CHAPTER 1

# Introduction

The 2007 Nobel Prize in economics honored a subject, mechanism design, fundamental to the study of incentives and information. Its importance is difficult to convey in a sound bite because it does not arise from a to-do list or a ten-point plan. Rather, it is an analytical framework for thinking clearly and carefully about the most fundamental of social problems: What exactly can a given institution achieve when the information necessary to make decisions is dispersed and privately held? The range of questions to which the approach can be applied is striking. To achieve a given reduction in carbon emissions, should one rely on taxes or a cap-and-trade system? Is it better to sell an Initial Public Offering (IPO) via auction or the traditional book-building approach? Would juries produce more informed decisions under a unanimity rule or that of simple majority? Mechanism design helps us understand how the answers to these questions depend on the details of the underlying environment. In turn, this helps us understand which details matter and which do not.

To get a sense of what mechanism design is, we begin with a fable, first told by the Nobelist, Ronald Coase. It involves, as all good fables do, a coal-burning locomotive and a farmer. The locomotive emits sparks that set fire to the farmer's crops. Suppose that running the locomotive yields $1,000 worth of profit for the railroad but causes $2,000 worth of crop damage. Should the railroad be made to pay for the damage it causes?

The sparks alone do no damage. One might say the farmer caused the damage by placing crops next to the railway line. It is the *juxtaposition* of sparks and crops that lead to the $2,000 worth of damage. Perhaps, then, the farmer is liable?

If you think this strange, suppose it costs the farmer $100 to ensure the safety of the crop. If we make the railroad liable for damage to the crop, what happens? The locomotive stops running.[1] Why spend $2,000 to get a return of $1,000? The farmer takes no precautions to secure the crop. As a society, we are out $1,000 – the profit the railroad would have made had it continued to run the locomotive. Now suppose we make the farmer liable. The locomotive runs.

[1] Assuming the absence of technology that would eliminate the sparks.

The farmer pays \$100 to safeguard the crop rather than \$2,000 in crop damage. On the balance, society is out only \$100. If we cared about avoiding damage in the most cost-effective way possible, we should make the farmer liable.

Suppose now the railroad had access to technology that would eliminate the sparks for a price of \$50. In this case, because it is cheaper for the railroad to avoid the damage, it should be made liable. If cost effectiveness is our lodestar, it puts us in a pickle, because the assignment of liability depends on the details of the particular situation. Coase's essential insight is that it does not matter how liability is assigned as long as the parties are permitted to trade the liability among themselves.

Suppose the railroad is made liable. What matters is whether or not the railroad can pay the farmer to shoulder the liability. Assume, as before, that the railroad cannot reduce the sparks emitted without shutting down the locomotive, and that the farmer can avoid the crop damage at a cost of \$100. Observe that the railroad is better off paying the farmer at least \$100 (and no more than \$1,000) to move the crops. The farmer will also be better off. In effect, the railroad pays the farmer to assume the liability – seemingly a win-win arrangement. Thus, as long as we allow the parties concerned to trade their liabilities, the party with the least cost for avoiding the damage will shoulder the liability. In terms of economic efficiency, it does not matter who is liable for what. It matters only that the liabilities be clearly defined, easily tradeable, and enforced. It is true that the farmer and railroad care a great deal about who is held liable for what. If it is the railroad, then it must pay the farmer. If it is the farmer, the railroad pays nothing. One may prefer, for reasons quite separate from economic efficiency, to hold one party liable rather than the other. However, the outcome in terms of who does what remains the same.

Coase recognizes there are transaction costs associated with bargaining over the transfer of liabilities. Because they might overwhelm the gains to be had from bargaining, it is of fundamental importance that such costs be minimized. Nevertheless, mutually beneficial bargains fail to be struck even when transaction costs are nonexistent. Personality, ego, and history conspire to prevent agreement. These are unsatisfying explanations for *why* mutually beneficial agreements are unmade *because* they are idiosyncratic and situation specific. Mechanism design suggests another reason: The actual cost incurred by each party to avoid the damage is private information known only to themselves.

To see why, suppose the railroad incurs a cost \$R of avoiding the damage whereas the farmer incurs a cost of \$F to do the same. Only the railroad knows \$R and only the farmer knows \$F. If \$F > \$R, economic efficiency dictates that the railroad should incur the cost of avoiding the damage. If \$F < \$R, efficiency requires the farmer to shoulder the cost of avoiding the damage. In the event that \$F = \$R, we are indifferent as to which one incurs the cost.

Now, let us – quite arbitrarily – make the railroad liable for the damage and trust that bargaining between railroad and farmer will result in the person with the lower cost of avoiding the damage undertaking the burden to avoid the damage. If \$R > \$F, the railroad should pay the farmer to take on the liability.

Furthermore, it would want to pay as little as possible – ideally no more than \$F. However, the railroad does not know the magnitude of F. So, how much should it offer? The lower the offer, the less likely it will be accepted. On the other hand, if accepted, the more profitable it is to the railroad. On the flip side, the farmer has every incentive to bluff the railroad into thinking that \$F is larger than it actually is so as to make a tidy profit. If the farmer is too aggressive in this regard, the railroad may walk away thinking that \$R < \$F. One can conceive of a variety of bargaining procedures that might mitigate these difficulties. Is there a bargaining protocol that will lead inexorably to the party with the lower cost of avoiding the damage assuming the liability?

Mechanism design approaches this question using the tools of game theory. Any such protocol can be modeled as a game that encourages each party to truthfully reveal its cost of avoiding the damage so that the correct assignment of liability can be made. The encouragement to truthfully reveal this private information is obtained with money. The monetary rewards must be generated internally, that is, there is no rich uncle waiting on the sidelines to come to the aid of either the farmer or the railroad. Thus, the question becomes a purely mathematical one: Is there a game with these properties? Myerson and Satterthwaite (1983) proved that the answer to this question was a resounding, de Gaulle – like, "NON." There is no bargaining protocol or trusted mediator that is guaranteed in all circumstances to ensure that the party with the lower cost of avoiding the damage assumes the liability. Hence, there is always the possibility that no bargain will be struck even when it is in the mutual interest of both parties to come to terms.

Thus, Coase's original observation that the assignment of liability is irrelevant because an incorrect assignment would be corrected by bargaining in the marketplace (provided transaction costs are small) is rendered false in the presence of private information. Mechanism design also suggests how liability should be assigned. Specifically, to ensure that the liability is assigned to the party with the lowest cost for avoiding the damage, the right to *avoid* the liability should be auctioned off to the highest bidder. How is this possible? Suppose our auction works as follows. We have a price clock initially set at zero. We then raise the price. At each price, we ask the bidders (railroad and farmer) whether they wish to buy the right to avoid liability at the current price. If both say "yes," continue raising the price. The instant one of them drops out, stop and sell the right to the remaining active bidder at the terminal price. Observe that the farmer will stay active as long as the current price is below \$F. The railroad will stay active as long as the current price is below \$R. If the farmer drops out first, it must be because \$F < \$R. In this case, the farmer assumes liability and the railroad pays the auctioneer \$F. In short, the farmer, who had the lower cost of avoiding the damage, is saddled with the liability. If \$R < \$F, the reverse happens.

The fable of the railroad and the farmer involved the allocation of liability. It could just as well have involved the allocation of a property right. One is the obverse of the other. Now, the punchline. When governments create new property rights or asset classes, these should be auctioned off to ensure they are allocated in an economically efficient manner. It is exactly this reasoning that supports the allocation of spectrum rights by auction. It is exactly this reasoning that supports the allocation of permits to pollute by auction. It is exactly this reasoning that will eventually propel the Federal Aviation Authority (FAA) to use auctions to allocate arrival and departure slots at airports. Keynes said it best: "I am sure that the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas."

It is not my ambition to provide a complete account of mechanism design and its implications. My goal is more modest. It is to provide a systematic account of the underlying mathematics of the subject. The novelty lies in the approach. The emphasis is on the use of linear programming as a tool for tackling the problems of mechanism design. This is at variance with custom and practice, which have relied on calculus and the methods of analysis.[2] There are three advantages of such an approach:

1. Simplicity. Arguments based on linear programming are both elementary and transparent.
2. Unity. The machinery of linear programming provides a way to unify results from disparate areas of mechanism design.
3. Reach. It provides the ability to solve problems that appear to be beyond the reach of traditional methods.

No claim is made that the approach advocated here should supplant the traditional mathematical machinery. Rather, it is an addition to the quiver of the economic theorist who purposes to understand economic phenomena through the lens of mechanism design.

It is assumed the reader has some familiarity with game theory, the basics of linear programming, and some convex analysis. This is no more than what is expected of a first-year student in a graduate economics program. No prior knowledge of mechanism design is assumed. However, the treatment offered here will be plain and unadorned. To quote Cassel, it lacks "the corroborative detail, intended to give artistic verisimilitude to an otherwise bald and unconvincing narrative."

The point of view that animates this monograph is the product of collaborations with many individuals, including Sushil Bikhchandani, Sven de Vries, Alexey Malakhov, Rudolf Müller, Mallesh Pai, Teo Chung Piaw, Jay Sethuraman, and James Schummer. However, they are not responsible for errors of commission or omission on my part.

It was William Thomson who first suggested that I put all this down on paper. The spur was an invitation from Luca Rigotti, Pino Lopomo, and Sasa Pekec to talk about these matters at Duke University's Fuqua School. My thanks to Daniele Condorelli, Antoine Loeper, Rudolf Müller, and John Weymark,

[2] My colleagues refer to this as the pre-Newtonian approach to mechanism design.

who provided comments on an earlier version. George Mailath and anonymous reviewers provided invaluable suggestions on focus and intuition. An invitation from Benny Moldovanu to spend time at the Hausdorff Institute during its program on mechanism design provided valuable time to complete this project. Michael Sara was very helpful in preparing the figures. My particular thanks to Simone Galperti and Gabriel Carroll, who helped ferret out numerous blushworthy mistakes.

## 1.1 OUTLINE

Here is a brief outline of the other chapters.

### Chapter 2

This chapter is devoted to classical social choice. There are two main results. The first is a linear inequality description of all social welfare functions that satisfy Arrows conditions. These inequalities are then employed to derive Arrow's celebrated Impossibility Theorem. The same inequalities can be employed to derive other results about social welfare functions.

The second result is a proof of the Gibbard-Satterthwaite Impossibility Theorem. A number of authors have commented on the similarities between Arrow's Theorem and the Gibbard-Satterthwaite Theorem. Reny (2001), for example, provides a unified proof of the two results. In this chapter, it is shown that the social-choice functions of the Gibbard-Satterthwaite Theorem must satisfy the same inequalities as the social welfare functions of Arrow's Theorem. Thus, impossibility in one translates immediately into impossibility in the other. This is one illustration of the unifying power of linear programming – based arguments.

The chapter closes with the revelation principle of mechanism design. Readers with prior exposure to mechanism design can skip this portion of the chapter without loss.

### Chapter 3

As noted in Chapter 2, attention in the remainder of this monograph is directed to the case when utilities are quasilinear. The incentive-compatibility constraints in this case turn out to be dual to the problem of finding a shortest-path in a suitable network. This chapter introduces basic properties of the problem of finding a shortest path in a network. In fact, a problem that is more general is considered: minimum cost network flow. The analysis of this problem is not much more elaborate than needed for the shortest-path problem. Because the minimum cost flow problem arises in other economic settings, the extra generality is worth it.

**Chapter 4**

This chapter applies the machinery of Chapter 3 to provide a characterization of allocation rules that can be implemented in dominant as well as Bayesian incentive-compatible strategies. In addition, a general form of the Revenue Equivalence Theorem is obtained.

**Chapter 5**

The focus in this chapter is on mechanisms that implement the efficient outcome. The celebrated Vickrey-Clarke-Groves mechanism is derived using the results from Chapter 4. Particular attention is devoted to indirect implementations of the Vickrey-Clarke-Groves scheme in the context of combinatorial auctions. Such indirect mechanisms have an interpretation as primal-dual algorithms for an appropriate linear programming problem.

**Chapter 6**

This chapter applies the machinery of linear programming to the problem of optimal mechanism design. Not only are some of the classical results duplicated, but some new results are obtained, illustrating the usefulness of the approach.

**Chapter 7**

The subject of this brief chapter has no apparent connection to mechanism design, but it is relevant. It considers an inverse question: Given observed choices from a menu, what can we infer about preferences? Interestingly, the same mathematical structure inherent in the study of incentives appears here.

Cambridge University Press
978-0-521-17946-1 - Mechanism Design: A Linear Programming Approach
Rakesh V. Vohra
Excerpt
More information

CHAPTER 2

# Arrow's Theorem and Its Consequences

By custom and tradition, accounts of mechanism design begin with a genuflection in the direction of Kenneth Arrow and his (im)possibility theorem.[1] The biblical Mas-Collel, Whinston, and Green (1995), for example, introduce mechanism design by reminding the reader of Arrow's theorem, introduced some chapters earlier. Weight of history aside, there is no logical reason for this. Not disposed to being bolshy, I bow to precedent and begin with an account of Arrow's theorem. Whereas the conceptual connection to mechanism design is tenuous, the mathematical connection, as the linear programming approach reveals, is remarkably close.[2]

The environment considered involves a set $\Gamma$ of alternatives (at least three). Let $\Sigma$ denote the set of all strict preference orderings, that is, permutations over $\Gamma$.[3] The set of admissible preference orderings or **preference domain** for a society of $n$-agents will be a subset of $\Sigma$ and denoted $\Omega$. Let $\Omega^n$ be the set of all $n$-tuples of preferences from $\Omega$, called **profiles**.[4] An element of $\Omega^n$ will typically be denoted as $\mathbf{P} = (\mathbf{p_1}, \mathbf{p_2}, \ldots, \mathbf{p_n})$, where $\mathbf{p_i}$ is interpreted as the preference ordering of agent $i$.

The objective is to identify for each profile $\mathbf{P}$ a strict preference ordering that will summarize it – a "median" or "mean" preference ordering, if you will.[5] The rule for summarizing a profile is called a social welfare function. Formally, an $n$-person social welfare function is a function $f : \Omega^n \mapsto \Sigma$. Thus for any $\mathbf{P} \in \Omega^n$, $f(\mathbf{P})$ is an ordering of the alternatives. We write $x f(\mathbf{P}) y$ if $x$ is ranked above $y$ under $f(\mathbf{P})$.

There are many social welfare functions that one could imagine. One could list each one and examine its properties. To avoid this botanical exercise, Arrow suggested conditions that a social welfare function should satisfy to make it an attractive way to summarize a profile. An $n$-person **Arrovian social welfare**

[1] Arrow called it a 'possibility' theorem. His intellectual heirs added the 'im'.
[2] The treatment given here is based on Sethuraman, Teo, and Vohra (2003).
[3] The results extend easily to the case of indifference. See Sato (2006).
[4] This is sometimes called the **common preference domain**.
[5] This is not the usual motivation for the construct to be introduced, but will do for our purposes.

**function** (ASWF) on $\Omega$ is a function $f : \Omega^n \mapsto \Sigma$ that satisfies the following two conditions:

(1) **Unanimity:** If for $\mathbf{P} \in \Omega^n$ and some $x, y \in \Gamma$ we have $x\mathbf{p_i}y$ for all $i$, then $xf(\mathbf{P})y$.

(2) **Independence of Irrelevant Alternatives:** For any $x, y \in \Gamma$, suppose $\exists \mathbf{P}, \mathbf{Q} \in \Omega^n$, such that $x\mathbf{p_i}y$ if and only if $x\mathbf{q_i}y$ for $i = 1, \ldots, n$. Then $xf(\mathbf{P})y$ if and only if $xf(\mathbf{Q})y$.

The first axiom is uncontroversial. It stipulates that if all agents prefer alternative $x$ to alternative $y$, then the social welfare function $f$ must rank $x$ above $y$. The second axiom states that the ranking of $x$ and $y$ by $f$ is not affected by how the agents rank the other alternatives. This is not a benign axiom. Much bile and ink had been spent debating its merits. The reader interested in philosophical diversions on this matter can refer to Saari (2003).

A social welfare function that satisfies the two conditions is the *dictatorial rule*: rank the alternatives in the order of the preferences of a particular agent (the dictator). Formally, an ASWF is **dictatorial** if there is an $i$ such that $f(\mathbf{P}) = \mathbf{p_i}$ for all $\mathbf{P} \in \Omega^n$. Clearly, the dictatorial rule is far from ideal as a rule for summarizing a profile.

An ordered pair $x, y \in \Gamma$ is called **trivial** if $x\mathbf{p}y$ for all $\mathbf{p} \in \Omega$. In view of unanimity, any ASWF must have $xf(\mathbf{P})y$ for all $\mathbf{P} \in \Omega^n$ whenever $x, y$ is a trivial pair. If $\Omega$ consists only of trivial pairs, then distinguishing between dictatorial and nondictatorial ASWF's becomes nonsensical, so we assume that $\Omega$ contains at least one nontrivial pair. The domain $\Omega$ is **Arrovian** if it admits a nondictatorial ASWF.

The goal is to derive an integer linear programming formulation of the problem of finding an $n$-person ASWF. For each $\Omega$, a set of linear inequalities is identified with the property that every feasible 0-1 solution corresponds to an $n$-person ASWF.[6] By examining the inequalities, we should be able to determine whether a given domain $\Omega$ is Arrovian.

## 2.1 THE INTEGER PROGRAM

Denote the set of all ordered pairs of alternatives by $\Gamma^2$. Let $E$ denote the set of all agents, and $S^c$ denote $E \setminus S$ for all $S \subseteq E$.

To construct an $n$-person ASWF, we exploit the independence of irrelevant alternatives, condition. The condition allows one to specify an ASWF in terms of which ordered pair of alternatives a particular subset, $S$, of agents is decisive over. A subset $S$ of agents is **decisive for** $x$ **over** $y$ with respect to the ASWF $f$, if whenever all agents in $S$ rank $x$ over $y$ and all agents in $S^c$ rank $y$ over $x$, the ASWF $f$ ranks $x$ over $y$.[7]

[6] The formulation is an extension of the **decomposability** conditions identified by Kalai and Muller (1977).

[7] In the literature, this is called *weakly* decisive.

For each nontrivial element $(x, y) \in \Gamma^2$, we define a 0-1 variable as follows:

$$d_S(x, y) = \begin{cases} 1, & \text{if the subset } S \text{ of agents is decisive for } x \text{ over } y; \\ 0, & \text{otherwise.} \end{cases}$$

If $(x, y) \in \Gamma^2$ is a trivial pair, then by default we set $d_S(x, y) = 1$ for all $S \neq \emptyset$.[8]

To each ASWF $f$, we can associate $d$ variables that can be determined as follows: for each $S \subseteq E$, and each nontrivial pair $(x, y)$, pick a $\mathbf{P} \in \Omega^n$ in which agents in $S$ rank $x$ over $y$, and agents in $S^c$ rank $y$ over $x$; if $xf(\mathbf{P})y$, set $d_S(x, y) = 1$, else set $d_S(x, y) = 0$.

The remainder of this section identifies conditions satisfied by the $d$ variables associated with an ASWF $f$.

**Unanimity**: To ensure unanimity, for all $(x, y) \in \Gamma^2$, we must have

$$d_E(x, y) = 1. \tag{2.1}$$

**Independence of Irrelevant Alternatives**: Consider a pair of alternatives $(x, y) \in \Gamma^2$, a $\mathbf{P} \in \Omega^n$, and let $S$ be the set of agents that prefer $x$ to $y$ in $\mathbf{P}$. (Thus, each agent in $S^c$ prefers $y$ to $x$ in $\mathbf{P}$.) Suppose $xf(\mathbf{P})y$. Let $\mathbf{Q}$ be any other profile such that all agents in $S$ rank $x$ over $y$ and all agents in $S^c$ rank $y$ over $x$. By the independence of irrelevant alternatives, condition $xf(\mathbf{Q})y$. Hence, the set $S$ is decisive for $x$ over $y$. However, if $yf(\mathbf{P})x$, a similar argument would imply that $S^c$ is decisive for $y$ over $x$. Thus, for all $S$ and nontrivial $(x, y) \in \Gamma^2$, we must have

$$d_S(x, y) + d_{S^c}(y, x) = 1. \tag{2.2}$$

A consequence of equations (2.1) and (2.2) is that $d_\emptyset(x, y) = 0$ for all $(x, y) \in \Gamma^2$.

**Transitivity**: To motivate the next class of constraints, consider majority rule. Suppose the number of agents is odd. Majority rule ranks alternative $x$ above alternative $y$ if a strict majority of the agents prefer $x$ to $y$. Thus, majority rule can be described using the following variables:

$$d_S(x, y) = \begin{cases} 1, & \text{if } \quad |S| > n/2, \\ 0, & \text{otherwise.} \end{cases}$$

These variables satisfy equations (2.1) and (2.2). However, if $\Omega$ admits a **Condorcet** triple (e.g., $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \in \Omega$ with $x\mathbf{p}_1 y\mathbf{p}_1 z$, $y\mathbf{p}_2 z\mathbf{p}_2 x$, and $z\mathbf{p}_3 x\mathbf{p}_3 y$), then such a rule does not always return an element of $\Sigma$ for each preference profile. The reader can verify that applying majority rule to a three-agent

[8] To accommodate indifferences in preferences as well as the social ordering, Sato (2006) proposes a modification of the decision variables. For each $S \subseteq E$ of agents and each $(x, y) \in \Gamma^2$, $d_S(x, y) = 1$ is interpreted to mean that in any profile where all agents in $S$ prefer $x$ to $y$ or are indifferent between them and all agents in $E \setminus S$ prefer $y$ to $x$ or are indifferent between them, then $x$ is, socially, at least as preferred as $y$.

Cambridge University Press
978-0-521-17946-1 - Mechanism Design: A Linear Programming Approach
Rakesh V. Vohra
Excerpt
More information



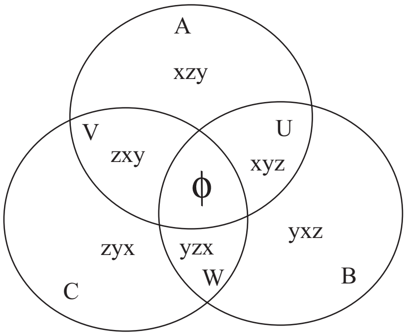


Figure 2.1 The sets and the associated orderings.

profile corresponding to a Condorcet triple does not return an ordering. The next constraint (**cycle elimination**) excludes this and similar possibilities.

For each triple $x$, $y$, $z$ and partition of the agents in up to six sets, $A$, $B$, $C$, $U$, $V$, and $W$,

$$d_{A\cup U\cup V}(x, y) + d_{B\cup U\cup W}(y, z) + d_{C\cup V\cup W}(z, x) \leq 2, \tag{2.3}$$

where the sets satisfy the following conditions (hereafter referred to as conditions [*]):

$A \neq \emptyset$ only if there exists $\mathbf{p} \in \Omega$, $x\mathbf{p}z\mathbf{p}y$,
$B \neq \emptyset$ only if there exists $\mathbf{p} \in \Omega$, $y\mathbf{p}x\mathbf{p}z$,
$C \neq \emptyset$ only if there exists $\mathbf{p} \in \Omega$, $z\mathbf{p}y\mathbf{p}x$,
$U \neq \emptyset$ only if there exists $\mathbf{p} \in \Omega$, $x\mathbf{p}y\mathbf{p}z$,
$V \neq \emptyset$ only if there exists $\mathbf{p} \in \Omega$, $z\mathbf{p}x\mathbf{p}y$,
$W \neq \emptyset$ only if there exists $\mathbf{p} \in \Omega$, $y\mathbf{p}z\mathbf{p}x$.

The constraint ensures that on any profile $\mathbf{P} \in \Omega^n$, the ASWF $f$ does not produce a ranking that "cycles."

Subsequently we prove that constraints (2.1–2.3) are both necessary and sufficient for the characterization of $n$-person ASWF's. Before that, it will be instructive to develop a better understanding of constraints (2.3), and their relationship to the constraints identified in Kalai and Muller (1977), called **decisiveness implications**, described later in the chapter.

Suppose there are $\mathbf{p}$, $\mathbf{q} \in \Omega$ and three alternatives $x$, $y$ and $z$ such that $x\mathbf{p}y\mathbf{p}z$ and $y\mathbf{q}z\mathbf{q}x$. Then,

$$d_S(x, y) = 1 \Rightarrow d_S(x, z) = 1,$$

and

$$d_S(z, x) = 1 \Rightarrow d_S(y, x) = 1.$$

The first implication follows from using a profile $\mathbf{P}$ in which agents in $S$ rank $x$ over $y$ over $z$ and agents in $S^c$ rank $y$ over $z$ over $x$. If $S$ is decisive for $x$