# Game Theory

**Lecture Notes By**


**Y. Narahari**

**Department of Computer Science and Automation**

**Indian Institute of Science**

**Bangalore, India**


**July 2012**

## The Mechanism Design Environment

**Note**: *This is a only a draft version, so there could be flaws. If you find any errors, please do send email to* `hari@csa.iisc.ernet.in`. *A more thorough version would be available soon in this space.*

## 1 Nature of Mechanism Design Problems

In mechanism design, problem solving involves implementing a *system-wide solution* that will satisfy certain *desirable properties*. In this process of problem solving, the following key common characteristics of these problems need to be taken into account:

- There is a set of decision makers or players who interact in a *strategic* way. The players have well defined payoff functions and are *rational* in the sense of having the sole objective of maximizing their own individual payoffs. The respective objectives of the individual players could be conflicting. Both conflict and cooperation could be possible during the interactions of these rational players.

- Each player holds certain information which is *private* and only this player would know it deterministically; other players do not know this information deterministically. Thus the information in the system is decentralized and each player only has incomplete information. Of course, there could be some information which all players know and all players know that all players know and so on. Such information is *common knowledge*.

- Each player has a choice of certain strategies that are available to them. The players have enough *intelligence* to determine their *best response strategies*.

Because of the above characteristics, the problems could be called *game theoretic problems*. Solving problems in such a setting would involve solving a *decision or optimization problem with incomplete information or incomplete specification*. Mechanism design offers a natural tool to solve such problems by providing an elegant way to do *reverse engineering* of games. Essentially, using techniques of mechanism design, we induce a game among the players in such a way that in an equilibrium of the induced game, the desired system-wide solution is implemented.

## 2 Mechanism Design

In the second half of the twentieth century, game theory and mechanism design have found widespread use in a gamut of applications in engineering. In particular, game theory and mechanism design have emerged as an important tool to model, analyze, and solve *decentralized design problems* in engineering involving multiple autonomous agents that interact strategically in a *rational* and *intelligent* way. The field of mechanism design has been in intense limelight in the recent times; the Nobel Prize in Economic Sciences for the year 2007 was jointly awarded to three economists, Leonid Hurwicz, Eric Maskin, and Roger Myerson *for having laid the foundations of mechanism design theory* [1]. Earlier, in 1996, William Vickrey, the inventor of the famous Vickrey auction had been awarded the Nobel Prize in Economic Sciences.

The theory of *mechanism design* is concerned with settings where a policy maker (or social planner) faces the problem of aggregating the *announced preferences* of multiple agents into a collective (or social) decision when the *actual preferences* are not publicly known. Mechanism design theory uses the framework of *non-cooperative games with incomplete information* and seeks to study how the privately held preference information can be elicited. The theory also clarifies the extent to which the preference elicitation problem constrains the way in which social decisions can respond to individual preferences. In fact, mechanism design can be viewed as *reverse engineering* of games or equivalently as the *art of designing the rules of a game to achieve a specific desired outcome*. The main focus of mechanism design is to design institutions or protocols that satisfy certain desired objectives, assuming that the individual agents, interacting through the institution, will act strategically and may hold private information that is relevant to the decision at hand.

### 2.1 Mechanisms: Some Simple Examples

Mechanisms have been used and practiced from times immemorial. Auctions provide a popular example of mechanisms; as is well known, auctions have been in vogue for a long time for selling, procuring, and exchanging goods and services.

Two simple, popular stories capture the idea behind mechanisms quite strikingly. The first story is that of a mother of two kids who has to design a mechanism to make her two kids share a cake equally. The mother is the social planner in this case and the mechanism she designs is the following: (1) One of the kids would slice the cake into two pieces and (2) the other kid would pick up one of the pieces, leaving the remaining piece to the kid who sliced the cake into two pieces. This mechanism implements the desirable outcome of the kids sharing the cake equally (of course, it would be interesting to see what a suitable mechanism would be if instead of two, there were more kids).

The second story is from ancient wisdom. This is attributed to several wise people. In India, it is attributed independently to Birbal who was an adviser to Emperor Akbar in the late 1500s and to Tenali Rama, who was a popular poet and adviser to the king in the court of the famous King Sri Krishna Devaraya of the Vijayanagara dynasty in the early 1500s. In this fable, two women come to the king with a baby, each claiming to be the baby's mother, seeking justice. The clueless king turns to his adviser for advice. Birbal (Tenali Rama) is supposed to have suggested that the baby be sliced into two pieces and the two pieces be equally shared by the two mothers. Upon which, one of the women (the real mother) immediately pleaded with the king not to resort to the cruelty. The king immediately ordered that the baby be handed over to that woman. This is an example of a truth elicitation mechanism.

Mechanisms such as above are ubiquitous in everyday life. The emergence of game theory during the 1940s and 1950s helped develop a formal theory of mechanism design starting from the 1960s.

### 2.2 Mechanism Design: A Brief History

Leonid Hurwicz (Nobel laureate in Economic Sciences in 2007) first introduced the notion of mechanisms with his work in 1960 [2]. He defined a mechanism as a communication system in which participants send messages to each other and perhaps to a *message center* and a pre-specified rule assigns an outcome (such as allocation of goods and payments to be made) for every collection of received messages. William Vickrey (Nobel laureate in Economic Sciences in 1996) wrote a classic paper in 1961 [3] which introduced the famous Vickrey auction (second price auction). To this day, the Vickrey auction continues to enjoy a special place in the annals of mechanism design. John Harsanyi (Nobel laureate in Economic Sciences in 1994 jointly with John Nash and Richard Selten) developed the theory of games with incomplete information, in particular Bayesian games, through a series of three seminal papers in 1967-68 [4, 5, 6]. Harsanyi's work later proved to be of foundational value to mechanism design. Hurwicz [7] introduced the key notion of incentive compatibility in 1972. This notion allowed mechanism design to incorporate the incentives of rational players and opened up mechanism design. Clarke [8] and Groves [9] came up with a generalization of the Vickrey mechanisms and helped define broad class of dominant strategy incentive compatible mechanisms in the quasi-linear environment.

There were two major advances in mechanism design in the 1970s. The first was the *revelation principle* which essentially showed that direct mechanisms are the same as indirect mechanisms. This meant that mechanism theorists needed to worry only about direct mechanisms, leaving the development of real-world mechanisms (which are mostly indirect mechanisms) to mechanism designers and practitioners. Gibbard [10] formulated the revelation principle for dominant strategy incentive compatible mechanisms. This was later extended to Bayesian incentive compatible mechanisms through several independent efforts [1] - Maskin and Myerson (both Nobel laureates in Economic Sciences in 2007) had a leading role to play in this. In fact, Myerson developed the revelation principle in its greatest generality [1]. The second major advance in mechanism design in the 1970s was on *implementation theory* which addresses the following problem: can a mechanism be designed so that all its equilibria are optimal? Maskin [11] gave the first general solution to this problem.

Mechanism design has made phenomenal advances during 1980s, 1990s, and during the past few years. It has found widespread applicability in a variety of disciplines. These include: design of markets and trading institutions [1, 12, 13], regulation and auditing [1], social choice theory [1], and computer science [14]. The above list is by no means exhaustive. In this monograph, our focus is on applying mechanism design in the area of Internet and network economics.

## 3 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 3.1 (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 1 (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 1 provides a pictorial representation of all the elements making up the mechanism design environment.

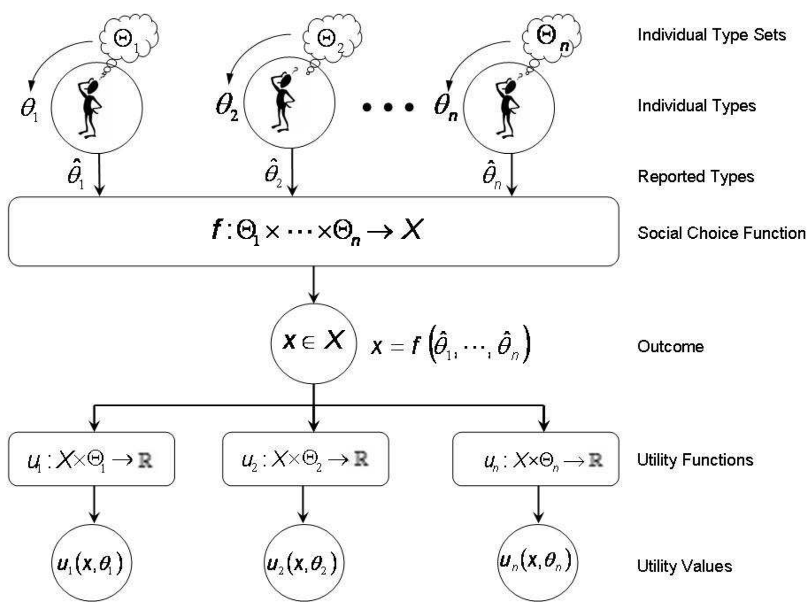


Figure 1: Mechanism design environment

### 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.

**Definition 3.2 (Direct Mechanism)** *Given a social choice function $f : \Theta_1 \times \Theta_2 \times \ldots \times \Theta_n \to X$, a direct mechanism (also called 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 3.3 (Indirect Mechanism)** *An indirect mechanism (also called 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 $g : S_1 \times S_2 \times \ldots \times S_n \to X$ 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 chapters of this book, 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.

### To Probe Further

The material discussed in this chapter draws upon mainly from three sources, namely the books by Myerson [15], Mascolell, Whinston, and Green [12], and Osborne and Rubinstein [16].

The following books also contain illustrative examples of strategic form games: Osborne [16], Straffin [17], and Binmore [18].

## References

[1] 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, 2007.

[2] L. Hurwicz. Optimality and informational efficiency in resource allocation processes. In K.J. Arrow and S. Karlin, editors, *Mathematical Methods in the Social Sciences*. Stanford University Press, Palo Alto, California, USA, 1960.

[3] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16(1):8–37, 1961.

[4] John C. Harsanyi. Games with incomplete information played by Bayesian players. Part I: The basic model. *Management Science*, 14:159–182, 1967.

[5] John C. Harsanyi. Games with incomplete information played by Bayesian players. Part II: Bayesian equilibrium points. *Management Science*, 14:320—334, 1968.

[6] John C. Harsanyi. Games with incomplete information played by Bayesian players. Part III: The basic probability distribution of the game. *Management Science*, 14:486–502, 1968.

[7] L. Hurwicz. On informationally decentralized systems. In C.B. McGuire and R. Radner, editors, *Decision and Organization*. North-Holland, Amsterdam, 1972.

[8] E. Clarke. Multi-part pricing of public goods. *Public Choice*, 11:17–23, 1971.

[9] T. Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973.

[10] A. Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–601, 1973.

[11] E. Maskin. Nash equilibrium and welfare optimality. *Review of Economic Studies*, 66:23–38, 1999.

[12] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Micoreconomic Theory*. Oxford University Press, 1995.

[13] Paul Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, Cambridge, UK, 2004.

[14] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay Vazirani (Editors). *Algorithmic Game Theory*. Cambridge University Press, 2007.

[15] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, Massachusetts, USA, 1997.

[16] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory*. Oxford University Press, 1994.

[17] Philip D. Straffin Jr. *Game Theory and Strategy*. The Mathematical Association of America, 1993.

[18] Ken Binmore. *Fun and Games : A Text On Game Theory*. D. C. Heath & Company, 1992.