism. First, we provide two definitions.

**Definition 2.43 (Choice Set Monotonicity).** We say that a mechanism $\mathscr{M}$ is choice set monotone if the set of feasible outcomes $X$ (weakly) increases as additional agents are introduced into the system. An implication of this property is $K_{-i} \subset K \; \forall \, i = 1, \ldots, n$.

**Definition 2.44 (No Negative Externality).** Consider a choice set monotone mechanism $\mathscr{M}$. We say that the mechanism $\mathscr{M}$ has no negative externality if for each agent $i$, each $\theta \in \Theta$, and each $k^*_{-i}(\theta_{-i}) \in B^*(\theta_{-i})$, we have

$$v_i(k^*_{-i}(\theta_{-i}), \theta_i) \geq 0.$$

We now state and prove a proposition which provides a sufficient condition for the ex-post individual rationality of the Clarke mechanism. Recall from Section 2.10.3 the notation $\overline{u_i}(\theta_i)$, which represents the utility that agent $i$ receives by withdrawing from the mechanism.

**Proposition 2.6 (Ex-Post Individual Rationality of Clarke Mechanism ).** *Let us consider a Clarke mechanism in which*

1. *$\overline{u_i}(\theta_i) = 0 \; \forall \theta_i \in \Theta_i; \; \forall \, i = 1, \ldots, n,$*
2. *The mechanism satisfies choice set monotonicity property,*
3. *The mechanism satisfies no negative externality property.*

*Then the Clarke mechanism is ex-post individual rational.*

**Proof:** Recall that utility $u_i(f(\theta), \theta_i)$ of an agent $i$ in Clarke mechanism is given by

$$\begin{aligned} u_i(f(\theta),\theta_i) &= v_i(k^*(\theta),\theta_i) + \left[\sum_{j\neq i} v_j(k^*(\theta),\theta_j)\right] - \left[\sum_{j\neq i} v_j(k^*_{-i}(\theta_{-i}),\theta_j)\right] \\ &= \left[\sum_{j} v_j(k^*(\theta),\theta_j)\right] - \left[\sum_{j\neq i} v_j(k^*_{-i}(\theta_{-i}),\theta_j)\right]. \end{aligned}$$

By virtue of choice set monotonicity, we know that $k^*_{-i}(\theta_{-i}) \in K$. Therefore, we have

$$\begin{aligned} u_i(f(\theta),\theta_i) &\geq \left[\sum_{j} v_j(k^*_{-i}(\theta_{-i}),\theta_j)\right] - \left[\sum_{j\neq i} v_j(k^*_{-i}(\theta_{-i}),\theta_j)\right] \\ &= v_i(k^*_{-i}(\theta_{-i}),\theta_i) \\ &\geq 0 = \overline{u_i}(\theta_i). \end{aligned}$$

The last step follows due to the fact that the mechanism has no negative externality. *Q.E.D.*

*Example 2.52 (Individual Rationality in Sealed Bid Auction).* Let us consider the example of first-price sealed bid auction. If for each possible type $\theta_i$, the utility $\overline{u_i}(\theta_i)$ derived by the agents $i$ from not participating in the auction is 0, then it is easy to see that the SCF used in this example would be ex-post IR.

Let us next consider the example of a second-price sealed bid auction. If for each possible type $\theta_i$, the utility $\overline{u_i}(\theta_i)$ derived by the agents $i$ from not participating in the auction is 0, then it is easy to see that the SCF used in this example would be ex-post IR. Moreover, the ex post IR of this example also follows directly from Proposition 2.6 because this is a special case of the Clarke mechanism satisfying all the required conditions in the proposition.

## 2.16 Examples of VCG Mechanisms

VCG mechanisms derive their popularity on account of the elegant mathematical and economic properties that they have and the revealing first level insights they provide during the process of designing mechanisms for a game theoretic problem. For this reason, invariably, mechanism design researchers try out VCG mechanisms first. However, VCG mechanisms do have many limitations. The virtues and limitations of VCG mechanisms are captured by Ausubel and Milgrom [19], whereas a recent paper by Rothkopf [20] summarizes the practical limitations of applying VCG mechanisms. In this section, we provide a number of examples to illustrate some interesting nuances of VCG mechanisms.

*Example 2.53 (Vickrey Auction for a Single Indivisible Item).* Consider 5 bidders $\{1,2,3,4,5\}$, with valuations $v_1 = 20; v_2 = 15; v_3 = 12; v_4 = 10; v_5 = 6$, participating in a sealed bid auction for a single indivisible item. If Vickrey auction is the mechanism used, then it is a dominant strategy for the agents to bid their valuations. Agent 1 with valuation 20 will be the winner, and the monetary transfer to agent 1

$$
\begin{aligned}
&= \sum_{j\neq 1} v_j(k^*(\theta),\theta_j) - \sum_{j\neq 1} v_j(k^*(\theta_{-1}),\theta_j) \\
&= 0 - 15 = -15.
\end{aligned}
$$

This means agent 1 would pay an amount equal to 15, which happens to be the second highest bid (in this case the second highest valuation). Note that 15 is the change in the total value of agents other than agent 1 when agent 1 drops out of the system. This is the externality that agent 1 imposes on the rest of the agents. This externality becomes the payment of agent 1 when he wins the auction.

Another way of determining the payment by agent 1 is to compute his marginal contribution to the system. The total value in the presence of agent 1 is 20, while the total value in the absence of agent 1 is 15. Thus the marginal contribution of agent 1 is 5. The above marginal contribution is given as a discount to agent 1 by the Vickrey payment mechanism, and agent 1 pays $20-5=15$. Such a discount is known as the *Vickrey discount*.

If we use the Clarke mechanism, we have

$$
\begin{aligned}
t_i(\theta_i,\theta_{-i}) &= \sum_{j\neq i} v_j(k^*(\theta),\theta_j) - \sum_{j\neq i} v_j(k^*_{-i}(\theta_{-i}),\theta_j) \\
&= \sum_{j\in N} v_j(k^*(\theta),\theta_j) - v_i(k^*(\theta),\theta_i) - \sum_{j\neq i} v_j(k^*_{-i}(\theta_{-i}),\theta_j) \\
&= \sum_{j\in N} v_j(k^*(\theta),\theta_j) - \sum_{j\neq i} v_j(k^*_{-i}(\theta_{-i}),\theta_j) - v_i(k^*(\theta),\theta_i).
\end{aligned}
$$

The difference in the first two terms represents the marginal contribution of agent $i$ to the system while the term $v_i(k^*(\theta),\theta_i)$ is the value received by agent $i$.

*Example 2.54 (Vickrey Auction for Multiple Identical Items).* Consider the same set of bidders as above but with the difference that there are 3 identical items available for auction. Each bidder wants only one item. If we apply the Clarke mechanism for this situation, bidders 1, 2, and 3 become the winners. The payment by bidder 1

$$
\begin{aligned}
&= \sum_{j\neq 1} v_j(k^*(\theta),\theta_j) - \sum_{j\neq 1} v_j(k^*_{-1}(\theta_{-1}),\theta_j) \\
&= (15+12) - (15+12+10) \\
&= -10.
\end{aligned}
$$

Thus bidder 1 pays an amount equal to the highest nonwinning bid. Similarly, one can verify that the payment to be made by the other two winners (namely agent 2 and agent 3) is also equal to 10. This payment is consistent with their respective marginal contributions.

$$\text{Marginal contribution of agent 1} = (20+15+12)-(15+12+10) = 10$$

$$\text{Marginal contribution of agent 2} = (20+15+12)-(20+12+10) = 5$$

$$\text{Marginal contribution of agent 3} = (20+15+12)-(20+15+10) = 2.$$

In the above example, let the demand by agent 1 be 2 units with the rest of agents continuing to have unit demand. Now the allocation will allocate 2 units to agent 1 and 1 unit to agent 2.

$$\text{Payment by agent 1} = 15-(15+12+10) = -22$$

$$\text{Payment by agent 2} = 40-(40+12) = -12.$$

This is because the marginal contribution of agent 1 and agent 2 are given by: agent 1: $55-37=18$; agent 2: $55-52=3$.

*Example 2.55 (Generalized Vickrey Auction).* Generalized Vickrey auction (GVA) refers to an auction that results when the Clarke mechanism is applied to a combinatorial auction. A combinatorial auction is one where the bids correspond to bundles or combinations of different items. In a *forward combinatorial auction*, a bundle of different types of goods is available with the seller; the buyers are interested in purchasing certain subsets of the items. In a *reverse combinatorial auction*, a bundle of different types of goods is required by the buyer; several sellers are interested in selling subsets of the goods to the buyer. There is a rich body of literature on combinatorial auctions, for example see the edited volume [21]. We discuss a simple example here. Let a seller be interested in auctioning two items A and B. Let there be three buying agents $\{1,2,3\}$. Let us abuse the notation slightly and denote the subsets $\{A\}$, $\{B\}$, $\{A,B\}$ by A, B, and AB, respectively. These are called combinations or bundles. Assume that the agents have *valuations* for the bundles as shown in Table 2.14. In the above table, a "*" indicates that the agent is not interested in

| | A | B | AB |
|---|---|---|---|
| Agent 1 | * | * | 10 |
| Agent 2 | 5 | * | * |
| Agent 3 | * | 5 | * |

**Table 2.14** Valuations of agents for bundles in scenario 1

that bundle. Note from Table 2.14 that agent 1 values bundle AB at 10 and does not have any valuation for bundle A and bundle B. Agent 2 is only interested in bundle A and has a valuation of 5 for this bundle. Agent 3 is only interested in bundle B and has a valuation of 5 for this bundle. If we apply the Clarke mechanism to this situation, the bids from the agents will be identical to the valuations because of the DSIC property of the Clarke mechanism. There are two allocatively efficient allocations, namely: (1) Allocate bundle AB to agent 1; (2) Allocate bundle A to agent

2 and bundle B to agent 3. Each of these allocations has a total value of 10. Suppose we choose allocation (2), which awards bundle A to agent 2 and bundle B to agent 3. To compute the payments to be made by agents 2 and 3, we have to use the Clarke payment rule. For this, we analyze what would happen in the absence of agent 2 and agent 3 separately. If agent 2 is absent, the allocation will award the bundle AB to agent 1 resulting in a total value of 10. Therefore, the Vickrey discount to agent 2 is $10-10=0$, which means payment to be made by agent 2 is $5+0=5$. Similarly the Vickrey discount to agent 3 is also 0 and the payment to be made by agent 3 is also equal to 5. The total revenue to the seller is $5+5=10$. Even if allocation (1) is chosen (that is, award bundle AB to agent 1), the total revenue to the seller remains as 10. This is a situation where the seller is able to capture the entire consumer surplus.

A contrasting situation will result if the valuations are as shown in Table 2.15. In

| | A | B | AB |
|---|---|---|---|
| Agent 1 | * | * | 10 |
| Agent 2 | 10 | * | * |
| Agent 3 | * | 10 | * |

**Table 2.15** Valuations of agents for bundles in scenario 2

this case, the winning allocation is: award bundle A to agent 2 and bundle B to agent 3, resulting in a total value of 20. If agent 2 is not present, the allocation will be to award bundle AB to agent 1, thus resulting in a total value of 10. Similarly, if agent 3 were not present, the allocation would be to award bundle AB to agent 1, thus resulting in a total value of 10. This would mean a Vickrey discount of 10 each to agent 2 and agent 3, which in turn means that the the payment to be made by agent 2 and agent 3 is 0 each! This represents a situation where the seller will end up with a zero revenue in the process of guaranteeing allocative efficiency and dominant strategy incentive compatibility. Worse still, if agent 2 and agent 3 are both the false names of a single agent, then the auction itself is seriously manipulated!

We now study a third scenario where the valuations are as described in Table 2.16. Here, the allocation is to award bundle AB to agent 1, resulting in a total value

| | A | B | AB |
|---|---|---|---|
| Agent 1 | * | * | 10 |
| Agent 2 | 2 | * | * |
| Agent 3 | * | 2 | * |

**Table 2.16** Valuations of agents for bundles in scenario 3

of 10. If agent 1 were absent, the allocation would be to award bundle A to agent 2 and bundle B to agent 3, which leads to a total value of 4. The Vickrey discount to agent 1 is therefore $10-4=6$, and the payment to be made by agent 1 is 4.

The revenue to the seller is also 4. Contrast this scenario with scenario 2, where the valuations of bidders 2 and 3 were higher, but they were able to win the bundles by paying nothing. This shows that the GVA mechanism is not foolproof against bidder collusion (in this case, bidders 2 and 3 can collude and deny the bundle to agent 1 and also seriously reduce the revenue to the seller.

*Example 2.56 (Strategy Proof Mechanism for the Public Project Problem).* Consider the public project problem discussed in Example 2.31. We shall compute the Clarke payments by each agent for each type of profile. We will also compute the utilities. First consider the type profile (20,20). Since $k = 0$, the values derived by either agent is zero. Hence the Clarke payment by each agent is zero, and the utilities are also zero.

Next consider the type profile (60, 20). Note that $k(60,20) = 1$. Agent 1 derives a value 35 and agent 2 derives a value $-5$. If agent 1 is not present, then agent 2 is left alone and the allocation will be 0 since its willingness to pay is 20. Thus the value to agent 2 when agent 1 is not present is 0. This means

$$t_1(60,20) = -5-0 = -5.$$

This implies agent 1 would pay an amount of 5 units in addition to 25 units, which is its contribution to the cost of the project. The above payment is consistent with the marginal contribution of agent 1, which is equal to $(60-25)+(20-25)-0 = 35-5 = 30$.

We can now determine the utility of agent 1, which will be

$$\begin{aligned} u_1(60,20) &= v_1(60,20)+t_1(60,20) \\ &= 35-5 = 30. \end{aligned}$$

To compute $t_2(60,20)$, we first note that the value to the agent 1 when agent 2 is not present is $(60-50)$. Therefore

$$t_2(60,20) = 35-10 = 25.$$

This means agent 2 receives 25 units of money; of course, this is besides the 25 units of money it pays towards the cost of the project. Now

$$\begin{aligned} u_2(60,20) &= v_2(60,20)+t_2(60,20) \\ &= -5+25 \\ &= 20. \end{aligned}$$

Likewise, we can compute the payments and utilities of the agents for all the type profiles. Table 2.17 provides these values. Note that this mechanism is ex-post individually rational assuming that the utility for not participating in the mechanism is zero.

*Example 2.57 (Strategy Proof Network Formation).* Consider the problem of forming a supply chain as depicted in Figure 2.13. The node S represents a start-

| $(\theta_1,\theta_2)$ | $t_1(\theta_1,\theta_2)$ | $t_2(\theta_1,\theta_2)$ | $u_1(\theta_1,\theta_2)$ | $u_2(\theta_1,\theta_2)$ |
|---|---|---|---|---|
| (20, 20) | 0 | 0 | 0 | 0 |
| (60, 20) | -5 | 25 | 30 | 20 |
| (20, 60) | 25 | -5 | 20 | 30 |
| (60, 60) | 25 | 25 | 60 | 60 |

**Table 2.17** Payments and utilities for different type profiles

ing state and T represents a target state; A and B are two intermediate states. $SP_1, SP_2, SP_3, SP_4$ are four different service providers. In the figure, the service

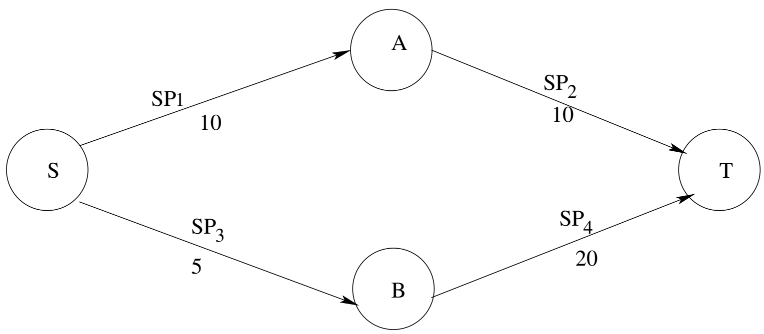



**Fig. 2.13** A network formation problem - case 1

providers are represented as owners of the respective edges. The cost of providing service (willingness to sell) is indicated on each edge. The problem is to procure a path from S to T having minimum cost. Let $(y_1, y_2, y_3, y_4)$ represent the allocation vector. The feasible allocation vectors are

$$K = \{(1,1,0,0),(0,0,1,1)\}.$$

Among these, the allocation $(1,1,0,0)$ is allocatively efficient since it minimizes the cost of allocation. We shall define the value as follows:

$$v_i((y_1,y_2,y_3,y_4);\theta_i) = -y_i\theta_i.$$

The above manner of defining the values reflects the fact that cost minimization is the same as value maximization. Applying Clarke's payment rule, we obtain

$$t_1(\theta) = -10-(-25) = 15$$
$$t_2(\theta) = -10-(-25) = 15.$$

Note that each agent gets a surplus of 5, being its marginal contribution. The utilities for these two agents are

$$u_1(\theta) = -10 + 15 = 5$$
$$u_2(\theta) = -10 + 15 = 5.$$

The payments and utilities for $SP_3$ and $SP_4$ are zero. Let us study the effect of changing the willingness to sell of $SP_4$. Let us make it as 15. Then, we find that both the allocations $(1,1,0,0)$ and $(0,0,1,1)$ are allocatively efficient. If we choose the allocation $(1,1,0,0)$, we get the payments as

$$t_1(\theta) = 10$$
$$t_2(\theta) = 10$$
$$u_1(\theta) = 0$$
$$u_2(\theta) = 0.$$

This means that the payments to the service providers are equal to the costs. There is no surplus payment to the winning agents. In this case, the mechanism is friendly to the buyer and unfriendly to the sellers.

If we make the willingness to sell of $SP_4$ as 95, the allocation $(1,1,0,0)$ is efficient and we get the payments as

$$t_1(\theta) = 90$$
$$t_2(\theta) = 90$$
$$u_1(\theta) = 80$$
$$u_2(\theta) = 80.$$

In this case, the mechanism is extremely unfriendly to the buyer but is very attractive to the sellers.

Let us introduce one more edge from B to A and see the effect. See Figure 2.14.

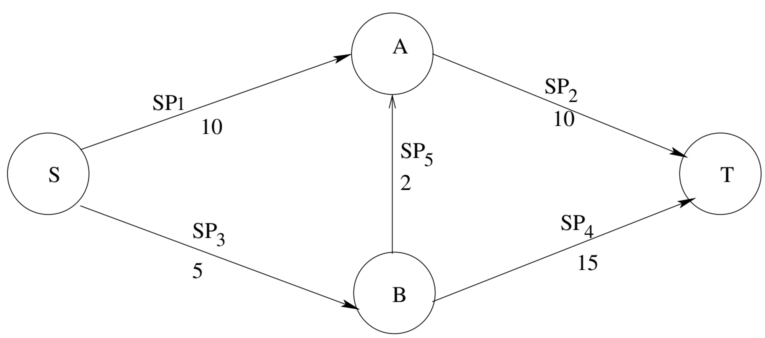



**Fig. 2.14** A network formation problem - case 2

The efficient allocation here is (0, 1, 1, 0, 1). The payments are

$$t_2(\theta) = 13$$
$$t_3(\theta) = 8$$
$$t_5(\theta) = 5$$
$$u_2(\theta) = 3$$
$$u_3(\theta) = 3$$
$$u_2(\theta) = 3.$$

This shows that the total payments to be made by the buyer is 26 whereas the total payment if the service provider $SP_5$ were absent is 20. Thus in spite of an additional agent being available, the payment to the buyer is higher. This shows a kind of non-monotonicity exhibited by the Clarke payment rule.

*Example 2.58 (A Groves (but not Clarke) Mechanism).* Consider a sealed bid auction for a single indivisible item where the bidder with the highest bid is declared as the winner and the winner pays an amount equal to equal to twice the bid of the bidder with the lowest valuation among the rest of the agents. Such a payment rule is not a Clarke payment rule but does belong to the Groves payment scheme.

As an example consider 5 bidders with valuations 20, 15, 12, 10, 8. The bidder with valuation 20 is declared the winner and will pay an amount = 16. On the other hand, if there are only three bidders with values 20, 15, 12, the first bidder wins but has to pay 24. It is clear that this is not individually rational.

## 2.17 Bayesian Implementation: The dAGVA Mechanism

Recall that we mentioned two possible routes to get around the Gibbard–Satterthwaite Impossibility Theorem. The first was to focus on restricted environments like the quasilinear environment, and the second one was to weaken the implementation concept and look for an SCF which is ex-post efficient, nondictatorial, and Bayesian incentive compatible. In this section, our objective is to explore the second route.

Throughout this section, we will once again be working with the quasilinear environment. As we saw earlier, the quasilinear environments have a nice property that every social choice function in these environments is nondictatorial. Therefore, while working within a quasilinear environment, we do not have to worry about the nondictatorial part of the social choice function. We can just investigate whether there exists any SCF in quasilinear environment, which is both *ex-post efficient* and *BIC*, or equivalently, which has three properties — *AE, BB,* and *BIC*. Recall that in the previous section, we have already addressed the question whether there exists any SCF in quasilinear environments that is *AE, BB*, and *DSIC*, and we found that no function satisfies all these three properties. On the contrary, in this section, we will show that a wide range of SCFs in quasilinear environments satisfy three properties — AE, BB, and BIC.

### *2.17.1 The dAGVA Mechanism*

The following theorem, due to d'Aspremont and Gérard-Varet [22] and Arrow [23] confirms that in quasilinear environments, there exist social choice functions that are both ex-post efficient and Bayesian incentive compatible. We refer to this theorem as the *dAGVA* Theorem.

**Theorem 2.12 (The dAGVA Theorem).** *Let the social choice function $f(\cdot) = (k^*(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ be allocatively efficient and the agents' types be statistically independent of each other (i.e. the density $\phi(\cdot)$ has the form $\phi_1(\cdot) \times \ldots \times \phi_n(\cdot)$). This function can be truthfully implemented in Bayesian Nash equilibrium if it satisfies the following payment structure, known as the dAGVA payment (incentive) scheme:*

$$t_i(\theta) = E_{\tilde{\theta}_{-i}}\left[\sum_{j \neq i} v_j(k^*(\theta_i, \tilde{\theta}_{-i}), \tilde{\theta}_j)\right] + h_i(\theta_{-i}) \;\; \forall\, i = 1, \ldots, n; \quad \forall \theta \in \Theta \quad (2.23)$$

*where $h_i(\cdot)$ is any arbitrary function of $\theta_{-i}$.*

**Proof:** Let the social choice function $f(\cdot) = (k^*(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ be allocatively efficient, i.e., it satisfies the condition (2.17) and also satisfies the dAGVA payment scheme (2.23). Consider

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] = E_{\theta_{-i}}[v_i(k^*(\theta_i, \theta_{-i}), \theta_i) + t_i(\theta_i, \theta_{-i})|\theta_i].$$

Since $\theta_i$ and $\theta_{-i}$ are statistically independent, the expectation can be taken without conditioning on $\theta_i$. This will give us

$$\begin{aligned} E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] &= E_{\theta_{-i}}\left[v_i(k^*(\theta_i, \theta_{-i}), \theta_i) + h_i(\theta_{-i}) + E_{\tilde{\theta}_{-i}}\left[\sum_{j \neq i} v_j(k^*(\theta_i, \tilde{\theta}_{-i}), \tilde{\theta}_j)\right]\right] \\ &= E_{\theta_{-i}}\left[\sum_{j=1}^{n} v_j(k^*(\theta_i, \theta_{-i}), \theta_j)\right] + E_{\theta_{-i}}[h_i(\theta_{-i})]. \end{aligned}$$

Since $k^*(\cdot)$ satisfies the condition (2.17),

$$\sum_{j=1}^{n} v_j(k^*(\theta_i, \theta_{-i}), \theta_j) \geq \sum_{j=1}^{n} v_j(k^*(\hat{\theta}_i, \theta_{-i}), \theta_j) \;\forall\; \hat{\theta}_i \in \Theta_i.$$

Thus we get, $\forall \;\hat{\theta}_i \in \Theta_i$

$$E_{\theta_{-i}}\left[\sum_{j=1}^{n} v_j(k^*(\theta_i, \theta_{-i}), \theta_j)\right] + E_{\theta_{-i}}[h_i(\theta_{-i})] \geq E_{\theta_{-i}}\left[\sum_{j=1}^{n} v_j(k^*(\hat{\theta}_i, \theta_{-i}), \theta_j)\right] + E_{\theta_{-i}}[h_i(\theta_{-i})].$$

Again by making use of statistical independence we can rewrite the above inequality in the following form

$$E_{\theta_{-i}}[u_i(f(\theta_i, \theta_{-i}), \theta_i)|\theta_i] \geq E_{\theta_{-i}}\left[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i\right] \;\forall\; \hat{\theta}_i \in \Theta_i.$$

This shows that when agents $j \neq i$ announce their types truthfully, agent $i$ finds that truth revelation is his optimal strategy, thus proving that the SCF is BIC.

*Q.E.D.*

After the results of d'Aspremont and Gérard-Varet [22] and Arrow [23], a direct revelation mechanism in which the SCF is allocatively efficient and satisfies the dAGVA payment scheme is called as *dAGVA mechanism/expected externality mechanism/expected Groves mechanism*.

**Definition 2.45 (dAGVA/expected externality/expected Groves Mechanisms).** A direct revelation mechanism, $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ in which $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ satisfies (2.17) and (2.23) is known as dAGVA/expected externality/expected Groves Mechanism.[2]

## 2.17.2 The dAGVA Mechanism and Budget Balance

We now show that the functions $h_i(\cdot)$ above can be chosen to guarantee $\sum_{i=1}^{n} t_i(\theta) = 0$. Let us define,

$$\xi_i(\theta_i) = E_{\tilde{\theta}_{-i}} \left[ \sum_{j \neq i} v_j(k^*(\theta_i, \tilde{\theta}_{-i}), \tilde{\theta}_j) \right] \; \forall \, i = 1, \ldots, n$$

$$h_i(\theta_{-i}) = -\left( \frac{1}{n-1} \right) \sum_{j \neq i} \xi_j(\theta_j) \; \forall \, i = 1, \ldots, n.$$

In view of the above definitions, we can say that

$$\begin{aligned} t_i(\theta) &= \xi_i(\theta_i) - \left( \frac{1}{n-1} \right) \sum_{j \neq i} \xi_j(\theta_j) \\ \Rightarrow \sum_{i=1}^{n} t_i(\theta) &= \sum_{i=1}^{n} \xi_i(\theta_i) - \left( \frac{1}{n-1} \right) \sum_{i=1}^{n} \sum_{j \neq i} \xi_j(\theta_j) \\ \Rightarrow \sum_{i=1}^{n} t_i(\theta) &= \sum_{i=1}^{n} \xi_i(\theta_i) - \left( \frac{1}{n-1} \right) \sum_{i=1}^{n} (n-1) \xi_j(\theta_j) \\ \Rightarrow \sum_{i=1}^{n} t_i(\theta) &= 0. \end{aligned}$$

The budget balanced payment structure of the agents in the above mechanism can be given a nice graph theoretic interpretation. Imagine a directed graph $G = (V, A)$ where $V$ is the set of $n+1$ vertices, numbered $0, 1, \ldots, n$, and $A$ is the set of $[n + n(n-1)]$ directed arcs. The vertices starting from 1 through $n$ correspond to the $n$

[2] *We will sometimes abuse the terminology and simply refer to a SCF $f(\cdot)$ satisfying (2.17) and (2.23) as dAGVA/expected externality/expected Groves Mechanism.*

agents involved in the system and the vertex number 0 corresponds to the social planner. The set $A$ consists of two types of the directed arcs:

1. Arcs $0 \rightarrow i \ \forall i = 1, \ldots, n$,
2. Arcs $i \rightarrow j \ \forall i, j \in \{1, 2, \ldots, n\}; i \neq j$.

Each of the arcs $0 \rightarrow i$ carries a flow of $t_i(\theta)$ and each of the arcs $i \rightarrow j$ carries a flow of $\frac{\xi_i(\theta_i)}{n-1}$. Thus the total outflow from a node $i \in \{1, 2, \ldots, n\}$ is $\xi_i(\theta_i)$ and total inflow to the node $i$ from nodes $j \in \{1, 2, \ldots, n\}$ is $-h_i(\theta_{-i}) = \left(\frac{1}{n-1}\right) \sum_{j \neq i} \xi_j(\theta_j)$ Thus for any node $i$, $t_i(\theta) + h_i(\theta_{-i})$ is the net outflow which it is receiving from node 0 in order to respect the flow conservation constraint. Thus, if $t_i(\cdot)$ is positive then the agent $i$ receives the money from the social planner and if it is negative, then the agent pays the money to the social planner. However, by looking at flow conservation equation for node 0, we can say that total payment received by the planner from the agents and total payment made by the planner to the agents will add up to zero. In graph theoretic terms, the flow from node $i$ to node $j$ can be justified as follows. Each agent $i$ first evaluates the expected total valuation that would be generated together by all his rival agents in his absence, which turns out to be $\xi_i(\theta_i)$. Now, agent $i$ divides it equally among the rival agents and pays to every rival agent an amount equivalent to this. The idea can be better understood with the help of Figure 2.15, which depicts the three agents case.

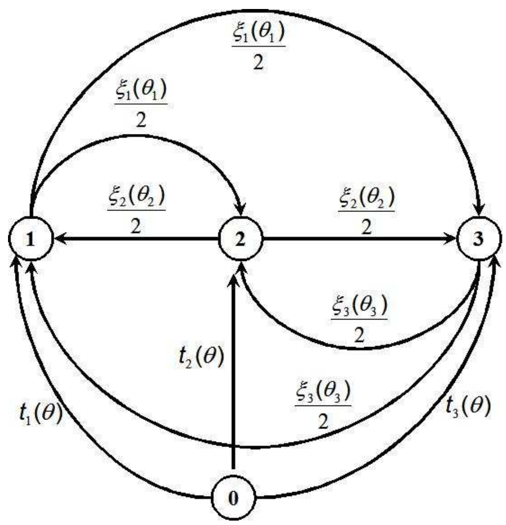



**Fig. 2.15** Payment structure showing budget balance in the expected externality mechanism

*Example 2.59 (dAGVA Mechanism for Sealed Bid Auction).* Consider a selling agent 0 and two buying agents 1, 2. The buying agents submit sealed bids to buy a single indivisible item. Let $\theta_1$ and $\theta_2$ be the willingness to pay of the buyers. Let us define the usual allocation function:

$$\begin{aligned} y_1(\theta_1,\theta_2) &= 1 \text{ if } \theta_1 \geq \theta_2 \\ &= 0 \text{ else} \\ y_2(\theta_1,\theta_2) &= 1 \text{ if } \theta_1 < \theta_2 \\ &= 0 \text{ else.} \end{aligned}$$

Let $\Theta_1 = \Theta_2 = [0,1]$ and assume that the bids from the bidders are i.i.d. uniform distributions on [0, 1]. Also assume that $\Theta_0 = \{0\}$. Assuming that the dAGVA mechanism is used, the payments can be computed as follows:

$$t_i(\theta_1,\theta_2) = E_{\theta_{-i}}\left[\sum_{j\neq i} v_j(k(\theta),\theta_j)\right] - \frac{1}{2}\left[\sum_{j\neq i} E_{\theta_{-i}}\left\{\sum_{l\neq j} v_l(k(\theta),\theta_j)\right\}\right].$$

It can be shown that

$$\begin{aligned} t_1(\theta) &= -\left(\frac{1}{12} - \frac{\theta_1}{2} + \frac{\theta_2}{2}\right) y_1(\theta) \\ t_2(\theta) &= -\left(\frac{1}{12} - \frac{\theta_2}{2} + \frac{\theta_1}{2}\right) y_2(\theta) \\ t_0(\theta) &= -(t_1(\theta) + t_2(\theta)) \end{aligned}$$

This can be compared to the first price auction in which case

$$\begin{aligned} t_1(\theta) &= -\frac{\theta_1}{2} y_1(\theta) \\ t_2(\theta) &= -\frac{\theta_2}{2} y_2(\theta). \end{aligned}$$

Also, one can compare with the second price auction, where

$$\begin{aligned} t_1(\theta) &= -\theta_2 y_1(\theta) \\ t_2(\theta) &= -\theta_1 y_2(\theta). \end{aligned}$$

### *2.17.3 The Myerson–Satterthwaite Theorem*

We have so far not seen a single example where we have all the desired properties in an SCF: AE, BB, BIC, and IR. This provides a motivation to study the feasibility of having all these properties in a social choice function.

The Myerson–Satterthwaite Theorem is a disappointing news in this direction, since it asserts that in a bilateral trade setting, whenever the gains from the trade are possible but not certain, then there is no SCF that satisfies AE, BB, BIC, and Interim IR all together. The precise statement of the theorem is as follows.

**Theorem 2.13 (Myerson–Satterthwaite Impossibility Theorem).** *Consider a bilateral trade setting in which the buyer and seller are risk neutral, the valuations* $\theta_1$ *and* $\theta_2$ *are drawn independently from the intervals* $[\underline{\theta_1}, \overline{\theta_1}] \subset \mathbb{R}$ *and* $[\underline{\theta_2}, \overline{\theta_2}] \subset \mathbb{R}$ *with strict positive densities, and* $(\underline{\theta_1}, \overline{\theta_1}) \bigcap (\underline{\theta_2}, \overline{\theta_2}) \neq \emptyset$*. Then there is no Bayesian incentive compatible social choice function that is ex-post efficient and gives every buyer type and every seller type nonnegative expected gains from participation.*

For a proof of the above theorem, refer to Proposition 23.E.1 of [6].

## *2.17.4 Mechanism Design Space in Quasilinear Environment*

Figure 2.16 shows the space of mechanisms taking into account all the results we have studied so far. A careful look at the diagram suggests why designing a mechanism that satisfies a certain combination of properties is quite intricate.

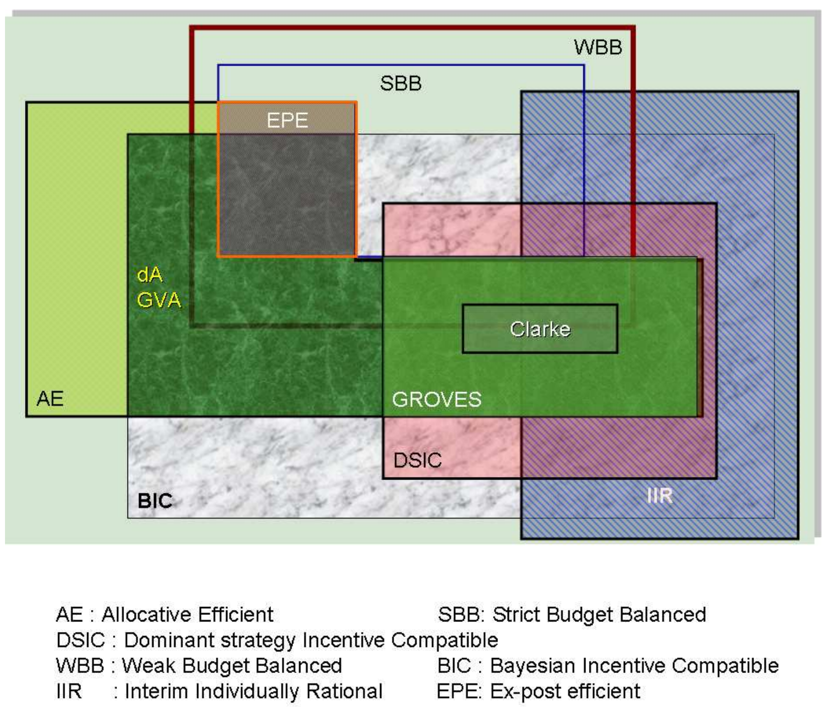



**Fig. 2.16** Mechanism design space in quasilinear environment

## 2.18 Bayesian Incentive Compatibility in Linear Environment

The linear environment is a special, but often-studied, subclass of the quasilinear environment. This environment is a restricted version of the quasilinear environment in the following sense.

1. Each agent $i$'s type lies in an interval $\Theta_i = [\underline{\theta}_i, \overline{\theta_i}] \subset \mathbb{R}$ with $\underline{\theta}_i < \overline{\theta_i}$.
2. Agents' types are statistically independent, that is, the density $\phi(\cdot)$ has the form $\phi_1(\cdot) \times \ldots \times \phi_n(\cdot)$.
3. $\phi_i(\theta_i) > 0 \; \forall \; \theta_i \in [\underline{\theta}_i, \overline{\theta_i}] \; \forall \; i = 1, \ldots, n$.
4. Each agent $i$'s utility function takes the following form

$$u_i(x, \theta_i) = \theta_i v_i(k) + m_i + t_i.$$

The linear environment has very interesting properties in terms of being able to obtain a characterization of the class of BIC social choice functions. Before we present Myerson's Characterization Theorem for BIC social choice functions in a linear environment, we would like to define the following quantities with respect to any social choice function $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ in this environment.

- Let $\overline{t_i}(\hat{\theta}_i) = E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})]$ be agent $i$'s expected transfer given that he announces his type to be $\hat{\theta}_i$ and that all agents $j \neq i$ truthfully reveal their types.
- Let $\overline{v_i}(\hat{\theta}_i) = E_{\theta_{-i}}[v_i(\hat{\theta}_i, \theta_{-i})]$ be agent $i$'s expected "benefits" given that he announces his type to be $\hat{\theta}_i$ and that all agents $j \neq i$ truthfully reveal their types.
- Let $U_i(\hat{\theta}_i|\theta_i) = E_{\theta_{-i}}[u_i(f(\hat{\theta}_i, \theta_{-i}), \theta_i)|\theta_i]$ be agent $i$'s expected utility when his type is $\theta_i$, he announces his type to be $\hat{\theta}_i$, and that all agents $j \neq i$ truthfully reveal their types. It is easy to verify from the previous two definitions that

$$U_i(\hat{\theta}_i|\theta_i) = \theta_i \overline{v_i}(\hat{\theta}_i) + \overline{t_i}(\hat{\theta}_i).$$

- Let $U_i(\theta_i) = U_i(\theta_i|\theta_i)$ be the agent $i$'s expected utility conditional on his type being $\theta_i$ when he and all other agents report their true types. It is easy to verify that

$$U_i(\theta_i) = \theta_i \overline{v_i}(\theta_i) + \overline{t_i}(\theta_i).$$

With the above discussion as a backdrop, we now present Myerson's [24] theorem for characterizing the BIC social choice functions in this environment.

**Theorem 2.14 (Myerson's Characterization Theorem).** *In linear environment, a social choice function $f(\cdot) = (k(\cdot), t_1(\cdot), \ldots, t_n(\cdot))$ is BIC if and only if, for all $i = 1, \ldots, n$,*

1. *$\overline{v_i}(\cdot)$ is nondecreasing,*
2. *$U_i(\theta_i) = U_i(\underline{\theta_i}) + \int_{\underline{\theta_i}}^{\theta_i} \overline{v_i}(s)ds \;\; \forall \; \theta_i$.*

For a proof of the above theorem, refer to Proposition 23.D.2 of [6]. The above theorem shows that to identify all BIC social choice functions in a linear environment, we can proceed as follows: First identify which functions $k(\cdot)$ lead every agent $i$'s expected benefit function $\overline{v_i}(\cdot)$ to be nondecreasing. Then, for each such function identify transfer functions $\overline{t_1}(\cdot),\ldots,\overline{t_n}(\cdot)$ that satisfy the second condition of the above proposition. Substituting for $U_i(\cdot)$ in the second condition above, we get that expected transfer functions are precisely those that satisfy, for $i = 1,\ldots,n$,

$$\overline{t_i}(\theta_i) = \overline{t_i}(\underline{\theta_i}) + \underline{\theta_i}\overline{v_i}(\underline{\theta_i}) - \theta_i\overline{v_i}(\theta_i) + \int_{\underline{\theta_i}}^{\theta_i} \overline{v_i}(s)ds$$

for some constant $\overline{t_i}(\underline{\theta_i})$. Finally, choose any set of transfer functions $t_1(\cdot),\ldots,t_n(\cdot)$ such that $E_{\theta_{-i}}[t_i(\theta_i,\theta_{-i})] = \overline{t_i}(\theta_i)$ for all $\theta_i$. In general, there are many such functions, $t_i(\cdot,\cdot)$; one, for example, is simply $t_i(\theta_i,\theta_{-i}) = \overline{t_i}(\theta_i)$.

In what follows we discuss two examples where the environment is linear and analyze the BIC property of the social choice function by means of Myerson's Characterization Theorem.

*Example 2.60 (First-Price Sealed Bid Auction in Linear Environment).* Consider the first-price sealed bid auction. Let us assume that $S_i = \Theta_i = \left[\underline{\theta_i},\overline{\theta_i}\right] \quad \forall\, i \in N$. In such a case, the first-price auction becomes a direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i\in N}, f(\cdot))$, where $f(\cdot)$ is an SCF that is the same as the outcome rule of the first-price auction. Let us impose the additional conditions on the environment to make it linear. We assume that

1. Bidders' types are statistically independent, that is, the density $\phi(\cdot)$ has the form $\phi_1(\cdot)\times\ldots\times\phi_n(\cdot)$
2. Let each bidder draw his type from the set $[\underline{\theta_i},\overline{\theta_i}]$ by means of a uniform distribution, that is $\phi_i(\theta_i) = 1/(\overline{\theta_i} - \underline{\theta_i})\ \forall\, \theta_i \in [\underline{\theta_i},\overline{\theta_i}]\ \ \forall\, i = 1,\ldots,n$.

Note that the utility functions of the agents in this example are given by

$$u_i(f(\theta),\theta_i) = \theta_i y_i(\theta) + t_i(\theta)\ \forall\, i = 1,\ldots,n.$$

Thus, observing $y_i(\theta) = v_i(k(\theta))$ will confirm that these utility functions also satisfy the fourth condition required for a linear environment. Now we can apply Myerson's Characterization Theorem to test the Bayesian incentive compatibility of the SCF involved here. It is easy to see that for any bidder $i$, we have

$$\begin{aligned}\overline{v_i}(\theta_i) &= E_{\theta_{-i}}[v_i(\theta_i,\theta_{-i})]\\ &= E_{\theta_{-i}}[y_i(\theta_i,\theta_{-i})]\\ &= 1.P\left\{(\theta_{-i})_{(n-1)} \le \theta_i\right\} + 0.\left(1 - P\left\{\theta_i < (\theta_{-i})_{(n-1)}\right\}\right)\\ &= P\left\{(\theta_{-i})_{(n-1)} \le \theta_i\right\} \qquad (2.24)\end{aligned}$$

where $P\left\{(\theta_{-i})_{(n-1)} \le \theta_i\right\}$ is the probability that the given type $\theta_i$ of the bidder $i$ is the highest among all the bidders' types. This implies that in the presence of the independence assumptions made above, $\overline{v_i}(\theta_i)$ is a nondecreasing function.

We know that for a first-price sealed bid auction, $t_i(\theta) = -\theta_i y_i(\theta)$. Therefore, we can claim that for a first-price sealed bid auction, we have

$$\overline{t_i}(\theta_i) = -\theta_i \overline{v_i}(\theta_i) \ \forall \ \theta_i \in \Theta_i.$$

The above values of $\overline{v_i}(\theta_i)$ and $\overline{t_i}(\theta_i)$ can be used to compute $U_i(\theta_i)$ in the following manner:

$$U_i(\theta_i) = \theta_i \overline{v_i}(\theta_i) + \overline{t_i}(\theta_i) = 0 \ \ \forall \ \theta_i \in [\underline{\theta_i}, \overline{\theta_i}]. \tag{2.25}$$

The above equation can be used to test the second condition of Myerson's Theorem, which requires

$$U_i(\theta_i) = U_i(\underline{\theta_i}) + \int_{\underline{\theta_i}}^{\theta_i} \overline{v_i}(s) ds.$$

In view of Equations (2.24) and (2.25), it is easy to see that this second condition of Myerson's Characterization Theorem is not being met by the SCF used in the first-price sealed bid auction. Therefore, we can finally claim that a first-price sealed bid auction is not BIC in linear environment.

*Example 2.61 (Second-Price Sealed Bid Auction in a Linear Environment).* Consider the second-price sealed bid auction. Let us assume that $S_i = \Theta_i = \left[\underline{\theta_i}, \overline{\theta_i}\right] \ \ \forall \ i \in N$. In such a case, the second-price auction becomes a direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$, where $f(\cdot)$ is an SCF that is the same as the outcome rule of the second-price auction. We have already seen that this SCF $f(\cdot)$ is DSIC in quasilinear environment, and a linear environment is a special case of a quasilinear environment; therefore, it is DSIC in the linear environment also. Moreover, we know that DSIC implies BIC. Therefore, we can directly claim that the SCF used in the second-price auction is BIC in a linear environment.

## 2.19 Revenue Equivalence Theorem

In this section, we prove two results that show the revenue equivalence of certain classes of auctions. The first theorem is a general result that shows the revenue equivalence of two auctions that satisfy certain conditions. The second result is a more specific result that shows the revenue equivalence of four different types of auctions (English auction, Dutch auction, first price auction, and second price auction) in the special context of an auction of a single indivisible item. The proof of the second result crucially uses the first result.

### *2.19.1 Revenue Equivalence of Two Auctions*

Assume that $y_i(\theta)$ is the probability of agent $i$ getting the object when the vector of announced types is $\theta = (\theta_1, \ldots, \theta_n)$. The expected payoff to the buyer $i$ with a type profile $\theta = (\theta_1, \cdots, \theta_n)$ will be $y_i(\theta)\theta_i + t_i(\theta)$. The set of allocations is given by

$$K = \left\{ (y_1, \cdots, y_n) : y_i \in [0,1] \forall i = 1, \cdots, n; \sum_{i=1}^{n} y_i \leq 1 \right\}.$$

As earlier, let $\overline{y_i}(\hat{\theta}_i) = E_{\theta_{-i}}[y_i(\hat{\theta}_i, \theta_{-i})]$ be the probability that agent $i$ gets the object conditional to announcing his type as $\hat{\theta}_i$, with the rest of the agents announcing their types truthfully. Similarly, $\overline{t_i}(\hat{\theta}_i) = E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})]$ denotes the expected payment received by agent $i$ conditional to announcing his type as $\hat{\theta}_i$, with the rest of the agents announcing their types truthfully. Let $\overline{v_i}(\hat{\theta}_i) = \overline{y_i}(\hat{\theta}_i)$. Then,

$$U_i(\theta_i) = \overline{y_i}(\theta_i)\theta_i + \overline{t_i}(\theta_i)$$

denotes the payoff to agent $i$ when all the buying agents announce their types truthfully. We now state and prove an important proposition.

**Theorem 2.15.** *Consider an auction scenario with:*

1. *n risk-neutral bidders (buyers)* $1, 2, \cdots, n$
2. *The valuation of bidder i* $(i = 1, \cdots, n)$ *is a real interval* $[\underline{\theta_i}, \overline{\theta_i}] \subset \mathbb{R}$ *with* $\underline{\theta_i} < \overline{\theta_i}$.
3. *The valuation of bidder i* $(i = 1, \cdots, n)$ *is drawn from* $[\underline{\theta_i}, \overline{\theta_i}]$ *with a strictly positive density* $\phi_i(.) > 0$. *Let* $\Phi_i(.)$ *be the cumulative distribution function.*
4. *The bidders' types are statistically independent.*

*Suppose that a given pair of Bayesian Nash equilibria of two different auction procedures are such that:*

- *For every bidder i, for each possible realization of* $(\theta_1, \cdots, \theta_n)$*, bidder i has an identical probability of getting the good in the two auctions.*
- *Every bidder i has the same expected payoff in the two auctions when his valuation for the object is at its lowest possible level.*

*Then the two auctions generate the same expected revenue to the seller.*

Before proving the theorem, we elaborate on the first assumption above, namely risk neutrality. A bidder is said to be:

- *risk-averse* if his utility is a *concave* function of his wealth; that is, an increment in the wealth at a lower level of wealth leads to an increment in utility that is higher than the increase in utility due to an identical increment in wealth at a higher level of wealth;

- *risk-loving* if his utility is a *convex* function of his wealth; that is, an increment in the wealth at a lower level of wealth leads to an increment in utility that is lower than the increase in utility due to an identical increment in wealth at a higher level of wealth; and
- *risk-neutral* if his utility is a *linear* function of his wealth; that is, an increment in the wealth at a lower level of wealth leads to the same increment in the utility as an identical increment would yield at a higher level of wealth.

**Proof:** By the revelation principle, it is enough that we investigate two Bayesian incentive compatible social choice functions in this auction setting. It is enough that we show that two Bayesian incentive compatible social choice functions having (a) the same allocation functions $(y_1(\theta), \cdots, y_n(\theta))\ \forall \theta \in \Theta$, and (b) the same values of $U_1(\underline{\theta_1}), \cdots, U_n(\underline{\theta_n})$ will generate the same expected revenue to the seller.

We first derive an expression for the seller's expected revenue given any Bayesian incentive compatible mechanism. Expected revenue to the seller

$$= \sum_{i=1}^{n} E_\theta[-t_i(\theta)]. \tag{2.26}$$

Now, we have:

$$\begin{aligned}
E_\theta[-t_i(\theta)] &= E_{\theta_i}[-E_{\theta_{-i}}[t_i(\theta)]] \\
&= \int_{\underline{\theta_i}}^{\overline{\theta_i}} [\overline{y_i}(\theta_i)\theta_i - U_i(\theta_i)]\phi_i(\theta_i)d\theta_i \\
&= \int_{\underline{\theta_i}}^{\overline{\theta_i}} \left[ [\overline{y_i}(\theta_i)\theta_i - U_i(\underline{\theta_i})] - \int_{\underline{\theta_i}}^{\theta_i} \overline{y_i}(s)ds \right] \phi_i(\theta_i)d\theta_i.
\end{aligned}$$

The last step is an implication of Myerson's characterization of Bayesian incentive compatible functions in linear environment. The above expression is now equal to

$$= \left[ \int_{\underline{\theta_i}}^{\overline{\theta_i}} \left( \overline{y_i}(\theta_i)\theta_i - \int_{\underline{\theta_i}}^{\theta_i} \overline{y_i}(s)ds \right) \phi_i(\theta_i)d\theta_i \right] - U_i(\underline{\theta_i}).$$

Now, applying integration by parts with $\int_{\underline{\theta_i}}^{\theta_i} \overline{y_i}(s)ds$ as the first function, we get

$$\int_{\underline{\theta_i}}^{\overline{\theta_i}} \left( \int_{\underline{\theta_i}}^{\theta_i} \overline{y_i}(s)ds \right) \phi_i(\theta_i)d\theta_i$$

$$\begin{aligned}
&= \int_{\underline{\theta_i}}^{\overline{\theta_i}} \overline{y_i}(\theta_i)d\theta_i - \int_{\underline{\theta_i}}^{\overline{\theta_i}} \overline{y_i}(\theta_i)\Phi_i(\theta_i)d\theta_i \\
&= \int_{\underline{\theta_i}}^{\overline{\theta_i}} \overline{y_i}(\theta_i)[1 - \Phi_i(\theta_i)]d\theta_i.
\end{aligned}$$

Therefore we get

$$E_{\theta_i}[-\overline{t_i}(\theta_i)] = -U_i(\underline{\theta_i}) + \left[ \int_{\underline{\theta_i}}^{\overline{\theta_i}} \overline{y_i}(\theta_i) \left\{ \theta_i - \frac{1-\Phi_i(\theta_i)}{\phi_i(\theta_i)} \right\} \phi_i(\theta_i) d\theta_i \right]$$

$$= -U_i(\underline{\theta_i}) + \left[ \int_{\underline{\theta_1}}^{\overline{\theta_1}} \cdots \int_{\underline{\theta_n}}^{\overline{\theta_i}} y_i(\theta_1, \cdots, \theta_n) \right.$$

$$\left. \times \left( \theta_i - \frac{1-\Phi_i(\theta_i)}{\phi_i(\theta_i)} \right) \left( \prod_{j=1}^{n} \phi_j(\theta_j) \right) d\theta_n \cdots d\theta_1 \right]$$

since

$$\overline{y_i}(\theta_i) = \int_{\underline{\theta_1}}^{\overline{\theta_1}} \cdots \int_{\underline{\theta_n}}^{\overline{\theta_n}} y_i(\theta_1, \cdots, \theta_n) \quad \underbrace{d\theta_n \cdots d\theta_1}_{\text{without } d\theta_i}.$$

Therefore the expected revenue of the seller

$$= \left[ \int_{\underline{\theta_1}}^{\overline{\theta_1}} \cdots \int_{\underline{\theta_n}}^{\overline{\theta_n}} \sum_{i=1}^{n} y_i(\theta_1, \cdots, \theta_n) \left( \theta_i - \frac{1-\Phi_i(\theta_i)}{\phi_i(\theta_i)} \right) \right] \left( \prod_{j=1}^{n} \phi_j(\theta_j) \right) d\theta_n \cdots d\theta_1$$
$$- \sum_{i=1}^{n} U_i(\underline{\theta_i}).$$

By looking at the above expression, we see that any two Bayesian incentive compatible social choice functions that generate the same functions $(y_1(\theta), \cdots, y_n(\theta))$ and the same values of $(U_1(\underline{\theta_1}), \cdots, U_n(\underline{\theta_n}))$ generate the same expected revenue to the seller.

As an application of the above theorem, we now state and prove a revenue equivalence theorem for a single indivisible item auction. The article by McAfee and McMillan [25] is an excellent reference for this topic and a part of the discussion in this section is inspired by this article.

### *2.19.2 Revenue Equivalence of Four Classical Auctions*

There are four basic types of auctions when a single indivisible item is to be sold:

1. **English auction**: This is also called oral auction, open auction, open cry auction, and ascending bid auction. Here, the price starts at a low level and is successively raised until only one bidder remains in the fray. This can be done in several ways: (a) an auctioneer announces prices, (b) bidders call the bids themselves, or (b) bids are submitted electronically. At any point of time, each bidder knows the level of the current best bid. The winning bidder pays the latest going price.
2. **Dutch auction**: This is also called a descending bid auction. Here, the auctioneer announces an initial (high) price and then keeps lowering the price iteratively until one of the bidders accepts the current price. The winner pays the current price.

3. **First price sealed bid auction**: Recall that in this auction, potential buyers submit sealed bids and the highest bidder is awarded the item. The winning bidder pays the price that he has bid.
4. **Second price sealed bid auction**: This is the classic Vickrey auction. Recall that potential buyers submit sealed bids and the highest bidder is awarded the item. The winning bidder pays a price equal to the second highest bid (which is also the highest losing bid).

When a single indivisible item is to be bought or procured, the above four types of auctions can be used in a *reverse* way. These are called *reverse auctions* or *procurement auctions*. In this section, we would be discussing the Revenue Equivalence Theorem as it applies to selling. The procurement version can be analyzed on similar lines.

#### 2.19.2.1 The Benchmark Model

There are four assumptions underlying the derivation of the Revenue Equivalence Theorem: (1) risk neutrality of bidders; (2) bidders have independent private values; (3) bidders are symmetric; (4) payments depend on bids alone. These are described below in more detail.

**(1) Risk Neutrality of Bidders**

It is assumed in the benchmark model that all the bidders are risk neutral. This immediately implies that the utility functions are linear.

**(2) Independent Private Values Model**

In the independent private values model, each bidder knows precisely how much he values the item. He has no doubt about the true value of the item to him. However, each bidder does not know anyone else's valuation of the item. Instead, he perceives any other bidder's valuation as a draw from some known probability distribution. Also, each bidder knows that the other bidders and the seller regard his own valuation as being drawn from some probability distribution. More formally, let $N = \{1, 2, \cdots, n\}$ be the set of bidders. There is a probability distribution $\Phi_i$ from which bidder $i$ draws his valuation $v_i$. Only bidder $i$ observes his own valuation $v_i$, but all other bidders and the seller know the distribution $\Phi_i$. Any one bidder's valuation is statistically independent from any other bidder's valuation.

An apt example of this assumption is provided by the auction of an antique in which the bidders are consumers buying for their own use and not for resale. Another example is government contract bidding when each bidder knows his own production cost if he wins the contract.

A contrasting model is the *common value model*. Here, if $V$ is the unobserved true value of the item, then the bidders' perceived values $v_i, i = 1, 2, \cdots, n$ are independent draws from some probability distribution $H(v_i|V)$. All the bidders know the distribution $H$. An example is provided by the sale of an antique that is being bid for by dealers who intend to resell it. The item has one single objective value, namely its market price. However, no one knows the true value. The bidders, perhaps having access to different information, have different guesses about how much the item is objectively worth. Another example is that of the sale of mineral rights to a particular tract of land. The objective value here is the amount of mineral actually lying beneath the ground. However no one knows its true value.

Suppose a bidder were somehow to learn another bidder's valuation. If the situation is described by the common value model, then the above provides useful information about the likely true value of the item, and the bidder would probably change his own valuation in light of this. If the situation is described by the independent private values model, the bidder knows his own mind, and learning about others' valuations will not cause him to change his own valuation (although he may, for strategic reasons, change his bid).

Real world auction situations are likely to contain aspects of both the independent private values model and the common value model. It is assumed in the benchmark model that the independent private values assumption holds.

**(3) Symmetry**

This assumption implies that all the bidders have the same set of possible valuations, and further they draw their valuations using the same probability density $\phi$. That is, $\phi_1 = \phi_2 = \ldots = \phi_n = \phi$.

**(4) Dependence of Payments on Bids Alone**

It is assumed that the payment to be made by the winner to the auctioneer is a function of bids alone.

**Theorem 2.16 (Revenue Equivalence Theorem for Single Indivisible Item Auctions).** *Consider a seller or an auctioneer trying to sell a single indivisible item in which n bidders are interested. For the benchmark model (bidders are risk neutral, bidders have independent private values, bidders are symmetric, and payments depend only on bids), all the four basic auction types (English auction, Dutch auction, first price auction, and second price auction) yield the same average revenue to the seller.*

The result looks counter intuitive: For example, it might seem that receiving the highest bid in a first price sealed bid auction must be better for the seller than receiving the second highest bid, as in second price auction. However, it is to be noted that bidders act differently in different auction situations. In particular, they bid more aggressively in a second price auction than in a first price auction.

**Proof**: The proof proceeds in three parts. In Part 1, we show that the first price auction and the second price auction yield the same expected revenue in their respective equilibria. In Part 2, we show that the Dutch auction and the first price auction produce the same outcome. In Part 3, we show that the English auction and the second price auction yield the same outcome.

### Part 1: Revenue Equivalence of First Price Auction and Second Price Auction

The first price auction and the second price auction satisfy the conditions of the theorem on revenue equivalence of two auctions.

- In both the auctions, the bidder with the highest valuation wins the auction.
- bidders' valuations are drawn from $[\underline{\theta_i}, \overline{\theta_i}]$ and a bidder with valuation at the lower limit of the interval has a payoff of zero in both the auctions.

Thus the theorem can be applied to the equilibria of the two auctions: Note that in the case of the first price auction, it is a Bayesian Nash equilibrium while in the case of the second price auction, it is a weakly dominant strategy equilibrium. In fact, it can be shown in any *symmetric* auction setting (where the bidders' valuations are independently drawn from identical distributions) that the conditions of the above proposition will be satisfied by any Bayesian Nash equilibrium of the first price auction and the weakly dominant strategy equilibrium of the second price scaled bid auction.

### Part 2: Revenue Equivalence of Dutch Auction and First Price Auction

To see this, consider the situation facing a bidder in these two auctions. In each case, the bidder must choose how high to bid without knowing the other bidders' decisions. If he wins, the price he pays equals his own bid. This result is true irrespective of which of the assumptions in the benchmark model apply. Note that the equilibrium in the underlying Bayesian game in the two cases here is a Bayesian Nash equilibrium.

### Part 3: Revenue Equivalence of English Auction and Second Price Auction

First we analyze the English auction. Note that a bidder drops out as soon the going price exceeds his valuation. The second last bidder drops out as soon as the price exceeds his own valuation. This leaves only one bidder in the fray and he wins the auction. Note that the winning bidder's valuation is the highest among all the bidders and he earns some payoff in spite of the monopoly power of the seller. Only the winning bidder knows how much payoff he receives because only he knows his own valuation. Suppose the valuations of the $n$ bidders are $v_{(1)}, v_{(2)}, \cdots, v_{(n)}$. Since the bidders are symmetric, these valuations are draws from the same distribution and without loss of generality, assume that these are in descending order. The winning bidder gets a payoff of $v_{(1)} - v_{(2)}$.

Next we analyze the second price auction. In the second price auction, the bidder's choice of bid determines only whether or not he wins; the amount he pays if he wins is beyond his control. We have already shown that each bidder's equilibrium best response strategy is to bid his own valuation for the item. The payment here is equal to the actual valuation of the bidder with the second highest valuation. Thus the expected payment and payoff are the same in English auction and the second price auction. This establishes Part 3 and therefore proves the Revenue Equivalence Theorem.

Note that the outcomes of the English auction and the second price auction satisfy a weakly dominant strategy equilibrium. That is, each bidder has a well defined best response bid regardless of how high he believes his rivals will bid. In the second price auction, the weakly dominant strategy is to bid true valuation. In the English auction, the weakly dominant strategy is to remain in the bidding process until the price reaches the bidder's own valuation.

#### 2.19.2.2 Some Observations

We now make a few important observations.

- The theorem does not imply that the outcomes of the four auction forms are always exactly the same. They are only equal on average.
  - Note that in the English auction or the second price auction, the price exactly equals the valuation of the bidder with the second highest valuation, $v_{(2)}$. In Dutch auction or the first price auction, the price is the expectation of the second highest valuation conditional on the winning bidder's own valuation. The above two prices will be equal only by accident; however, they are equal on average.
- Bidding strategy is very simple in the English auction and the second price auction. In the former, a bidder remains in bidding until the price reaches his valuation. In the latter, he submits a sealed bid equal to his own valuation.
- On the other hand, the bidding logic is quite complex in the Dutch auction and the first price auction. Here the bidder bids some amount less than his true valuation. Exactly how much less depends upon the probability distribution of the other bidders' valuations and the number of competing bidders. Finding the Nash equilibrium bid is a non-trivial computational problem.
- The Revenue Equivalence Theorem for the single indivisible item is devoid of empirical predictions about which type of auction will be chosen by the seller in any particular set of circumstances. However when the assumptions of the benchmark model are relaxed, particular auction forms emerge as being superior.

- The variance of revenue is lower in English auction or second price auction than in Dutch auction or first price auction. Hence if the seller were risk averse, he would choose English or second price rather than Dutch or first price.

For more details on the revenue equivalence theorems, the reader is referred to the papers by Myerson [24], McAfee and McMillan [25], Klemperer [26], and the books by Milgrom [27] and Krishna [28].

## 2.20 Myerson Optimal Auction

A key problem that faces a social planner is to decide which direct revelation mechanism (or equivalently, social choice function) is *optimal* for a given problem. We now attempt to formalize the notion of optimality of social choice functions and optimal mechanisms. For this, we first define the concept of a *social utility function*.

**Definition 2.46 (Social Utility Function).** A social utility function is a function $w : \mathbb{R}^n \to \mathbb{R}$ that aggregates the profile $(u_1, \ldots, u_n) \in \mathbb{R}^n$ of individual utility values of the agents into a social utility.

Consider a mechanism design problem and a direct revelation mechanism $\mathscr{D} = ((\Theta_i)_{i \in N}, f(\cdot))$ proposed for it. Let $(\theta_1, \ldots, \theta_n)$ be the actual type profile of the agents and assume for a moment that they will all reveal their true types when requested by the planner. In such a case, the social utility that would be realized by the social planner for a type profile $\theta$ of the agents is given by:

$$w(u_1(f(\theta), \theta_1), \ldots, u_n(f(\theta), \theta_n)). \tag{2.27}$$

However, recall the implicit assumption behind a mechanism design problem, namely, that the agents are autonomous, and they would report a type as dictated by their rational behavior. Therefore, the assumption that all the agents will report their true types is not true in general. In general, rationality implies that the agents report their types according to a strategy suggested by a Bayesian Nash equilibrium $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ of the underlying Bayesian game. In such a case, the social utility that would be realized by the social planner for a type profile $\theta$ of the agents is given by

$$w(u_1(f(s^*(\theta)), \theta_1), \ldots, u_n(f(s^*(\theta)), \theta_n)). \tag{2.28}$$

In some instances, the above Bayesian Nash equilibrium may turn out to be a dominant strategy equilibrium. Better still, truth revelation by all agents could turn out to be a Bayesian Nash equilibrium or a dominant strategy equilibrium.

### 2.20.1 Optimal Mechanism Design Problem

In view of the above notion of a social utility function, it is clear that the objective of a social planner would be to look for a social choice function $f(\cdot)$ that would maximize the expected social utility for a given social utility function $w(\cdot)$. However, being the social planner, it is always expected of him to be fair to all the agents. Therefore, the social planner would first put a few desirable constraints on the set of social choice functions from which he can probably choose. The desirable constraints may include any combination of all the previously studied properties of a social choice function, such as ex-post efficiency, incentive compatibility, and individual rationality. This set of social choice functions is known as a *set of feasible social choice functions* and is denoted by $F$. Thus, the problem of a social planner can now be cast as an optimization problem where the objective is to maximize the expected social utility, and the constraint is that the social choice function must be chosen from the feasible set $F$. This problem is known as the *optimal mechanism design* problem and the solution of the problem would be social choice function $f^*(\cdot) \in F$, which is used to define the optimal mechanism $\mathscr{D}^* = ((\Theta_i)_{i\in N}, f^*(\cdot))$ for the problem that is being studied.

Depending on whether the agents are loyal or autonomous rational entities, the optimal mechanism design problem may take two different forms.

$$\underset{f(\cdot)\in F}{\text{maximize}}\; E_\theta\left[w(u_1(f(\theta),\theta_1),\ldots,u_n(f(\theta),\theta_n))\right] \tag{2.29}$$

$$\underset{f(\cdot)\in F}{\text{maximize}}\; E_\theta\left[w(u_1(f(s^*(\theta)),\theta_1),\ldots,u_n(f(s^*(\theta)),\theta_n))\right] \tag{2.30}$$

The problem (2.29) is relevant when the agents are loyal and always reveal their true types whereas the problem (2.30) is relevant when the agents are rational. At this point of time, one may ask how to define the set of feasible social choice functions $F$. There is no unique definition of this set. The set of feasible social choice functions is a subjective judgment of the social planner. The choice of the set $F$ depends on the desirable properties the social planner would wish to have in the optimal social choice function $f^*(\cdot)$. If we define

$$\begin{aligned}
F_{\text{DSIC}} &= \{f:\Theta\to X | f(\cdot) \text{ is dominant strategy incentive compatible}\}\\
F_{\text{BIC}} &= \{f:\Theta\to X | f(\cdot) \text{ is Bayesian incentive compatible}\}\\
F_{\text{EPIR}} &= \{f:\Theta\to X | f(\cdot) \text{ is ex-post individual rational}\}\\
F_{\text{IIR}} &= \{f:\Theta\to X | f(\cdot) \text{ is interim individual rational}\}
\end{aligned}$$

$$\begin{aligned}
F_{\text{EAIR}} &= \{f : \Theta \to X | f(\cdot) \text{ is ex-ante individual rational}\}\\
F_{\text{EAE}} &= \{f : \Theta \to X | f(\cdot) \text{ is ex-ante efficient}\}\\
F_{\text{IE}} &= \{f : \Theta \to X | f(\cdot) \text{ is interim efficient}\}\\
F_{\text{EPE}} &= \{f : \Theta \to X | f(\cdot) \text{ is ex post efficient}\}.
\end{aligned}$$

The set of feasible social choice functions $F$ may be either any one of the above sets or intersection of any combination of the above sets. For example, the social planner may choose $F = F_{BIC} \bigcap F_{IIR}$. In the literature, this particular feasible set is known as *incentive feasible set* due to Myerson [1]. Also, note that if the agents are loyal then the sets $F_{DSIC}$ and $F_{BIC}$ will be equal to the whole set of all the social choice functions.

### *2.20.2 Myerson's Optimal Reverse Auction*

We now consider the problem of procuring a single indivisible item from among a pool of suppliers and present Myerson's optimal auction that minimizes the expected cost of procurement subject to Bayesian incentive compatibility and interim individual rationality of all the selling agents. The classical Myerson auction [24] is for maximizing the expected revenue of a selling agent who wishes to sell an indivisible item to a set of prospective buying agents. We present it here for the reverse auction case.

Each bidder $i$'s type lies in an interval $\Theta_i = [\underline{\theta_i}, \overline{\theta_i}]$. We impose the following additional conditions on the environment.

1. The auctioneer and the bidders are risk neutral.
2. Bidders' types are statistically independent, that is, the joint density $\phi(\cdot)$ has the form $\phi_1(\cdot) \times \ldots \times \phi_n(\cdot)$.
3. $\phi_i(\cdot) > 0 \; \forall \; i = 1, \ldots, n$.
4. We generalize the outcome set $X$ by allowing a random assignment of the good. Thus, we now take $y_i(\theta)$ to be seller $i$'s probability of selling the good when the vector of announced types is $\theta = (\theta_1, \ldots, \theta_n)$. Thus, the new outcome set is given by

$$X = \Bigg\{ (y_0, y_1 \ldots, y_n, t_0, t_1, \ldots, t_n) : y_0 \in [0,1], t_0 \leq 0, y_i \in [0,1], t_i \geq 0 \; \forall \; i = 1, \ldots, n,$$
$$\sum_{i=1}^{n} y_i \leq 1; \; \sum_{i=0}^{n} t_i = 0 \Bigg\}.$$

Recall that the utility functions of the agents in this example are given by, $\forall \; i = 1, \ldots, n$,

$$u_i(f(\theta), \theta_i) = u_i(y_0(\theta), \ldots, y_n(\theta), t_0(\theta), \ldots, t_n(\theta), \theta_i) = -\theta_i y_i(\theta) + t_i(\theta).$$

Thus, viewing $y_i(\theta) = v_i(k(\theta))$ in conjunction with the second and third conditions above, we can claim that the underlying environment here is linear.

In the above example, we assume that the auctioneer (buyer) is the social planner and he is looking for an optimal direct revelation mechanism to buy the good. Myerson's [24] idea was that the auctioneer must use a social choice function that is Bayesian incentive compatible and interim individual rational and at the same time minimizes the cost to the auctioneer. Thus, in this problem, the set of feasible social choice functions is given by $F = F_{BIC} \bigcap F_{IIR}$. The objective function in this case would be to minimize the total expected cost of the buyer, which would be given by

$$E_\theta \left[ w(u_1(f(\theta), \theta_1), \ldots, u_n(f(\theta), \theta_n)) \right] = E_\theta \left[ \sum_{i=1}^{n} t_i(\theta) \right].$$

Note that in the above objective function we have used $f(\theta)$ and not $f(s^*(\theta))$. This is because in the set of feasible social choice functions we are considering only BIC social choice functions, and for these functions we have $s^*(\theta) = \theta \ \forall \ \theta \in \Theta$. Thus, the Myerson's optimal auction design problem can be formulated as the following optimization problem:

$$\underset{f(\cdot) \in F}{\text{minimize}} \ E_\theta \left[ \sum_{i=1}^{n} t_i(\theta) \right] \tag{2.31}$$

where

$$F = \{ f(\cdot) = (y_0(\cdot), y_1(\cdot), \ldots, y_n(\cdot), t_0(\cdot), t_1(\cdot), \ldots, t_n(\cdot)) : f(\cdot) \text{ is BIC and interim IR} \}.$$

We have seen Myerson's Characterization Theorem (Theorem 2.12) for BIC SCFs in linear environment. Similarly, we can say that an SCF $f(\cdot)$ in the above context would be BIC iff it satisfies the following two conditions:

1. $\overline{y_i}(\cdot)$ is nonincreasing for all $i = 1, \ldots, n$.
2. $U_i(\theta_i) = U_i(\overline{\theta_i}) + \int_{\theta_i}^{\overline{\theta_i}} \overline{y_i}(s) ds \ \forall \ \theta_i \in \Theta_i; \ \forall \ i = 1, \ldots, n.$

Also, we can invoke the definition of interim individual rationality to claim that the an SCF $f(\cdot)$ in the above context would be interim IR iff it satisfies the following conditions:

$$U_i(\theta_i) \geq 0 \ \forall \theta_i \in \Theta_i; \ \forall \ i = 1, \ldots, n$$

where

- $\overline{t_i}(\hat{\theta}_i) = E_{\theta_{-i}}[t_i(\hat{\theta}_i, \theta_{-i})]$ is bidder $i$'s expected transfer given that he announces his type to be $\hat{\theta}_i$ and that all the bidders $j \neq i$ truthfully reveal their types.

- $\overline{y_i}(\hat{\theta}_i) = E_{\theta_{-i}}[y_i(\hat{\theta}_i, \theta_{-i})]$ is the probability that object will be procured from bidder $i$ given that he announces his type to be $\hat{\theta}_i$ and all bidders $j \neq i$ truthfully reveal their types.
- $U_i(\theta_i) = -\theta_i \overline{y_i}(\theta_i) + \overline{t_i}(\theta_i)$ ( we can take unconditional expectation because types are independent).

Based on the above, problem (2.31) can be rewritten as follows:

$$\underset{(y_i(\cdot), U_i(\cdot))_{i \in N}}{\text{minimize}} \sum_{i=1}^{n} \int_{\underline{\theta_i}}^{\overline{\theta_i}} \left(\theta_i \overline{y_i}(\theta_i) + U_i(\theta_i)\right) \phi_i(\theta_i) d\theta_i \tag{2.32}$$

subject to

(i) $\overline{y_i}(\cdot)$ is nonincreasing $\forall\, i = 1, \ldots, n$

(ii) $y_i(\theta) \in [0,1], \sum_{i=1}^{n} y_i(\theta) \leq 1\ \forall i = 1, \ldots, n, \forall\, \theta \in \Theta$

(iii) $U_i(\theta_i) = U_i(\overline{\theta_i}) + \int_{\theta_i}^{\overline{\theta_i}} \overline{y_i}(s) ds\ \ \forall\, \theta_i \in \Theta_i;\ \forall\, i = 1, \ldots, n$

(iv) $U_i(\theta_i) \geq 0\ \forall\, \theta_i \in \Theta_i;\ \forall\, i = 1, \ldots, n.$

We first note that if constraint (iii) is satisfied then constraint (iv) will be satisfied iff $U_i(\overline{\theta_i}) \geq 0\ \ \forall\, i = 1, \ldots, n$. As a result, we can replace the constraint (iv) with

$$(\text{iv}')\quad U_i(\overline{\theta_i}) \geq 0\ \forall\, i = 1, \ldots, n$$

Next, substituting for $U_i(\theta_i)$ in the objective function from constraint (iii), we get

$$\sum_{i=1}^{n} \int_{\underline{\theta_i}}^{\overline{\theta_i}} \left(\theta_i \overline{y_i}(\theta_i) + U_i(\overline{\theta_i}) + \int_{\theta_i}^{\overline{\theta_i}} \overline{y_i}(s) ds\right) \phi_i(\theta_i) d\theta_i.$$

Integrating by parts the above expression, the auctioneer's problem can be written as one of choosing the $y_i(\cdot)$ functions and the values $U_1(\overline{\theta_1}), \ldots, U_n(\overline{\theta_n})$ to minimize

$$\int_{\underline{\theta_1}}^{\overline{\theta_1}} \cdots \int_{\underline{\theta_n}}^{\overline{\theta_n}} \left[\sum_{i=1}^{n} y_i(\theta_i) J_i(\theta_i)\right] \left[\prod_{i=1}^{n} \phi_i(\theta_i)\right] d\theta_n \ldots d\theta_1 + \sum_{i=1}^{n} U_i(\overline{\theta_i})$$

subject to constraints (i), (ii), and (iv'), where

$$J_i(\theta_i) = \left(\theta_i + \frac{\Phi_i(\theta_i)}{\phi_i(\theta_i)}\right).$$

It is evident that the solution must have $U_i(\overline{\theta_i}) = 0$ for all $i = 1, \ldots, n$. Hence, the auctioneer's problem reduces to choosing functions $y_i(\cdot)$ to minimize

$$\int_{\underline{\theta_1}}^{\overline{\theta_1}} \cdots \int_{\underline{\theta_n}}^{\overline{\theta_n}} \left[\sum_{i=1}^{n} y_i(\theta_i) J_i(\theta_i)\right] \left[\prod_{i=1}^{n} \phi_i(\theta_i)\right] d\theta_n \ldots d\theta_1$$

subject to constraints (i) and (ii).

Let us ignore constraint (i) for the moment. Then inspection of the above expression indicates that $y_i(\cdot)$ is a solution to this relaxed problem iff for all $i = 1, \ldots, n$, we have

$$y_i(\theta) = \begin{cases} 0 & : \quad \text{if } J_i(\theta_i) > \min\left\{\overline{\theta_0}, \min_{h \neq i} J_h(\theta_h)\right\} \\ 1 & : \quad \text{if } J_i(\theta_i) < \min\left\{\overline{\theta_0}, \min_{h \neq i} J_h(\theta_h)\right\}. \end{cases} \tag{2.33}$$

Note that $J_i(\theta_i) = \min\left\{\overline{\theta_0}, \min_{h \neq i} J_h(\theta_h)\right\}$ is a zero probability event.

In other words, if we ignore the constraint (i) then $y_i(\cdot)$ is a solution to this relaxed problem iff the good is allocated to a bidder who has the lowest nonnegative value for $J_i(\theta_i)$. Now, recall the definition of $\overline{y_i}(\cdot)$. It is easy to write down the following expression:

$$\overline{y_i}(\theta_i) = E_{\theta_{-i}}\left[y_i(\theta_i, \theta_{-i})\right]. \tag{2.34}$$

Now, if we assume that $J_i(\cdot)$ is nondecreasing in $\theta_i$ then it is easy to see that the above solution $y_i(\cdot)$, given by (2.33), will be nonincreasing in $\theta_i$, which in turn implies, by looking at expression (2.34), that $\overline{y_i}(\cdot)$ is nonincreasing in $\theta_i$. Thus, the solution to this relaxed problem actually satisfies constraint (i) under the assumption that $J_i(\cdot)$ is nondecreasing. Assuming that $J_i(\cdot)$ is nondecreasing, the solution given by (2.33) seems to be the solution of the optimal mechanism design problem for single unit-single item procurement auction. The condition that $J_i(\cdot)$ is nondecreasing in $\theta_i$ is met by most of the distribution functions such as Uniform and Exponential.

So far we have computed the allocation rule for the optimal mechanism and now we turn our attention toward the payment rule. The optimal payment rule $t_i(\cdot)$ must be chosen in such a way that it satisfies

$$\overline{t_i}(\theta_i) = E_{\theta_{-i}}[t_i(\theta_i, \theta_{-i})] = U_i(\theta_i) + \theta_i \overline{y_i}(\theta_i) = \int_{\theta_i}^{\overline{\theta_i}} \overline{y_i}(s) ds + \theta_i \overline{y_i}(\theta_i). \tag{2.35}$$

Looking at the above formula, we can say that if the payment rule $t_i(\cdot)$ satisfies the following formula (2.36), then it would also satisfy the formula (2.35).

$$t_i(\theta_i, \theta_{-i}) = \int_{\theta_i}^{\overline{\theta_i}} y_i(s, \theta_{-i}) ds + \theta_i y_i(\theta_i, \theta_{-i}) \ \ \forall\, \theta \in \Theta. \tag{2.36}$$

The above formula can be rewritten more intuitively as follows. For any vector $\theta_{-i}$, let us define

$$z_i(\theta_{-i}) = \sup\left\{\theta_i : J_i(\theta_i) < \overline{\theta_0} \text{ and } J_i(\theta_i) \leq J_j(\theta_j)\ \forall\ j \neq i\right\}.$$

Then $z_i(\theta_{-i})$ is the supremum of all winning bids for bidder $i$ against $\theta_{-i}$, so

$$y_i(\theta_i, \theta_{-i}) = \begin{cases} 1 & : \quad \text{if } \theta_i < z_i(\theta_{-i}) \\ 0 & : \quad \text{if } \theta_i > z_i(\theta_{-i}). \end{cases}$$

This gives us

$$\int_{\theta_i}^{\overline{\theta_i}} y_i(s, \theta_{-i}) ds = \begin{cases} z_i(\theta_{-i}) - \theta_i & : \quad \text{if } \theta_i \leq z_i(\theta_{-i}) \\ 0 & : \quad \text{if } \theta_i > z_i(\theta_{-i}). \end{cases}$$

Finally, the formula (2.36) becomes

$$t_i(\theta_i, \theta_{-i}) = \begin{cases} z_i(\theta_{-i}) & : \quad \text{if } \theta_i \leq z_i(\theta_{-i}) \\ 0 & : \quad \text{if } \theta_i > z_i(\theta_{-i}). \end{cases}$$

That is, bidder $i$ will receive payment only when the good is procured from him, and then he receives an amount equal to his highest possible winning bid.

We make a few interesting observations:

1. When the various bidders have differing distribution function $\Phi_i(\cdot)$ then, the bidder who has the smallest value of $J_i(\theta_i)$ is *not* necessarily the bidder who has bid the lowest amount for the good. Thus Myerson's optimal auction need not be allocatively efficient, and therefore, need not be ex-post efficient.
2. If the bidders are symmetric, that is,
    - $\Theta_1 = \ldots = \Theta_n = \Theta$
    - $\Phi_1(\cdot) = \ldots = \Phi_n(\cdot) = \Phi(\cdot)$,

    then the allocation rule would be precisely the same as that of first-price reverse auction and second-price reverse auction. In such a case the object would be allocated to the lowest bidder. In such a situation, the optimal auction would also become allocatively efficient, and the payment rule described above would coincide with the payment rules in second-price reverse auction. In other words, the second price reverse auction would be an optimal auction when the bidders are symmetric. Therefore, many times, the optimal auction is also known as *modified Vickrey auction*.

Riley and Samuelson [30] also have studied the problem of design of an optimal auction for selling a single unit of a single item. They assume the bidders to be symmetric. Their work is less general than that of Myerson [24].

## 2.21 Further Topics in Mechanism Design

Mechanism design is a rich area with a vast body of knowledge. So far in this chapter, we have seen essential aspects of game theory, followed by key results in mechanism design. We now provide a brief description of a few topics in mechanism design. The topics have been chosen, with an eye on their possible application to network economics problems of the kind discussed in the monograph. We have not followed any particular logical order while discussing the topics. We also caution the reader that the treatment is only expository. Pointers to the relevant literature are provided wherever appropriate.

### 2.21.1 Characterization of DSIC Mechanisms

We have seen that a direct revelation mechanism is specified as $\mathscr{D} = ((\Theta_i)_{i \in N}, f(.))$, where $f$ is the underlying social choice function and $\Theta_i$ is the type set of agent $i$. A valuation function of each agent $i$, $v_i(.)$, associates a value of the allocation chosen by $f$ to agent $i$, that is, $v_i : K \to \mathbb{R}$, where $K$ is the set of project choices.

In the case of an auction for selling a single indivisible item, suppose each agent $i$ has a valuation for the object $\theta_i \in [\underline{\theta_i}, \overline{\theta_i}]$. If agent $i$ gets the object, $v_i(., \theta_i) = \theta_i$. Otherwise the valuation is zero. Thus for the agent $i$, the set of valuation functions over the set of allocations $K$ can be written as $\Theta_i = [\underline{\theta_i}, \overline{\theta_i}]$. Thus $\Theta_i$ is single dimensional in this environment.

In a general setting, $\Theta_i$ may not be single dimensional. If we consider all real valued functions on $X$ and allow each agent to have a valuation function to be any of these functions, we say $\Theta_i$ is unconstrained. Suppose $|K| = m$, then $\theta_i \in \Theta_i$ is an $m$-dimensional vector:

$$\theta_i = (\theta_{i_1}, \ldots, \theta_{i_j}, \ldots, \theta_{i_m}).$$

Note that $\theta_{i_j}$ will be the valuation of agent $i$ if the $j^{th}$ allocation from $K$ is selected. In other words, $v_i(j) = \theta_{i_j}$. With such unconstrained type sets/valuation functions, an elegant characterization of DSIC social choice functions has been obtained by Roberts [31]. The work of Roberts generalizes the the Green–Laffont Theorem (Theorem 2.8) in the following way. Recall that the Green–Laffont Theorem asserts that an allocatively efficient and DSIC social choice function in the above unconstrained setting has to be necessarily a VCG mechanism. The result of Roberts asserts that all DSIC mechanisms are variations of the VCG mechanism. These variants are often referred to as the *weighted VCG mechanisms*. In a weighted VCG mechanism, weights are given to the agents and to the outcomes. The resulting social choice function is said to be an *affine maximizer*. The notion of an affine maximizer is defined below. Next we state the Roberts' Theorem.

**Definition 2.47.** A social choice function $f$ is called an affine maximizer if for some subrange $A' \subset X$, for some agent weights $w_1, w_2, \ldots, w_n \in \mathbb{R}^+$, and for some outcome weights $c_x \in \mathbb{R}$, and for every $x \in A'$, we have that

$$f(\theta_1, \theta_2, \ldots, \theta_n) \in \arg\max_{x \in A'} (c_x + \sum_i w_i v_i(x)).$$

**Theorem 2.17 (Roberts' Theorem).** *If $|X| \geq 3$ and for each agent $i \in N$, $\Theta_i$ is unconstrained, then any DSIC social choice function $f$ has nonnegative weights $w_1, w_2, \ldots, w_n$ (not all of them zero) and constants $\{c_x\}_{x \in X}$, such that for all $\theta \in \Theta$,*

$$f(\theta) \in \arg\max_{x \in X} \left\{ \sum_{i=1}^{n} w_i v_i(x) + c_x \right\}.$$

For a proof of this important theorem, we refer the reader to the article by Roberts [31]. Lavi, Mu'alem, and Nisan have provided two more proofs for the theorem — interested readers might refer to their paper [32] as well.

### *2.21.2 Dominant Strategy Implementation of BIC Rules*

Clearly, dominant strategy incentive compatibility is stronger and much more desirable than Bayesian incentive compatibility. A striking reason for this is any Bayesian implementation assumes that the private information structure is common knowledge. It also assumes that the social planner knows a common prior distribution. In many cases, this requirement might be quite demanding. Also, a slight misspecification of the common prior may lead the equilibrium to shift quite dramatically. This may result in unpredictable effects; for example it might cause an auction to to be highly nonoptimal.

A dominant strategy implementation overcomes these problems in a simple way since the equilibrium strategy does not depend upon the common prior distribution. We would therefore always wish to have a DSIC implementation. Since the class of BIC social choice functions is much richer than DSIC social choice functions, one would like to ask the question: Can we implement a BIC SCF as a DSIC rule with the same expected interim utilities to all the players? Mookherjee and Stefan [33] have answered this question by characterizing BIC rules that can be equivalently implemented in dominant strategies. When these sufficient conditions are satisfied, a BIC social choice function could be implemented without having to worry about a common prior. The article by Mookherjee and Stefan [33] may be consulted for further details.

### *2.21.3 Implementation in Ex-Post Nash Equilibrium*

Dominant strategy implementation and Bayesian implementation are widely used for implementing a social choice function. There exists another notion of implementation, called ex-post Nash implementation, which is stronger than Bayesian implementation but weaker than dominant strategy implementation. This was formalized by Maskin [34]. Dasgupta, Hammond, and Maskin [35] generalized this to the Bayesian Nash implementation.

**Definition 2.48.** A profile of strategies $\left(s_1^*(.), s_2^*(.), \ldots, s_n^*(.)\right)$ is an ex-post Nash equilibrium if for every $\theta = (\theta_1, \ldots \theta_n) \in \Theta$, the profile $\left(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)\right)$ is a Nash equilibrium of the complete information game defined by $(\theta_1, \ldots, \theta_n)$. That is, for all $i \in N$ and for all $\theta \in \Theta$, we have

$$u_i(s_i^*(\theta_i), s_{-i}^*(\theta_{-i}), \theta_i) \geq u_i(s_i'(\theta_i), s_{-i}^*(\theta_{-i}), \theta_i) \; \forall \, s_i' \in S_i.$$

In a Bayesian Nash equilibrium, the equilibrium strategy is played by the agents after observing their own private types and computing an expectation over others' types; it is an equilibrium only in the expected sense. On the other hand, in ex-post Nash equilibrium, even after the players are informed of the types of the other players, it is still a Nash equilibrium for each agent $i$ to play an action according to $s_i^*(\cdot)$. This is called the *lack of regret* feature. That is, even if agents come to know about the others' types, the agent need not regret playing this action. Bayesian Nash equilibrium may not have this feature since the agents may want to revise their strategies after knowing the types of the other agents.

For example, consider the first price sealed bid auction with two bidders. Let $\Theta_1 = \Theta_2 = [0, 1]$ and $\theta_1$ denote the valuation of the first agent and $\theta_2$ that of the agent 2. It can be shown that it is a Bayesian Nash equilibrium for each bidder to bid according to the strategy $(b_1^*(\theta_1), b_2^*(\theta_2)) = (\frac{\theta_1}{2}, \frac{\theta_2}{2})$. Now suppose agent 1 is informed that the other agent values the object at 0.6. If agent 1 has a valuation of 0.8, say, it is not a Nash equilibrium for him to bid 0.4 even if agent 2 is still following a Bayesian Nash strategy.

**Definition 2.49.** We say that the mechanism $\mathscr{M} = ((S_i)_{i \in N}, g(\cdot))$ implements the social choice function $f(\cdot)$ in ex-post Nash equilibrium if there is a pure strategy ex-post Nash equilibrium $s^*(\cdot) = (s_1^*(\cdot), \ldots, s_n^*(\cdot))$ of the game $\Gamma^b$ induced by $\mathscr{M}$ such that

$$g\left(s_1^*(\theta_1), \ldots, s_n^*(\theta_n)\right) = f\left(\theta_1, \ldots, \theta_n\right) \; \forall \left(\theta_1, \ldots, \theta_n\right) \in \Theta.$$

Though ex-post implementation is stronger than Bayesian Nash implementation, it is still much weaker than dominant strategy implementation.

### *2.21.4 Interdependent Values*

We have so far assumed that the private values or signals observed by the agents are independent of one another. This is a reasonable assumption in many situations. However, in the real world, there are environments where the valuation of agents might depend upon the information available or observed by the other agents. We will look at two examples.

*Example 2.62.* Consider an auction for an antique painting. There is no guarantee that the painting is an original one or a plagiarized version. If all the agents knew that the painting is not an original one, they would have a very low value for it independent of one another, whereas on the other hand, they would have a high value for it when it is a genuine piece of work. But suppose they have no knowledge about its authenticity. In such a case, if a certain bidder happens to get information about its genuineness, the valuations of all the other agents will naturally depend upon this signal (indicating the authenticity of the painting) observed by this agent.

*Example 2.63.* Consider an auction for oil drilling rights. At the time of the auction, buyers usually conduct geological tests and their private valuations depend upon the results of these tests. If a prospective bidder knew the results of the tests of the others, his own willingness to pay for the drilling rights would be modulated suitably based on the information available.

The interdependent private value models have been studied in the mechanism design literature. For example, there exists a popular model called the *common value model* (which we have already seen in Section 2.19.2). As another example, consider a situation when a seller is trying to sell an indivisible good or a fixed quantity of a divisible good. The value of the received good for the bidders depends upon each others' private signals. Also, the private signals observed by the agents are interdependent of specified properties. In such a scenario, Cremer and McLean [36] have designed an auction that extracts a revenue from the bidders, which is equal to what could have been extracted when the actual signals of the bidders are known. In this auction, it is an ex-post Nash equilibrium for the agents to report their true types. This auction is interim individually rational but may not be ex-post individually rational.

### *2.21.5 Implementation Theory*

Dominant strategy incentive compatibility ensures that reporting true types is a weakly dominant strategy equilibrium. Bayesian incentive compatibility ensures that reporting true types is a Bayesian Nash equilibrium. Typically, the Bayesian game underlying a given mechanism may have multiple equilibria, in fact, could have infinitely many equilibria. These equilibria typically will produce different outcomes. Thus it is possible that nonoptimal outcomes are produced by truth revelation.

The *implementation problem* addresses the above difficulty caused by multiple equilibria. The implementation problem seeks to design mechanisms in which all the equilibrium outcomes are optimal. This property is called the *weak implementation property*. If it also happens that every optimum outcome is also an equilibrium, we call the property as *full implementation property*. Maskin [34] provided a general characterization of Nash implementable social choice functions using a monotonicity property, which is now called *Maskin Monotonicity*. This property has a striking similarity to the property of independence of irrelevant alternatives, which we encountered during our discussion on Arrow's impossibility theorem (Section 2.12). Maskin's work shows that Maskin monotonicity, in conjunction with another property called *no-veto-power* will guarantee that all Nash equilibria will produce an optimal outcome. His work has led to development of implementation theory. Dasgupta, Hammond, and Maskin [35] have summarized many important results in implementation theory, and they discuss incentive compatibility in detail. Maskin's results have now been generalized in many directions; for example, see the references in [37].

### *2.21.6 Computational Issues in Mechanism Design*

We have seen several possibility and impossibility results in the context of mechanism design. While every possibility result is good news, there could be still be challenges involved in actually implementing a mechanism that is possible. For example, we have seen that the GVA mechanism (Example 2.55) is an allocatively efficient and dominant strategy incentive compatible mechanism for combinatorial auctions. A major difficulty with GVA is the computational complexity involved in determining the allocation and the payments. Both the allocation and payment determination problems are NP-hard, being instances of the weighted set packing problem (in the case of forward GVA) or the weighted set covering problem (in the case of reverse GVA). In fact, if there are $n$ agents, then in the worst case, the payment determination will involve solving as many as $n$ NP-hard problems, so overall, as many as $(n+1)$ NP-hard problems will need to be solved for implementing the GVA mechanism. Moreover, approximately solving any one of these problems may compromise properties such as efficiency and/or incentive compatibility of the mechanism.

In mechanism design, computations are involved at two levels: first, at the agent level and secondly at the mechanism level [38, 39]. Complexity at the agent level involves strategic complexity (complexity of computing an optimal strategy) and valuation complexity (computation required to provide preference information within a mechanism). Complexity at the mechanism level includes communication complexity (how much communication is required between agents and the mechanism to compute an outcome) and winner determination complexity (computation required to determine an outcome given the strategies of the agents). Typically, insufficient computation leading to approximate solutions hinders mechanism design since properties such as incentive compatibility, allocative efficiency, individual rationality, etc., may be compromised. Novel algorithms and high computing power surely lead to better mechanisms.

For a detailed description of computational complexity issues in mechanism design, the reader is referred to the excellent survey articles [40, 41, 38, 39].

## 2.22 To Probe Further

For a microeconomics oriented treatment of mechanism design, the readers are requested to refer to textbooks, such as the ones by Mas-Colell, Whinston, and Green [6] (Chapter 23); Green and Laffont [18]; and Laffont [42]. There is an excellent recent survey article by Nisan [43], which targets a computer science audience. There are many other informative survey papers on mechanism design — for example by Myerson [44], Serrano [45], and Jackson [46, 47]. The Nobel Prize website has a scholarly technical summary of mechanism design theory [37]. The recent edited volume on *Algorithmic Game Theory* by Nisan, Roughgarden, Tardos, and Vazirani [48] also has valuable articles related to mechanism design.

This chapter is not to be treated as a survey on auctions in general. There are widely popular books (for example, by Milgrom [27], Krishna [28], and Klemperer [53]) and surveys on auctions (for example, [25, 49, 50, 51, 38]) that deal with auctions in a comprehensive way.

A related area where an extensive amount of work has been carried out in the past decade is combinatorial auctions. Exclusive surveys on combinatorial auctions include the articles by de Vries and Vohra [40, 41], Pekec and Rothkopf [52], and Narahari and Pankaj Dayama [54]. Cramton, Ausubel, and Steinberg [21] have brought out a comprehensive edited volume containing expository and survey articles on varied aspects of combinatorial auctions.

## References

1. R.B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, Massachusetts, 1997.
2. J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
3. R. Aumann. Agreeing to disagree. *The Annals of Statistics*, 4(6):1236–1239, 1976.
4. E. Tardos and V.V. Vazirani. Basic solution concepts and computational issues. In N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani, (eds.), *Algorithmic Game Theory*, pages 3–28. Cambridge University Press, New York, 2007.
5. J. Bertrand. Book review of theorie mathematique de la richesse sociale and of recherches sur les principles mathematiques de la theorie des richesses. *Journal de Savants*, 67:499–508, 1883.
6. A. Mas-Colell, M.D. Whinston, and J.R. Green. *Microeconomic Theory*. Oxford University Press, New York, 1995.
7. R.D. McKelvey and A. McLennan. Computation of equilibria in finite games. In J. Rust H.M. Amman, D.A. Keudrick, (eds.), *Handbook of Computational Economics, Handbooks in Economics (13), Volume 1*, pages 87–142. North-Holland, Amsterdam, 1996.

8. C.H. Papadimitriou. The complexity of finding nash equilibria. In N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani, (eds.), *Algorithmic Game Theory*, pages 29–52. Cambridge University Press, New York, 2007.
9. B.V. Stengel. Equilibrium computation for two-player games in strategic and extensive form. In N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani, (eds.), *Algorithmic Game Theory*, pages 53–78. Cambridge University Press, New York, 2007.
10. L. Hurwicz. On informationally decentralized systems. In C.B. McGuire and R. Radner, (eds.), *Decision and Organization*, North-Holland, Amsterdam, 1972.
11. L. Hurwicz. Optimality and informational efficiency in resource allocation processes. In K.J. Arrow, S. Karlin and P. Suppes (eds.), *Mathematical Methods in the Social Sciences*. Stanford University Press, Palo Alto, California, USA, 1960.
12. B. Holmstrom and R.B. Myerson. Efficient and durable decision rules with incomplete information. *Econometrica*, 51(6):1799–1819, 1983.
13. A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–601, 1973.
14. M.A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorem for voting procedure and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.
15. W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16(1):8–37, 1961.
16. E. Clarke. Multi-part pricing of public goods. *Public Choice*, 11:17–23, 1971.
17. T. Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973.
18. J.R. Green and J.J. Laffont. *Incentives in Public Decision Making*. North-Holland, Amsterdam, 1979.
19. L.M. Ausubel and P. Milgrom. The lovely but lonely vickrey auction. In P. Cramton, Y. Shoham, and R. Steinberg, (eds.), *Combinatorial Auctions*, pages 17–40. The MIT Press, Cambridge, Massachusetts, 2006.
20. M. Rothkopf. Thirteen reasons why the Vickrey-Clarke-Groves process is not practical. *Operations Research*, 55(2):191–197, 2007.
21. C. Caplice and Y. Sheffi. Combinatorial auctions for truckload transportation. In P. Cramton, Y. Shoham, and R. Steinberg, (eds.), *Combinatorial Auctions*, pages 539–572. The MIT Press, Cambridge, Massachusetts, 2005.
22. C. d'Aspremont and L.A. Gérard-Varet. Incentives and incomplete information. *Journal of Public Economics*, 11:25–45, 1979.
23. K. Arrow. The property rights doctrine and demand revelation under incomplete information. In M. Boskin, (eds.), *Economics and Human Welfare*. Academic Press, New York, 1979.
24. R.B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.
25. P.R. McAfee and J. McMillan. Auctions and bidding. *Journal of Economic Literature*, 25(2):699–738, 1987.
26. P. Klemperer. Why every economist should learn some auction theory. In M. Dewatripont, L. Hansen, and S. Turnovsky, (eds.), *Advances in Economics and Econometrics: Invited Lectures to 8th World Congress of the Econometric Society*. Cambridge University Press, Cambridge, UK, 2003.
27. P. Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, Cambridge, UK, 2004.
28. V. Krishna. *Auction Theory*. Academic Press, San Diego, California, USA, 2002.
29. D. Garg, Y. Narahari, and S. Gujar. Foundations of mechanism design: A tutorial – Part 1: Key Concepts and Classical Results. *Sadhana — Indian Academy Proceedings in Engineering Sciences*, 33(2):83–130, 2008.
30. J.G. Riley and W.F. Samuelson. Optimal auctions. *American Economic Review*, 71(3):383–92, 1981.
31. K. Roberts. The characterization of implementable choice rules. In J.J. Laffont, (eds.), *Aggregation and Revelation of Preferences*, pages 321–349, Amsterdam, 1979.

32. R. Lavi, A. Mu'alem, and N. Nisan. Two simplified proofs for Roberts' theorem. Technical report, Working Paper, School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel, 2004.
33. D. Mookherjee and S. Reichelstein. Dominant strategy implementation of Bayesian incentive compatible allocation rules. *Journal of Economic Theory*, 56(2):378–399, 1992.
34. E. Maskin. Nash equilibrium and welfare optimality. *Review of Economic Studies*, 66:23–38, 1999.
35. P. Dasgupta, P. Hammond, and E. Maskin. The implementation of social choice rules: some general results on incentive compatibility. *Review of Economic Studies*, 46:181–216, 1979.
36. J. Cremer and R.P McLean. Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica*, 53(2):345–61, 1985.
37. The Nobel Foundation. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2007: Scientific Background. Technical report, The Nobel Foundation, Stockholm, Sweden, December 2007.
38. J.K. Kalagnanam and D.C. Parkes. Auctions, bidding, and exchange design. In D. Simchi-Levi, S.D. Wu, and Z.J. Shen, (eds.), *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*. Kluwer Academic Publishers, New York, 2005.
39. T. Sandholm. Computing in mechanism design. In S.N. Durlauf and L.E. Blume, (eds.), *The New Palgrave Dictionary of Economics*. Second Edition, Palgrave Macmillan, 2008.
40. S. de Vries and R.V. Vohra. Combinatorial auctions: A survey. *INFORMS Journal of Computing*, 15(1):284–309, 2003.
41. S. de Vries and R.V. Vohra. Design of combinatorial auctions. In *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*, pages 247–292. International Series in Operations Research and Management Science, Kluwer Academic Publishers, Norwell, MA, USA, 2005.
42. J.J. Laffont. *Fundamentals of Public Economics*. The MIT Press, Cambridge, Massachusetts, 1988.
43. N. Nisan. Introduction to mechanism design (for computer scientists). In N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani (eds.), *Algorithmic Game Theory*, pages 209–242. Cambridge University Press, New York, 2007.
44. R. Myerson. Mechanism design. In J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave Dictionary of Economics*, pages 191–206. Norton, New York, 1989.
45. R. Serrano. The theory of implementation of social choice rules. *SIAM Review*, 46:377–414, 2004.
46. M.O. Jackson. A crash course in implementation theory. *Social Choice and Welfare*, 18:655–708, 2001.
47. M.O. Jackson. Mechanism theory. In U. Derigs (ed.), *Encyclopedia of Life Support Systems*. EOLSS Publishers, Oxford, U.K., 2003.
48. N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani (eds.). *Algorithmic Game Theory*. Cambridge University Press, New York, 2007.
49. P. Milgrom. Auctions and bidding: A primer. *Journal of Economic Perspectives*, 3(3):3–22, 1989.
50. P. Klemperer. *Auctions: Theory and Practice. The Toulouse Lectures in Economics*. Princeton University Press, Princeton, NJ, USA, 2004.
51. E. Wolfstetter. Auctions: An introduction. *Economic Surveys*, 10:367–421, 1996.
52. A. Pekec and M.H. Rothkopf. Combinatorial auction design. *Management Science*, 49:1485–1503, 2003.
53. P. Klemperer Auctions: Theory and Practice. Princeton University Press, Princeton, NJ, USA, 2004.
54. Y. Narahari and P. Dayama. Combinatorial auctions for electronic business. *Sadhana — Indian Academy Proceedings in Engineering Sciences*, 30(2-3):179–212, 2005.