Mechanism Theory

Matthew O. Jackson*

Humanities and Social Sciences 228-77, California Institute of Technology

Pasadena, California 91125, U.S.A.

October 12, 2000, revised December 8, 2003

An abridged version of this appears in *Optimization and Operations Research*, edited by Ulrich Derigs, in the *Encyclopedia of Life Support Systems*, EOLSS Publishers: Oxford UK,[http://www.eolss.net], 2003.

Keywords: Mechanism, Mechanism Design, Dominant Strategy, Public Goods, Auction, Bargaining, Bayesian Equilibrium, Bayesian Incentive Compatibility, Revelation Principle, Efficiency, Individual Rationality, Balance, Strategy-Proof, Direct Mechanism, Social Choice Function, Single-Peaked Preferences, Implementation.

Contents:

*This survey is written to conform to the strict textbook-like style of the EOLSS. An addendum is attached to this working paper version with a more comprehensive bibliography and some bibliographic notes. Some of the contents of this paper are based on lectures Jackson gave at the CORE-Francqui Summer School in May and June of 2000. He thanks the organizers Claude d'Aspremont, Michel Le Breton and Heracles Polemarchakis, as well as CORE and the Francqui Foundation, and the participants of the summer school for their support and feedback. He also thanks Ulrike Ervig, Benny Moldovanu, and In-Uck Park for comments on earlier drafts.

Glossary:

Balanced Transfers: Transfers that sum to zero across individuals in all states.

Bayesian Equilibrium: A specification of individual strategies as a function of information such that no individual can gain by a unilateral change of strategies.

Bayesian Incentive Compatibility: A property of a direct mechanism requiring that truth be a Bayesian equilibrium.

Direct Mechanism: A mechanism where the message space is the type space and the outcome function is the social choice function.

Dominant Strategy: A strategy that is a best regardless of the strategies chosen by other individuals.

Efficiency: A requirement that a social decision maximize the sum of utilities of the individuals in a society.

Implementation: A mechanism implements a social choice function if its equilibrium outcomes correspond to the outcomes of the social choice function for each vector of types.

Individual Rationality Constraint: A requirement that each individual weakly prefers participation in a mechanism to not participating.

Mechanism: A specification of a message space for each individual and an outcome function that maps vectors of messages into social decisions and transfers.

Revelation Principle: To each mechanism and equilibrium one can associate a direct mechanism that is incentive compatible.

Single-Peaked Preferences: When alternatives are ordered on a line, these are preferences such that there is a unique most preferred alternative and alternatives get worse as one moves away from the peak.

Social Choice Function: A function mapping types into social decisions and transfers.

Transfer Function: A specification of transfers across individuals as a function of the type vector.

Type: The private information held by an individual relating to preferences of that individual.

Summary: Some of the basic results and insights of the literature on mechanism design are presented. In that literature game theoretic reasoning is used to model social institutions as varied as voting systems, auctions, bargaining protocols, and methods for deciding on public projects. A theme that comes out of the literature is the difficulty of finding mechanisms compatible with individual incentives that simultaneously result in efficient decisions (maximizing total welfare), the voluntary participation of the individuals, and balanced transfers (taxes and subsidies that net to zero across individuals). This is explored in the context of various incentive compatibility requirements, public and private goods settings, small and large societies, and forms of private information held by individuals.

1 Introduction

The design of the institutions through which individuals interact can have a profound impact on the results of that interaction. For instance, whether an auction is conducted with sealed bids versus oral ascending bids can have an impact on what bidders learn about each other's valuations and ultimately how they bid. Different methods of queuing jobs and charging users for computer time can affect which jobs they submit and when they are submitted. The way in which costs of a public project are spread across a society can affect the decision of whether or not the project is undertaken.

The theory of mechanism design takes a systematic look at the design of institutions and how these affect the outcomes of interactions. The main focus of mechanism design is on on the design of institutions that satisfy certain objectives, assuming that the individuals interacting through the institution will act strategically and may hold private information that is relevant to the decision at hand. In bargaining between a buyer and a seller, the seller would like to act as if the item is very costly thus raising the price, and the buyer would like to pretend to have a low value for the object to keep the price down. One question is whether one can design a mechanism through which the bargaining occurs (in this application, a bargaining protocol) to induce efficient trade of the good - so that successful trade occurs whenever the buyer's valuation exceeds that of the seller. Another question is whether there exists such a mechanism so that the buyer and seller voluntarily participate in the mechanism.

The mechanism design literature models the interaction of the individuals using game theoretic tools, where the institutions governing interaction are modeled as mechanisms. In a mechanism each individual has a message (or strategy) space and decisions result as a function of the messages chosen. For instance, in an auction setting the message space would be the possible bids that can be submitted and the outcome function would specify who gets the object and how much each bidder pays as a function of the bids submitted. Different sorts of assumptions can be examined concerning how individuals choose messages as a function of their private information, and the analysis can be applied to a wide variety of contexts. The analysis also allows for transfers to be made among the individuals, so that some might be taxed and others subsidized (as a function of their private information) to bring their incentives into alignment.

A theme that comes out of the literature is that it is often impossible to find mechanisms compatible with individual incentives that simultaneously result in efficient decisions (maximizing total welfare), the voluntary participation of the individuals, and balanced transfers (taxes and subsidies that always net out across individuals). Nevertheless, there are some important settings where incentives and efficiency are compatible and in other settings a "second-best" analysis is still possible. This is described in detail in what follows in the context of different incentive compatibility requirements, public and private goods settings, small and large societies, and forms of private information held by individuals.

2 A General Mechanism Design Setting

Individuals

A finite group of *individuals* interact. This set is denoted $N = \{1, 2, \ldots, n\}$ and generic individuals are represented as i, j, and k.

Decisions

The set of potential social *decisions* is denoted D, and generic elements are represented as d and d'.

The set of decisions may be finite or infinite depending on the application.

Preferences and Information

Individuals hold private information. Individual i's information is represented by a type θ_i which lies in a set Θ_i. Let $\theta = (\theta_1, \ldots, \theta_n)$ and $\Theta = \times_i \Theta_i$.

Individuals have preferences over decisions that are represented by a utility function $v_i : D \times \Theta_i \to I\!R$. So, $v_i(d, \theta_i)$ denotes the benefit that individual i of type $\theta_i \in \Theta_i$ receives from a decision $d \in D$. Thus, $v_i(d, \theta_i) > v_i(d', \theta_i)$ indicates that i of type θ_i prefers decision d to decision d'.

The fact that i's preferences depend only on θ_i is commonly referred to as being a case of *private values*. In private values settings θ_i represents information about i's preferences over the decisions. More general situations are discussed in section 4.7 below.

Example 1 *A Public Project.*

A society is deciding on whether or not to build a public project at a cost c. For example the project might be a public swimming pool, a public library, a park, a defense system, or any of many public goods. The cost of the public project is to be equally divided equally. Here $D = \{0, 1\}$ with 0 representing not building the project and 1 representing building the project.

Individual i's value from use of the public project is represented by θ_i. In this case, i's net benefit is 0 from not having a project built and $\theta_i - \frac{c}{n}$ from having a project built. The utility function of i can then be represented as

$$v_i(d, \theta_i) = d\theta_i - d\frac{c}{n}.$$

Example 2 *A Continuous Public Good Setting.*

In Example 1 the public project could only take two values: being built or not. There was no question about its scale. It could be that the decision is to choose a scale of a public project, such as how large to make a park, and also to choose an allocation of the costs. Let $y \in I\!R_+$ denote the scale of the public project and $c(y)$ denote the total cost of the project as it depends on the scale. Here $D = \{(y, z_1, \ldots, z_n) \in I\!R_+ \times I\!R^n \mid \sum_i z_i = c(y)\}$.

Example 3 *Allocating a Private Good.*

An indivisible good is to be allocated to one member of society. For instance, the rights to an exclusive license are to be allocated or an enterprise is to be privatized. Here, $D = \{d \in \{0,1\}^n \ : \ \sum_i d_i = 1\}$, where $d_i = 1$ denotes that individual i gets the object. If individual i is allocated the object, then i will benefit by an amount θ_i, so $v_i(d, \theta_i) = d_i\theta_i$.

Clearly, there are many other examples that can be accommodated in the mechanism design analysis as the formulation of the space D has no restrictions.

Decision Rules and Efficient Decisions

It is clear from the above examples that the decision a society would like to make will depend on the θ_i's. For instance, a public project should only be built if the total value it generates exceeds its cost.

A *decision rule* is a mapping $d : \Theta \to D$, indicating a choice $d(\theta) \in D$ as a function of θ.

A decision rule $d(\cdot)$ is *efficient* if

$$\sum_i v_i(d(\theta), \theta_i) \geq \sum_i v_i(d', \theta_i)$$

for all θ and $d' \in D$.[1]

[1] This notion of efficiency takes an ex-post perspective. That is, it looks at comparisons given that that θ_i's are already realized, and so may ignore improvements that are obtainable due to risk sharing in applications where the d's may involve some randomization.

This notion of efficiency looks at maximization of total value and then coincides with Pareto efficiency only when utility is transferable across individuals. Transferability is the case treated in most of the literature.

In the public project example (Example 1), the decision rule where $d(\theta) = 1$ when $\sum_i \theta_i > c$ and $d(\theta) = 0$ when $\sum_i \theta_i < c$ (and any choice at equality) is efficient.

Transfer Functions

In order to provide the incentives necessary to make efficient choices, it may be necessary to tax or subsidize various individuals. To see the role of such transfers, consider the example of the public project above. Any individual i for whom $\theta_i < \frac{c}{n}$ would rather not see the project built and any individual for whom $\theta_i > \frac{c}{n}$ would rather not see the project built. Imagine that the government simply decides to poll individuals to ask for their θ_i's and then build the project if the sum of the announced θ_i's is greater than c. This would result in an efficient decision if the θ_i's were announced truthfully. However, individuals with $\theta_i < \frac{c}{n}$ have an incentive to underreport their values and say they see no value in a project, and individuals for whom $\theta_i > \frac{c}{n}$ have an incentive to overreport and say that they have a very high value from the project. This could result in a wrong decision.[2] To get a truthful revelation of the θ_i's, some adjustments need to be made so that individuals are taxed or subsidized based on the announced θ_i's and individuals announcing high θ_i's expect to pay more.

Adjustments are made by a transfer function $t : \Theta \to I\!R^n$. The function $t_i(\theta)$ represents the payment that i receives (or makes if it is negative) based on the announcement of types θ.

Social Choice Functions

A pair d, t will be referred to as a *social choice function*, and at times denoted by f. So, $f(\theta) = (d(\theta), t(\theta))$.

The utility that i receives if $\widehat{\theta}$ is the "announced" vector of types (that operated on by $f = (d, t)$) and i's true type is θ_i is

$$u_i(\widehat{\theta}, \theta_i, d, t) = v_i(d(\widehat{\theta}), \theta_i) + t_i(\widehat{\theta}).$$

This formulation of preferences is said to be *quasi-linear*.

Feasibility and Balance

[2]Similarly, if the decision is simply made by a majority vote of the population, the number who vote yes will simply be the number for whom $\theta_i > \frac{c}{n}$. This can easily result in not building the project when it is socially efficient, or building it when it is not socially efficient.

A transfer function t is said to be *feasible* if $0 \geq \sum_i t_i(\theta)$ for all θ.

If t is not feasible then it must be that transfers are made into the society from some outside source. If the t is feasible, but results in a sum less than zero in some circumstances, then it generates a surplus which would either have to be wasted or returned to some outsider.[3]

A transfer function t is *balanced* if $\sum_i t_i(\theta) = 0$ for all θ.

Balance is an important property if we wish the full (d, t) pair to be efficient rather than just d. If $\sum_i t_i < 0$, then there is some net loss in utility to society relative to an efficient decision with no transfers.

Mechanisms

A *mechanism* is a pair M, g, where $M = M_1 \times \cdots \times M_n$ is a cross product of message or strategy spaces and $g : M \to D \times I\!R^n$ is an outcome function. Thus, for each profile of messages $m = (m_1, \ldots, m_n)$, $g(m) = (g_d(m), g_{t,1}(m), \ldots, g_{t,n}(m))$ represents the resulting decision and transfers.

A mechanism is often also referred to as a *game form*. The terminology game form distinguishes it from a game (see game theory), as the consequence of a profile of messages is an outcome rather than a vector of utility payoffs. Once the preferences of the individuals are specified, then a game form or mechanism induces a game. Since in the mechanism design setting the preferences of individuals vary, this distinction between mechanisms and games is critical.

3 Dominant Strategy Mechanism Design

The mechanism design problem is to design a mechanism so that when individuals interact through the mechanism, they have incentives to choose messages as a function of their private information that leads to socially desired outcomes. In order to make predictions of how individuals will choose messages as a function of their private information, game theoretic reasoning is used (see game theory). We start, as much of the literature on mechanism design did, by looking at the notion of dominant strategies, which identifies situations in which individuals have unambiguously best strategies (messages).

[3] It is important that the surplus not be returned to the society. If it were returned to the society, then it would result in a different transfer function and different incentives.

3.1 Dominant Strategies

A strategy $m_i \in M_i$ is a *dominant strategy* at $\theta_i \in \Theta_i$, if

$$v_i(g_d(m_{-i}, m_i), \theta_i) + g_{t,i}(m_{-i}, m_i) \geq v_i(g_d(m_{-i}, \widehat{m}_i), \theta_i) + g_{t,i}(m_{-i}, \widehat{m}_i)$$

for all m_{-i} and $\widehat{m}_i$.

A dominant strategy has the strong property that it is optimal for a player no matter what the other players do. When dominant strategies exist, they provide compelling predictions for strategies that players should employ. However, the strong properties required of dominant strategies limits the set of situations where they exist.

A social choice function $f = (d, t)$ is *implemented* in dominant strategies by the mechanism (M, g) if there exist functions $m_i : \Theta_i \to M_i$ such that $m_i(\theta_i)$ is a dominant strategy for each i and $\theta_i \in \Theta_i$ and $g(m(\theta)) = f(\theta)$ for all $\theta \in \Theta$.

3.2 Direct Mechanisms and the Revelation Principle

Note that a social choice function $f = (d, t)$ can be viewed as a mechanism, where $M_i = \Theta_i$ and $g = f$. This is referred to as a *direct mechanism*.

A direct mechanism (or social choice function) $f = (d, t)$ is *dominant strategy incentive compatible* if θ_i is a dominant strategy at θ_i for each i and $\theta_i \in \Theta_i$. A social choice function is also said to be *strategy-proof* if it is dominant strategy incentive compatible.

The usefulness of the class of direct mechanisms as a theoretical tool in mechanism design is a result of the well-known, simple, and yet powerful *revelation principle*.

The Revelation Principle for Dominant Strategies: If a mechanism (M, g) implements a social choice function $f = (d, t)$ in dominant strategies, then the direct mechanism f is dominant strategy incentive compatible.

The Revelation Principle follows directly from noting that $f(\theta) = g(m(\theta))$ for each θ. The powerful implication of the revelation principle is that if we wish to find out the social choice functions can implemented in dominant strategies, we can restrict our attention to the set of direct mechanisms.

3.3 The Gibbard-Satterthwaite Theorem

Given that the specification of the space of decisions D can be quite general, it can keep track of all the aspects of a decision that are salient to a society. Thus, the transfer functions t are an extra that may be needed to provide correct incentives, but might best be avoided if possible. So, we start by exploring the set of decisions that can be implemented in dominant strategies without having to resort to transfers (beyond any that society already wished to specify inside the decisions), or in other words with t set to 0.

A decision rule d is dominant strategy incentive compatible (or strategy- proof) if the social choice function $f = (d, t^0)$ is dominant strategy incentive compatible, where t^0 is the transfer function that is identically 0.

A decision rule d is *dictatorial* if there exists i such that $d(\theta) \in \text{argmax}_{d \in R_d} v_i(d, \theta_i)$ for all θ, where $R_d = \{d \in D \mid \exists \theta \in \Theta : d = d(\theta)\}$ is the range of d.

Theorem 1 *Suppose that D is finite and type spaces include all possible strict orderings over D.*[4] *A decision rule with at least three elements in its range is dominant strategy incentive compatible (strategy-proof) if and only if it is dictatorial.*

The condition that type spaces allow for all possible strict orderings over D, is quite natural in situations such as when the set of decisions is a set of candidates, one of whom is to be chosen to represent or govern the society. But this condition may not be appropriate in settings where the decisions include some allocation of private goods and individuals each prefer to have more of the private good, as in an auction setting.

The Gibbard-Satterthwaite theorem has quite negative implications for the hopes of implementing non-trivial decision rules in dominant strategies in a general set of environments. It implies that transfer functions will be needed for dominant strategy implementation of non-dictatorial decision rules in some settings. Before discussing the role of transfer functions, let us point out some prominent settings where the preferences do not satisfy the richness of types assumption of the Gibbard-Satterthwaite theorem and there exist non-dictatorial strategy-proof social choice functions that do not rely on transfer functions.

3.4 Single-Peaked Preferences and Other Restricted Domains

Consider a setting where decisions fall on a single dimension, say $D \subset I\!R$.

Individuals have *single-peaked* preferences over D if for each i and $\theta_i \in \Theta_i$ there exists $p(\theta_i) \in D$, called the *peak* of i's preferences, such that $p(\theta_i) \geq d > d'$ or

[4] For any ordering $h : D \to \{1, \ldots, \#D\}$ (where h is onto) of elements of D and $i \in N$ there exists a type $\theta_i \in \Theta_i$ such that $v_i(d, \theta_i) < v_i(d', \theta_i)$ when $h(d) < h(d')$.

$d' > d \geq p(\theta_i)$ imply that

$$v_i(d, \theta_i) > v_i(d', \theta_i).$$

Single peaked preference domains are used in modeling in some voting and political science applications such that a "left to right" interpretation is appropriate. A single peaked preference domain can also arise as by-product in other domains, as discussed below.

In a single peaked preference domain there are dominant strategy incentive compatible decision rules that have quite nice properties. For instance having each individual declare their peak and then selecting the median (with a tie-break in the case of an even number of individuals) results in truthful announcements of peaks as a dominant strategy. Moreover, such a median voting rule is anonymous, Pareto efficient[5], and immune to manipulation of announced peaks even by coalitions of individuals. The same is true of variations on the median voting rule, such as taking the maximum of the peaks, the minimum, or any order statistic. It turns out that in a single-peaked setting the class of all anonymous and strategy-proof decision rules that are onto D are the class of "phantom voting rules" discovered by Hervé Moulin. These rules are described as follows. First $n-1$ "phantom" peaks in D are fixed.[6] Next, individuals declare their n peaks and the median is taken over the set of $2n-1$ peaks. A rich variety of rules is obtained by varying the location of the phantom peaks.

Although single-peaked preference domains are described above in the context of a left to right political choice, such settings can arise even in private good settings. For instance, consider 2 individuals in a two good classical exchange economy. A price for trade is fixed at $p > 0$ units of good 1 for good 2, and each individual starts with an endowment $e_i \in I\!R^2_+$ of the goods. Here a decision is a scalar $d \in I\!R$ such that $e_1 + (-pd, d) \in I\!R^2_+$ and $e_2 + (pd, -d) \in I\!R^2_+$. Thus, d represents the amount of good 2 that individual 1 buys from individual 2 at price p, and must be feasible given the endowments in the society. If each individual has preferences that are that

[5]It results in a decision such that there is no other decision that is improving for all individuals. Pareto efficiency is a weaker condition than the efficiency defined previously. Median voting will not generally result in a decision that maximizes the sum of utilities. If one wants to implement efficient decisions in this stronger sense, then transfers are needed even in the context of single-peaked preference domains.

[6]If D is not compact, then add points $-\infty$ and ∞ as possible points for the phantoms. Then, for example, to get the minimum peak rule, set the phantoms at the lowest point and then the median of the phantom peaks and the n peaks of the individuals will always be the lowest peak announced by an individual.

are continuous, strictly convex, and increasing over their final consumption,[7] are then their utility functions $v_i(d, \theta_i)$ are single-peaked over d. A strategy-proof decision rule in this setting is to median vote over d with a phantom voter at 0. It turns out that the class of all strategy-proof decision rules in exchange economies (with any number of agents and goods) satisfying an anonymity, individual rationality, and additional technical condition, are of a similar form, as shown by Salvador Barbera and Matthew Jackson. In particular, goods must be traded in fixed proportions, where individuals declare their preferred trades in these proportions.

3.5 Groves' Schemes

While there are settings where dominant strategy incentive compatible decision rules exist, even in those settings implementation of an efficient decision rule may require transfers. Moreover, there are many settings where the only hope for dominant strategy incentive compatibility is in the use of transfers.

The first approach we follow takes that of most of the literature (we will come back to another approach below). It starts with some efficient decision rule d and then asks what form of transfers are necessary so that d, t is dominant strategy incentive compatible. The resulting social choice functions are referred to as Groves' schemes after Theodore Groves, who first pointed out this full class of dominant strategy incentive compatible social choice functions with efficient decisions. The first part of the following theorem is due to Groves, while the converse was first established by Jerry Green and Jean-Jacques Laffont.

Theorem 2 *(I) If d be an efficient decision rule and for each i there exists a function $x_i : \times_{j \neq i} \Theta_j \to I\!R$ such that*

$$t_i(\theta) = x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta), \theta_j), \tag{1}$$

then (d, t) is dominant strategy incentive compatible.

(II) Conversely, if d is an efficient decision rule, (d, t) is dominant strategy incentive compatible, and the type spaces are complete in the sense that $\{v_i(\cdot, \theta_i) \mid \theta_i \in \Theta_i\} = \{v : D \to I\!R\}$ for each i, then for each i there exists a function $x_i : \times_{j \neq i} \Theta_j \to I\!R$ such that the transfer function t satisfies (1).

[7] Preferences are continuous if upper and lower contour sets are closed, strictly convex if upper contour sets are convex, and increasing if $c \geq c'$ and $c \neq c'$ (where $c \in I\!R^2_+$ is the vector of goods consumed by i) implies that c is preferred to c'.

As this is one of the fundamental results in mechanism design theory, and the proof is relatively easy and instructive, let us go through the proof.

Proof of Theorem 2: Let us first show (I). Suppose to the contrary that d is an efficient decision rule and for each i there exists a function $x_i : \times_{j \neq i} \Theta_j \to I\!R$ such that the transfer function t satisfies (1), but that (d, t) is not dominant strategy incentive compatible. Then there exists i, θ and $\widehat{\theta}_i$ such that

$$v_i(d(\theta_{-i}, \widehat{\theta}_i), \theta_i) + t_i(\theta_{-i}, \widehat{\theta}_i) > v_i(d(\theta), \theta_i) + t_i(\theta).$$

From (1) this implies that

$$v_i(d(\theta_{-i}, \widehat{\theta}_i), \theta_i) + x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta_{-i}, \widehat{\theta}_i), \theta_j) > v_i(d(\theta), \theta_i) + x_i(\theta_{-i}) + \sum_{j \neq i} v_j(d(\theta), \theta_j),$$

or that

$$v_i(d(\theta_{-i}, \widehat{\theta}_i), \theta_i) + \sum_{j \neq i} v_j(d(\theta_{-i}, \widehat{\theta}_i), \theta_j) > v_i(d(\theta), \theta_i) + \sum_{j \neq i} v_j(d(\theta), \theta_j).$$

This contradicts the efficiency of d and so our supposition was incorrect.

Next, let us show that (II) holds. Let d is an efficient decision rule, (d, t) is dominant strategy incentive compatible, and the type spaces are complete. Note that there exists a function $x_i : \Theta \to I\!R$ for each i such that

$$t_i(\theta) = x_i(\theta) + \sum_{j \neq i} v_j(d(\theta), \theta_j).$$

We need only show that x_i is independent of θ_i. Suppose to the contrary, that there exists i, θ and $\widehat{\theta}_i$ such that (without loss of generality) $x_i(\theta) > x_i(\theta_{-i}, \widehat{\theta}_i)$. Let $\varepsilon = \frac{1}{2}[x_i(\theta) - x_i(\theta_{-i}, \widehat{\theta}_i)]$. By dominant strategy incentive compatibility, it follows that $d(\theta) \neq d(\theta_{-i}, \widehat{\theta}_i)$. Given the completeness of type spaces, there exists $\widetilde{\theta}_i \in \Theta_i$ such that

$$v_i(d(\theta_{-i}, \widehat{\theta}_i), \widetilde{\theta}_i) + \sum_{j \neq i} v_j(d(\theta_{-i}, \widehat{\theta}_i), \theta_j) = \varepsilon$$

and

$$v_i(d, \widetilde{\theta}_i) + \sum_{j \neq i} v_j(d, \theta_j) = 0$$

for any $d \neq d(\theta_{-i}, \widehat{\theta}_i)$. The efficiency of d together with these conditions on $\widetilde{\theta}_i$ imply that $d(\theta_{-i}, \widetilde{\theta}_i) = d(\theta_{-i}, \widehat{\theta}_i)$. Then by dominant strategy incentive compatibility it follows that $t_i(\theta_{-i}, \widetilde{\theta}_i) = t_i(\theta_{-i}, \widehat{\theta}_i)$. Thus, the utility to i from truthful announcement at $\widetilde{\theta}_i$ is

$$u_i(\theta_{-i}, \overline{\theta}_i, \overline{\theta}_i, d, t) = \varepsilon + x_i(\theta_{-i}, \widehat{\theta}_i)$$

and by lying and saying θ_i at $\widetilde{\theta}_i$, i gets

$$u_i(\theta_{-i}, \theta_i, \widetilde{\theta}_i, d, t) = x_i(\theta).$$

This contradicts dominant strategy incentive compatibility since $x_i(\theta) > \varepsilon + x_i(\theta_{-i}, \widehat{\theta}_i)$. Thus our supposition was incorrect. ∎

3.6 The Pivotal Mechanism and Vickrey Auctions

One version of the Groves schemes characterized in Theorem 2 is simple and has a nice intuition and properties. It is the pivotal mechanism that was described by Clarke (independently of Groves). Let $x_i(\theta_{-i}) = -\max_{d \in D} \sum_{j \neq i} v_j(d, \theta_j)$. In this case, i's transfer becomes

$$t_i(\theta) = \sum_{j \neq i} v_j(d(\theta), \theta_j) - \max_{d \in D} \sum_{j \neq i} v_j(d, \theta_j).$$

This transfer is always non-positive and so the pivotal mechanism is always feasible. Moreover, the transfers have a nice interpretation. If i's presence makes no difference in the maximizing choice of d, then $t_i(\theta) = 0$. Otherwise, we can think of i as being "pivotal", and then t_i represents the loss in value that is imposed on the other individuals due to the change in decision that results from i's presence in society. The pivotal mechanism then has a very simple intuition behind its incentives: each individual's transfer function takes into account the marginal social impact (on other individuals) made by his announcement of θ_i. When looking at this social impact together with his own selfish utility, the individual has exactly the total social value in mind when deciding on a strategy. This leads to efficient decision making.

The pivotal mechanism has other nice properties besides always being feasible and having a simple intuition. Hervé Moulin has shown that the pivotal mechanism is characterized among the class of mechanisms satisfying (1) as being the only one for which $u_i(\theta, \theta_i) \geq \min_{d \in D} v_i(d, \theta_i)$ for all i and $\theta \in \Theta$.

Moreover, the pivotal mechanism reduces to a well-known auction form in the context of the allocation of indivisible objects. In that context there is a simple auction that is dominant strategy incentive compatible, as first noticed by William Vickrey. It turns out that this auction form, commonly referred to as a Vickrey auction, corresponds to the pivotal mechanism in this setting. Let us explore this relationship in the case of a single good. There are n individuals who each have a valuation θ_i for the object and $d \in \{1, \ldots, n\} = D$ indicates the individual to whom the object is allocated. In that case, the efficient decision is such that $d(\theta) \in \text{argmax}_i \theta_i$. The pivotal mechanism takes an easy form. If $d(\theta) = i$ then $t_i(\theta) = -\max_{j \neq i} \theta_j$ and if $d(\theta) \neq i$ then $t_i(\theta) = 0$. This means the object is allocated to the individual with the highest valuation and he pays an amount equal to the second highest valuation. No other payments are made. Now consider a second price (Vickrey) auction, where the high bidder is awarded the object and pays the second highest price. It is easy to see that bidding one's value is a dominant strategy in this auction, as it is the same reasoning as that behind the pivotal mechanism. The pivotal mechanism and Vickrey auction implement the same social choice function.[8]

3.7 The Tension between Balance and Incentive Compatibility

Requiring efficiency of the decision is only part of the efficiency that society should desire. If the transfer functions are not balanced, then there is waste and it can be considerable. To see this problem, let us reexamine the public project example (Example 1) in a simple case.

Example 4 *Lack of Balance in a Public Goods Setting*

Consider a situation where there are two individuals ($n = 2$) and $\Theta_1 = \Theta_2 = I\!R$ - so individuals' might have any value for the public project. By Theorem 2 there exist x_1 and x_2 so that (1) is satisfied. Let the cost of the project be $c = 3/2$.

Let us investigate the requirements imposed by feasibility of the transfers. Applying feasibility when $\theta_1 = \theta_2 = 1$ implies that $0 \geq t_1(1,1) + t_2(1,1)$, Noting that $d(1,1) = 1$, equation (1) implies that

$$-\frac{1}{2} \geq x_1(1) + x_2(1).$$

Similarly, $0 \geq t_1(0,0) + t_2(0,0)$, and noting that $d(0,0) = 0$ implies that

$$0 \geq x_1(0) + x_2(0).$$

Together these imply that at least one of

$$-\frac{1}{4} \geq x_1(1) + x_2(0) \quad \text{or} \quad -\frac{1}{4} \geq x_1(0) + x_2(1)$$

[8] This also corresponds to the ascending oral or English auction, which is one of the most common auction forms used. It is a dominant strategy to remain in the bidding until the bid hits one's valuations and then to drop out. The highest valuation individual will get the object at the second highest bidder's valuation.

are satisfied. Under (1), noting that $d(0,1) = d(1,0) = 0$, this implies that at least one of

$$-\frac{1}{4} \geq t_1(1,0) + t_2(1,0) \quad \text{or} \quad -\frac{1}{4} \geq t_1(0,1) + t_2(0,1)$$

holds. However, this means that the sum of transfers is negative in at least one case. Thus, to have a dominant strategy incentive compatible mechanism with an efficient decision rule, one cannot satisfy balance.

At this point it is important to emphasize a distinction between efficient decisions and efficient social choice functions. If transfers are not balanced, then the social choice function cannot be efficient among those that are feasible. This means that overall efficiency (taking transfers into account) is incompatible with dominant strategy incentive compatibility in some settings.

The above conclusion depends on the richness of the type space. Suppose that in the above example the type spaces only admitted two valuations, $\Theta_1 = \Theta_2 = \{0, 1\}$. In that case a simple voting mechanism would induce efficient decisions and no transfers would be necessary. Each individual would cast a vote and the project would be built if both individuals vote yes and not otherwise. In such a mechanism it is a dominant strategy to vote yes if $\theta_i = 1$ and no if $\theta_i = 0$. This mechanism works in this stark situation because of the simplicity of the efficient decision rule. There is only one scenario in which the project should be built and it requires unanimous high valuations. The lack of balance with richer type spaces results from the many different scenarios over which the decision varies. Then individuals can influence the scenario in their favor from over or under-reporting their valuations. Providing the incentives to each individual at the same time to truthfully reveal their information requires that transfers be made. If one individual is taxed, those taxes may be difficult to redistribute to the other individuals without distorting their incentives, as they will have an incentive to try to increase the amount that is being redistributed.

3.8 Large Numbers and Approximate Efficiency

The balance problem discussed above can be overcome if there is some individual whose valuation for the project is either known or fixed. This individual can then serve as a residual claimant.footnote As argued above, however, it is important that the transfers not end up being redistributed to the rest of the society as this would end up distorting the incentives. In the case where there is no such agent, the balance problem can be eliminated while still retaining *approximate* efficiency in large societies. This can be done in a simple way as follows. Take an individual from the society, say individual 1. Then operate a pivotal mechanism on the society of the remaining $n-1$ individuals as if individual 1 were not present. The result will be feasible and any surplus generated can be given to individual 1. The overall mechanism will be balanced, but the decision will not be efficient. However, the per-capita loss in terms of efficiency will be at most $\max_{d,d',\theta_1}[v_1(d,\theta_1) - v_1(d',\theta_1)]/n$. If utility is bounded, this tends to 0 as n becomes large. While this obtains approximate efficiency, it still retains some difficulties in terms of individual rationality. This is discussed next.

3.9 Lack of Individual Rationality in Groves' Schemes

In addition to the balance problems that are present with trying to achieve an efficient decision with a dominant strategy mechanism, there are related problems. One is the violation of what is commonly referred to as an *individual rationality* or voluntary participation condition.[9] That requires that $v_i(d(\theta),\theta_i) + t_i(\theta) \geq 0$ for each i and θ, assuming a proper normalization of utility. As we argued above in the example above, at least one of $-\frac{1}{4} \geq t_1(0,1)+t_2(0,1)$ or $-\frac{1}{4} \geq t_1(1,0)+t_2(1,0)$ holds. Since no project is built in these cases, this implies that some individual ends up with a negative total utility. That individual would have been better off by not participating and obtaining a 0 utility.

3.10 Inefficient Decisions

Groves' schemes impose efficient decision making and then set potentially unbalanced transfers to induce incentive compatibility. An alternative approach is to impose balance and then set decision making to induce incentive compatibility. This approach was taken by Hervé Moulin and strengthened by Shigehiro Serizawa in the context of a public good decision. It results in a completely different set of social choice functions from the Groves' schemes, which is outlined here in a special case.

Let $D = [0,1] \times I\!R^n_+$ where $d = (y, z_1, \ldots, z_n)$ is such that $\sum_i z_i = cy$, where $c > 0$ is a cost per unit of production of y. The interpretation is that y is the level of public good chosen and z_1 to z_n are the allocations of cost of producing y paid by the individuals.

[9] Another problem is that Groves' schemes can be open to coalitional manipulations even though they are dominant strategy incentive compatible. For example in a pivotal mechanism individuals may be taxed even when the project is not built. They can eliminate those taxes by jointly changing their announcements.

The class of admissible preferences differs from the quasi-linear case focused on above. An individual's preferences are represented by a utility function over (d, t, θ_i) that takes the form $w_i(y, t_i - z_i, \theta_i)$ where w_i is continuous, strictly quasi-concave, and monotonic in its first two arguments, and all such functions are admitted as θ_i varies across Θ_i. In the situation where no transfers are made and costs are split equally, the resulting $w_i(y, -cy/n, \theta_i)$ is single-peaked over y, with a peak denoted $\widehat{y}_i(\theta_i)$.

Theorem 3 *In the above described public good setting, a social choice function (d, t) is balanced, anonymous,*[10] *has a full range of public good levels, and dominant strategy incentive compatible if and only if it is of the form $t_i(\theta) = 0$ for all i and $d(\theta) = (y(\theta), cy(\theta)/n, \ldots, cy(\theta)/n)$, where there exists $(p_1, \ldots, p_{n-1}) \in [0,1]^{n-1}$*

$$y(\theta) = \text{median}\left[p_1, \ldots, p_{n-1}, \widehat{y}_1(\theta_1), \ldots, \widehat{y}_n(\theta_n)\right].$$

If in addition individual rationality is required, then the mechanism must be a minimum demand mechanism($y(\theta) = \min_i \widehat{y}_i(\theta_i)$).

The intuition Theorem 3 relates back to the discussion of single-peaked preferences. The anonymity and balance conditions, together with incentive compatibility, restrict the t_i's to be 0 and the z_i's to be an even split of the cost. Given this, the preferences of individuals over choices y are then single-peaked. Then the phantom voting methods described in section 3.4 govern how the choice. The expression for $y(\theta)$ fits the phantom voting methods, where $p_1, \ldots, p_{n-1}$ are the phantoms.

The answer to the question of which approach - that of fixing decisions to be efficient and solving for transfers, or that of fixing transfers to be balanced and solving for decisions - results in "better" mechanisms is ambiguous.[11] There are preference profiles for which the minimum demand mechanism Pareto dominates the pivotal mechanism, and vice versa.

4 Bayesian Mechanism Design

Dominant strategy incentive compatibility is a very strong condition as it requires that truthful revelation of preferences be a best response, regardless of the potential

[10] A permutation of the labels of the individuals results in a corresponding permutation of the resulting decision and transfers.

[11] Kevin Roberts has provided a characterization of the set of all dominant strategy mechanisms under some conditions, and there are some mechanisms that have neither efficient decisions nor balanced transfers.

announcements of the others. Some of the limitations of the schemes outlined above are due to this strong requirement. Weaker forms of incentive compatibility may often be appropriate

Claude d'Aspremont and Louis André Gerard-Varet and independently Kenneth Arrow showed that the balance difficulties exhibited by Groves' schemes could be overcome in a setting where individuals have probabilistic beliefs over the types of other individuals. This allows us to weaken the requirement of dominant strategy incentive compatibility to a Bayesian incentive compatibility condition.

For simplicity, (as in most of the literature) assume that Θ is a finite set and that $\theta \in \Theta$ is randomly chosen according to a distribution P, where the marginal of P, observes Θ_i, has full support. Each individual knows P and θ_i and has beliefs over the other individuals' types described by Bayes' rule. To distinguish random variables from their realizations, $\overline{\theta}_i$ will denote the random variable and θ_i, θ_i' will denote realizations.

4.1 A Bayesian Revelation Principle

Following John Harsanyi, we define Bayesian equilibrium for a mechanism (M, g). A Bayesian strategy is a mapping $m_i : \Theta_i \to M_i$. [12] A profile of Bayesian strategies $m : \Theta \to M$ forms a *Bayesian equilibrium* if

$$E\left[v_i(g_d(m_{-i}(\overline{\theta}_{-i}), m_i(\theta_i)), \theta_i) + g_{t,i}(m_{-i}(\overline{\theta}_{-i}), m_i(\theta_i)) \ \mid \ \theta_i\right]$$

$$\geq E\left[v_i(g_d(m_{-i}(\overline{\theta}_{-i}), \widehat{m}_i), \theta_i) + g_{t,i}(m_{-i}(\overline{\theta}_{-i}), \widehat{m}_i) \ \mid \ \theta_i\right]$$

for each i, $\theta_i \in \Theta_i$, and $\widehat{m}_i \in M_i$.

A direct mechanism (i.e., social choice function) $f = (d, t)$ is *Bayesian incentive compatible* if truth is a Bayesian equilibrium. This is expressed as

$$E\left[v_i(d(\overline{\theta}_{-i}, \theta_i), \theta_i) + t_i(\overline{\theta}_{-i}, \theta_i) \ \mid \ \theta_i\right] \geq E\left[v_i(d(\overline{\theta}_{-i}, \theta_i'), \theta_i) + t_i(\overline{\theta}_{-i}, \theta_i') \ \mid \ \theta_i\right]$$

for all i, $\theta_i \in \Theta_i$ and $\theta_i' \in \Theta_i$.

A mechanism (M, g) *realizes* a social choice function f in Bayesian equilibrium if there exists a Bayesian equilibrium $m(\cdot)$ of (M, g) such that $g(m(\theta)) = f(\theta)$ for all θ (that occur with positive probability).

[12] Only pure strategies are treated here. To consider mixed strategies (see game theory), given the finite type spaces, simply map Θ_i into distributions over M_i.

The Revelation Principle for Bayesian Equilibrium: If a mechanism (M, g) realizes a social choice function $f = (d, t)$ in Bayesian equilibrium, then the direct mechanism f is Bayesian incentive compatible.

Again, the proof is straightforward, and the usefulness of the Bayesian version of the revelation principle parallels that of dominant strategies.

4.2 A Balanced Mechanism with Independent Types

To get a feeling for the implications of the weakening of dominant strategy incentive compatibility to that of Bayesian incentive compatibility, let us examine the case of independent types. The mechanisms of d'Aspremont and Gerard-Varet, and of Arrow, can then be expressed as follows.

Theorem 4 *If types are independent ($\overline{\theta}_{-i}$ and $\overline{\theta}_i$ are independent for each i), d is efficient, and*

$$t_i(\theta) = E\left[\sum_{j \neq i} v_j(d(\overline{\theta}), \overline{\theta}_j) \;\mid\; \theta_i\right] - \frac{1}{n-1}\sum_{k \neq i} E\left[\sum_{j \neq k} v_j(d(\overline{\theta}), \overline{\theta}_j) \;\mid\; \theta_k\right],$$

then (d, t) is Bayesian incentive compatible and t is balanced.

Theorem 4 has a converse just as Theorem 2 did. Here it is that if (d, t) is Bayesian incentive compatible, d is efficient, and t is balanced, then t is of the form above plus a function $x_i(\theta)$ such that $\sum_i x_i(\theta) = 0$ and $E[x_i(\overline{\theta}) \mid \theta_i]$ does not depend on θ_i.[13]

Proof of Theorem 4: The balance of t follows directly from its definition. Let us verify that (d, t) is Bayesian incentive compatible.
$E\left[v_i(d(\overline{\theta}_{-i}, \theta_i'), \theta_i) + t_i(\overline{\theta}_{-i}, \theta_i') \;\mid\; \theta_i\right]$

$$= E\left[v_i(d(\overline{\theta}_{-i}, \theta_i'), \theta_i) \;\mid\; \theta_i\right] + E\left[\sum_{j \neq i} v_j(d(\overline{\theta}), \overline{\theta}_j) \;\mid\; \theta_i'\right]$$

$$-\frac{1}{n-1}\sum_{k \neq i} E\left[E\left[\sum_{j \neq k} v_j(d(\overline{\theta}), \overline{\theta}_j) \;\mid\; \overline{\theta}_k\right] \;\mid\; \theta_i\right].$$

[13] Note that it is possible for $E[x_i(\overline{\theta}) \mid \theta_i]$ not to depend on θ_i and yet $x_i(\theta)$ to depend on θ. For instance, suppose that each $\Theta_k = \{-1, 1\}$ and that $x_i(\theta) = \times_k \theta_k$.

Under independence, this expression becomes

$$= E\left[v_i(d(\overline{\theta}_{-i}, \theta_i'), \theta_i) + \sum_{j \neq i} v_j(d(\overline{\theta}_{-i}, \theta_i'), \overline{\theta}_j) \ \mid \ \theta_i\right] - \frac{1}{n-1}\sum_{k \neq i} E\left[\sum_{j \neq k} v_j(d(\overline{\theta}), \overline{\theta}_j)\right].$$

The second expression is independent of the announced θ_i', and so maximizing $E\left[v_i(d(\overline{\theta}_{-i}, \theta_i'), \theta_i) + t_i(\overline{\theta}_{-i}, \theta_i') \ \mid \ \theta_i\right]$ with respect to θ_i' boils down to maximizing:

$$E\left[v_i(d(\overline{\theta}_{-i}, \theta_i'), \theta_i) + \sum_{j \neq i} v_j(d(\overline{\theta}_{-i}, \theta_i'), \overline{\theta}_j) \ \mid \ \theta_i\right].$$

Since d is efficient, this expression is maximized when $\theta_i' = \theta_i$. ∎

Note that truth remains a best response even after θ_{-i} is known to i. Thus, the incentive compatibility is robust to any leakage or sharing of information among the individuals. Nevertheless, the design of the Bayesian mechanisms outlined in Theorem 4 still requires knowledge of $E[\cdot \mid \theta_i]$'s, and so such mechanisms are sensitive to particular ways on the distribution of uncertainty in the society.

4.3 Dependent Types

The independence condition in Theorem 4 is important in providing the simple structure of the transfer functions, and is critical to the proof. Without independence, it is still possible to find efficient, balanced, Bayesian incentive compatible mechanisms in "most" settings. The extent of "most" has been made precise by d'Aspremont, Crémer, and Gerard-Varet by showing that "most" means except those where the distribution of types is degenerate in that the matrix of conditional probabilities does not have full rank, which leads to the following theorem.

Theorem 5 *Fix the finite type space Θ and the decisions D. If $n \geq 3$, then the set of probability distributions P for which there exists a Bayesian incentive compatible social choice function that has an efficient decision rule and balanced transfers, is an open and dense subset of the set of all probability distributions on Θ.*[14]

To get a feeling for how correlation can be used in structuring transfers, see Example 9 below.

[14]Given the finite set Θ, a probability distribution can be written as a finite vector and so definitions of open and dense are standard.

4.4 Individual Rationality or Voluntary Participation

An overarching theme of mechanism design is that it is costly to provide correct incentives to individuals who hold private information that is valuable in society's decision making. That cost manifests itself in various shortcomings of the mechanisms or social choice functions that are incentive compatible.

Weakening the incentive hurdle from a dominant strategy perspective to a Bayesian perspective helps. It eliminates the inefficiency that results either in decisions or unbalanced transfers that plagues dominant strategy implementation. But it does not completely overcome some of the other difficulties that are present, such as providing a mechanism that satisfies individual rationality constraints. Let us examine this issue in some detail.

There are various time perspectives that one can take on individual rationality, which correspond to different timings at which individuals become bound to a mechanism.[15]

First let us normalize the utility functions v_i so that 0 represents the value that individual i would get from not participating in the mechanism for any θ_i.

The strongest form of individual rationality constraint is that no individual wishes to walk away from a mechanism after all information has been revealed and the decision and transfers fully specifed, regardless of the realization of θ. This is called *ex-post individual rationality* and requires that

$$v_i(d(\theta), \theta_i) + t_i(\theta) \geq 0$$

for all θ and i. This was the form of individual rationality that we discussed in the dominant strategy setting, since no beliefs were specified.

A weaker form of individual rationality is that no individual wishes to walk away from the mechanism at a point where they know their own type θ_i, but only have expectations over the other individuals' types and the resulting decision and transfers. This is called *interim individual rationality* and requires that

$$E\left[v_i(d(\overline{\theta}), \theta_i) + t_i(\overline{\theta}) \;\;|\;\; \theta_i\right] \geq 0$$

for all i and $\theta_i \in \Theta_i$.

[15]The time perspective can give rise to different versions of efficiency as well. However, in terms of maximizing $\sum_i v_i$ the time perspective is irrelevant as maximizing $E[\sum_i v_i(d(\overline{\theta}), \overline{\theta}_i)]$ is equivalent to maximizing $\sum_i v_i(d(\theta), \theta_i)$ at each θ (given the finite type space). If one instead considers Pareto efficiency, then the ex-ante, interim, and ex-post perspectives are no longer equivalent.

The weakest form of individual rationality is that no individual wishes to walk away from the mechanism before they know their own type θ_i, and only have expectations over all the realizations of types and the resulting decision and transfers. This is called *ex-ante individual rationality* and requires that

$$E\left[v_i(d(\overline{\theta}), \overline{\theta}_i) + t_i(\overline{\theta})\right] \geq 0$$

for all i.

Consider any social choice function (d, t) such that $\sum_i E[v_i(d(\overline{\theta}), \overline{\theta}_i) + t_i(\overline{\theta})] \geq 0$. This will generally be satisfied, as otherwise it would be better not to run the mechanism at all. Then if (d, t) does not satisfy ex-ante individual rationality, it is easy to alter transfers (simply adding or subtracting a constant to each individual's transfer function) to reallocate utility so that each individual's expected value from participating in the mechanism is nonnegative. Thus, ex-ante individual rationality is generally vacuously satisfied.

Interim individual rationality is more difficult to satisfy. Let us examine the problem in the context of simple public goods and private goods examples.

Example 5 *Lack of Interim Individual Rationality in a Public Goods Setting*

Let us consider the public project setting described in Example 1 when $n = 2$, $\Theta_i = \{0, 1\}$, and $c = 3/4$. Types are equally likely and independent across individuals. In this case, the efficient decision is to build the project when either $\theta_i = 1$. Split costs among the i's who announce $\theta_i = 1$.[16] So the rule $d(\cdot)$ takes on four values: not build; build and $i = 1$ pays $3/4$; build and $i = 2$ pays $3/4$; and build and both pay $3/8$.

Interim individual rationality implies that

$$\frac{1}{2}(t_1(0,1) + t_1(0,0)) \geq 0 \quad \text{and} \quad \frac{1}{2}(t_2(1,0) + t_2(0,0)) \geq 0 \tag{2}$$

Incentive compatibility evaluated at $\theta_1 = 1$ implies that

$$\frac{1}{2}(1 - \frac{3}{8} + t_1(1,1) + 1 - \frac{3}{4} + t_1(1,0)) \geq \frac{1}{2}(1 + t_1(0,1) + t_1(0,0)).$$

This coupled with (2) implies that

$$t_1(1,1) + t_1(1,0)) > 0. \tag{3}$$

[16] The specification of who pays the cost is actually irrelevant to this example as it comes out in an adjustment of the transfers.

Similarly

$$t_2(1,1) + t_2(0,1)) > 0. \tag{4}$$

Together the inequalities 2, 3, and 4 imply that $\sum_{i,\theta} t_i(\theta) > 0$, which cannot be satisfied if balance or even just feasibility is satisfied. Thus, we have shown that interim individual rationality is not compatible with Bayesian incentive compatibility, efficient decision making and feasibility.

Finding mechanisms that take efficient decisions, are incentive compatible, balanced and interim individually rational, is not simply a problem in public goods settings, but is also a problem in private goods settings as was pointed out by Roger Myerson and Mark Satterthwaite. That point is illustrated here in the context of a simple example.

Example 6 *Lack of Interim Individual Rationality in a Bargaining Setting*

A seller ($i = 1$) has an indivisible object worth θ_1 which takes on values in $\Theta_1 = \{0, 3/4\}$ with equal probability. A buyer has a value for the object of of θ_2 that takes on values in $\Theta_2 = [0,1]$ according to a uniform distribution. A decision is a specification in $D = \{0,1\}$ where 0 indicates that the object stays in the seller's hands, while 1 indicates that the object is traded to the buyer.

An efficient decision is $d(\theta) = 1$ if $\theta_2 > \theta_1$, and $d(\theta) = 0$ if $\theta_2 < \theta_1$. Interim individual rationality requires that if $\theta_2 < 3/4$ (noting that these types only trade 1/2 of the time), then $\frac{1}{2}\theta_2 + E[t_2(\overline{\theta}_1, \theta_2)] \geq 0$, or

$$E[t_2(\overline{\theta}_1, \theta_2)] \geq -\frac{1}{2}\theta_2. \tag{5}$$

Since an efficient decision is the same for any $0 < \theta_2 < 3/4$ and $0 < \theta_2' < 3/4$, Bayesian incentive compatibility implies that $t_2(\theta_1, \theta_2)$ is constant across $0 < \theta_2 < 3/4$. Then (5) implies that for any $0 < \theta_2 < 3/4$

$$E[t_2(\overline{\theta}_1, \theta_2)] \geq 0. \tag{6}$$

Interim individual rationality for sellers of type $\theta_1 = 3/4$ (who expect to trade 1/4 of the time) implies that $(\frac{3}{4})^2 + E[t_1(3/4, \overline{\theta}_2)] \geq \frac{3}{4}$ or $E[t_1(3/4, \overline{\theta}_2)] \geq 3/16$. Then incentive compatibility for type $\theta_1 = 0$ implies that $E[t_1(0, \overline{\theta}_2)] \geq 3/16$. Thus, $E[t_1(\overline{\theta})] \geq 3/16$. Feasibility then implies that $-3/16 \geq E[t_2(\overline{\theta})]$. Then by (6) it follows that

$$-\frac{3}{4} \geq E[t_2(\overline{\theta}_1, \theta_2)]$$

for some $\theta_2 \geq 3/4$. However, this, $1 \geq \theta_2 > 3/4$, and (6) then imply that

$$\frac{\theta_2}{2} + E[t_2(\overline{\theta}_1, 0)] \geq \theta_2 + E[t_2(\overline{\theta}_1, \theta_2)],$$

which violates incentive compatibility.

Thus, there does not exist a mechanism that satisfies interim individual rationality, feasibility, efficiency, and Bayesian incentive compatibility in this setting.

The Myerson-Satterthwaite theorem shows that there does not exist a mechanism that satisfies interim individual rationality, balance, efficiency, and Bayesian incentive compatibility for a general set of distributions. That theorem extracts the full implications of the incentive compatibility constraints and develops expressions for the interim expected utility of the buyer and seller in a similar setting as this example while allowing for general forms of uncertainty (with independent types). Those expressions show the impossibility of satisfying interim individual rationality, balance, efficiency and Bayesian incentive compatibility for a wide range of distributions over types.

To see the importance of the interim perspective on individual rationality in the above example[17], consider satisfying only an ex-ante individual rationality constraint in the context of Example 6. Consider the following mechanism: $d(\theta) = 1$ if $\theta_1 < \theta_2$ and $d(\theta) = 0$ if $\theta_1 \geq \theta_2$; $t_1(0, \theta_2) = -t_2(0, \theta_2) = 3/16$ and $t_1(3/4, \theta_2) = -t_2(3/4, \theta_2) = \frac{3}{4} d(3/4, \theta_2)$. It is easy to check that this mechanism is incentive compatible, balanced and efficient, and that it is ex-ante individually rational. It is not interim individually rational as some θ_2's that are less than 3/16 end up making expected payments (3/32) higher than the value they get from the object that they get 1/2 of the time.

Which time perspective is appropriate depends on the application and in particular on the time at which a mechanisms prescriptions become binding. If contracting occurs at an ex-ante time (as in many long term contracts and forward contracts), then the ex-ante perspective would be appropriate and there will not be problems in finding a mechanism that individuals will participate in that will provide efficient decisions and balanced transfers. If contracting occurs at an interim time and individuals cannot be forced to participate, then there will exist problems in a wide variety of settings.

[17] Note that any impossibility that holds under interim individual rationality implies that the same is true of the stronger ex-post individual rationality constraint.

4.5 Large Societies

The private information that an individual holds has value, as it is needed in order to make efficient decisions. Adding interim individually rationality effectively requires that the individual must be compensated for his or her information, which gives some power to the individual. When each individual exercises that power and earns rents, and the outcome is distorted and inefficiency results.

A reasonable conjecture is that as the size of the society increases, each individual's private information becomes less important and so the distortions should tend to become negligible. This could allow for mechanisms that are interim individually rational to become approximately efficient and balanced in large societies. It turns out that the validity of this reasoning depends on the setting, and the conclusion is different in public and private goods settings. Let us examine two different settings that illustrate this point.

Example 7 *Approximate Efficiency and Balance in Large Double-Auctions*

Consider a bargaining setting similar to that described in Example 6 except with a an even number n of individuals, with $n/2$ buyers and $n/2$ sellers. Recall that θ_i represents i's valuation for the object. The θ_i's of buyers and sellers are all independently distributed over $[0,1]$. A decision is $d \in \{0,1\}^n$ such that $\sum_i d_i = n/2$ where $d_i = 1$ indicates that i ends up allocated an object. The efficient decision is to place the objects in the hands of the $n/2$ individuals who have the highest valuations for the objects (whether they be buyers or sellers).

The following mechanism is due to Preston McAfee. It is feasible and Bayesian incentive compatible (in fact dominant strategy incentive compatible), ex-post individually rational (and thus interim and ex-ante individually rational), and is approximately balanced and efficient in large societies. Rank the announced θ_i's (of all individuals). Let $A(\theta) \subset \{1,\ldots,n\}$ denote the set of the i's who have the $n/2$ highest θ_i's. Break any ties in favor of lowest indexed i's. Let $i_b(\theta) \in A(\theta)$ be the buyer who has the lowest θ_i among buyers in $A(\theta)$. Let $i_s(\theta) \notin A(\theta)$ be the seller who has the highest θ_i among sellers who are not in $A(\theta)$. Set

$$d_i(\theta) = \begin{cases} 1 & \text{if } i \in A(\theta) \cup i_s(\theta) \setminus i_b(\theta) \\ 0 & \text{otherwise.} \end{cases}$$

and

$$t_i(\theta) = \begin{cases} -\theta_{i_b} & \text{if } i \text{ is a buyer, other than } i_b\text{, who is in } A(\theta) \\ \theta_{i_s} & \text{if } i \text{ is a seller, other than } i_s\text{, who is not in } A(\theta) \\ 0 & \text{otherwise.} \end{cases}$$

As this is essentially a two-sided version of a Vickrey auction, it easily checked that truth is a dominant strategy for this mechanism. The loss of efficiency of the mechanism is only $\theta_{i_s} - \theta_{i_b}$ which tends to zero in per-capita terms as n grows, and tends to zero in absolute terms (with probability 1) for some distributions over buyer and seller types. The lack of balance of the mechanism is at most $\frac{n}{2}(\theta_{i_s} - \theta_{i_b})$ which tends to zero in per-capita terms (with probability 1), for a wide variety of distributions on buyers and sellers types.[18]

Thus, in a simple market setting large numbers can help alleviate (at least in an approximate sense) some of the difficulty of satisfying balance and efficiency due to the constraints of incentive compatibility coupled with interim individual rationality.[19] The news is not quite as optimistic in a public goods setting.

Example 8 *Inefficiency in Large Public Goods Settings*

Let us examine a variation of Example 1 in larger societies. Each individual $i \in N$ has a value for the public project $\theta_i \in \{0, 1\}$, and these are distributed with equally probability and independently across individuals. The cost of the project is c. A decision is $d = (y, z_1, \ldots, z_n)$ where $y \in \{0, 1\}$ indicates whether or not the public project is built and $z_i \in [0, c]$ indicates the cost of the good paid by i, where $\sum_i z_i = c$ if $y = 1$ and $\sum_i z_i = 0$ if $y = 0$. A decision rule $d(\theta) = (y(\theta), z(\theta))$ is efficient if the project is built when $\#\{i \mid \theta_i = 1\} > c$, and not when $\#\{i \mid \theta_i = 1\} < c$.

For simplicity, consider only mechanisms for which $y(\theta)$ depends only on $\#\{i \mid \theta_i = 1\}$ and is non-decreasing in this number. Of course, this is true of the efficient decision. In addition, consider mechanisms that treat individuals symmetrically so that z and t are symmetric functions.[20] In this context, consider a mechanism that satisfies feasibility, interim individual rationality, and Bayesian incentive compatibility. Interim individual rationality implies that for any i

$$E[-z_i(\overline{\theta}) + t_i(\overline{\theta}) \mid \overline{\theta}_i = 0] \geq 0.$$

[18] Work by Mark Satterthwaite and Steve Williams, and by Aldo Rustichini, Mark Satterthwaite, and Steve Williams shows that a class of mechanisms called k-double auctions converge to efficiency at an even faster rate, and satisfy balance exactly, but are only Bayesian incentive compatible.

[19] The context does make a difference, and the simplicity of indivisible goods with single unit demands is important here. These results do not extend to more general settings where richer preferences over the goods being exchanged are admissible, without some additional conditions.

[20] If θ' is found from permuting θ_i and θ_j in θ, then $t_i(\theta') = t_j(\theta)$. The difficulties here hold in a variety of contexts and in a wider class of mechanisms as has been shown by a number of authors.

Symmetry and feasibility then imply that for any i

$$0 \geq E[t_i(\overline{\theta}) \mid \overline{\theta}_i = 1].$$

Bayesian incentive compatibility evaluated at $\theta_i = 1$ implies that

$$E[y(\overline{\theta}) - z_i(\overline{\theta}) + t_i(\overline{\theta}) \mid \overline{\theta}_i = 1] \geq E[y(\overline{\theta}_{-i}, 0) - z_i(\overline{\theta}_{-i}, 0) + t_i(\overline{\theta}_{-i}, 0) \ \mid \ \overline{\theta}_i = 1].$$

Given independence and the signs of t_i established above, this implies that

$$E[y(\overline{\theta}_{-i}, 1) - y(\overline{\theta}_{-i}, 0)] \geq E[z_i(\overline{\theta}_{-i}, 1)] \tag{7}$$

Note that under the symmetry requirement, y can be written as a function of k where k is the number of individuals who have $\theta_i = 1$. Then the nondecreasing nature of y implies that $y(k) - y(k-1) = 1$ for at most one value of k. So, letting s be the number of individuals besides i who have $\theta_j = 1$, inequality (7) implies that

$$\max_s \left[\left(\frac{1}{2}\right)^{n-1} \frac{(n-1)!}{(n-1-s)!s!} \right] \geq E[z_i(\overline{\theta}_{-i}, 1)]$$

Using Stirling's formula one can bound the left-hand expression, and find that

$$\frac{2e^{\frac{1}{12n}}}{\sqrt{2\pi n}} \geq E[z_i(\overline{\theta}_{-i}, 1)]$$

Let p denote the (unconditional) probability that the project is undertaken under the decision rule. A lower bound on $E[z_i(\overline{\theta}_{-i}, 1)]$ is $p\frac{c}{n}$. This implies that

$$\frac{2e^{\frac{1}{12n}}}{\sqrt{2\pi n}} \geq p\frac{c}{n}$$

This can be satisfied if (and only if) pc grows with n at a rate slower than $\sqrt{n}$. So, for large numbers to help in getting approximate efficiency and balance under incentive compatibility and interim individual rationality, it has to be that the per capita cost of the project becomes negligible in the limit; or if not then the probability of building the public project under the efficient decision must go to 0. In that second case, approximate efficiency is obtained simply by never building the project, and so in cases of interest the per capita cost of the project must be going to 0.

An important remark is that in the above analysis, an individual still benefits from the public project even if they say $\theta_i = 0$. This is the nature of the "free-rider" problem. If the public good is excludable, so that individuals who say $\theta_i = 0$ can be excluded from using the public facility (as in a toll road), then there exists an ex-post individually rational, dominant strategy incentive compatible, efficient and balanced mechanism. Use the efficient decision rule, exclude individuals who have $\theta_i = 0$, and split costs equally among those with $\theta_i = 1$ when the project is undertaken.[21]

4.6 Correlated Types

Having correlation between the types of the individuals can help in constructing mechanisms. This was originally pointed out in work by Jacques Crémer and Richard McLean and has been extended in a number of directions. A rough intuition for this is that correlation helps in extracting the private information of an individual. Knowing θ_{-i} alters the conditional distribution of θ_i, and adjusting transfers as a function of θ so that i is taxed in instances that are less likely conditional on θ_{-i} and rewarded in instance that are more likely, can induce i to correctly announced θ_i. To get a feeling for this, let us reconsider Example 8 with correlation among types.

Example 9 *Interim Individual Rationality with Correlated Types*

Consider the setting of Example 8 when $n = 3$. Let $c = 7/4$ and consider the variation on the uncertainty, where $\theta = (0, 0, 0)$ and $\theta = (1, 1, 1)$ occur with probability $\frac{1}{8} + 3\varepsilon$, and the other realizations of θ each occur with probability $\frac{1}{8} - \varepsilon$, where $0 < \varepsilon < \frac{1}{8}$. So, ε is a parameter that measures the strength of correlation and adjusts away from the fully independent case. Given the cost, it is efficient to build the project when at least two individuals have $\theta_i = 1$ and not otherwise.

Design a mechanism as follows. Use the efficient decision rule and split costs equally among those with $\theta_i = 1$ when the project is undertaken. Set t as follows.

$$t_i(\theta) = \begin{cases} \text{x} & \text{if } \theta_i = \theta_{i+1} \neq \theta_{i+2} \\ \text{-x} & \text{if } \theta_i \neq \theta_{i+1} = \theta_{i+2} \\ 0 & \text{otherwise,} \end{cases}$$

where $i + 1$ and $i + 2$ are taken modulo 3, and requirements on x are described below. It is easily seen that t is balanced, and that the mechanism is efficient and interim individually rational. Let us examine the conditions relating x to ε that result from requiring Bayesian incentive compatibility. It is easily checked that a type $\theta_i = 0$ does

[21] While this works in the simple two type case above, and excludability can help more generally, the extent to which approximate efficiency and balance are achievable depends on the range of valuations for the public good that are possible.

not want to announced $\theta_i = 1$, as that would increase the expected cost paid and decrease the expected transfer. We need only check incentives that type $\theta_i = 1$ not desire to announce $\theta_i = 0$, which requires that

$$(\frac{1}{4} + 6\varepsilon)(1 - \frac{c}{3}) + (\frac{1}{4} - 2\varepsilon)(1 - \frac{c}{2} + x) + (\frac{1}{4} - 2\varepsilon)(1 - \frac{c}{2}) + (\frac{1}{4} - 2\varepsilon)(-x)$$

$$\geq (\frac{1}{4} + 6\varepsilon)(1 - x) + (\frac{1}{4} - 2\varepsilon)(0) + (\frac{1}{4} - 2\varepsilon)(x) + (\frac{1}{4} - 2\varepsilon)(0).$$

This reduces to

$$x \geq \frac{1}{8\varepsilon}\left(\frac{c}{3} - \frac{1}{2} + 6\varepsilon\right).$$

This is satisfied for large enough x.

Note that $x \to \infty$ as $\varepsilon \to 0$, which points out one weakness of this approach to exploiting correlation. To take advantage of small amounts of correlation, the size of the transfers has to grow arbitrarily large. If there is some bound on these transfers, or a bankruptcy constraint, then for small amounts of correlation these conditions cannot all be satisfied.

4.7 Interdependent Valuations

The analysis discussed to this point has assumed that the type of an individual i only affects that individual's preferences. That is an assumption known as private values. There are many instances where the information that one individual holds may concern the valuation of other individuals as well. For instance, if a number of bidders are bidding in an auction for the rights to drill for oil in a given tract of land, one bidder may have some information about how much oil is contained on some part (or all) of the tract. That information concerns the value of the land to any of the bidders in the auction.

In settings with interdependent valuations, difficulties arise in getting efficient decisions together with incentive compatibility, regardless of issues of balance and individual rationality. The following example, due to Eric Maskin, shows the difficulties that can arise.

Example 10 *Inefficiency with Interdependent Valuations.*

A single indivisible object is to be allocated to one of two individuals. Only individual $i = 1$ observes information described by $\theta_1 \in [0, 2]$. Individual 1's value for the object is $2\theta_1 + 1$, while individual 2's value for the object is $3\theta_1$. The object should be given to individual 1 if $\theta_1 < 1$ and individual 2 if $\theta_1 > 1$. In this case there is no uncertainty for individual 1, and so dominant strategy and Bayesian incentive compatibility constraints coincide. Applying those incentive constraints to $\theta_1 > 1 > \theta_1'$, truth being a best response for type θ_1 imples that

$$t_1(\theta_1) \geq 2\theta_1 + 1 + t_1(\theta_1')$$

and truth being a best response for type θ_1' imples that

$$2\theta_1' + 1 + t_1(\theta_1') \geq t_1(\theta_1).$$

These imply that $2\theta_1' \geq 2\theta_1$, which is not possible given that $\theta_1 > 1 > \theta_1'$.

There is an aspect to the above example that is important. Changes in the information of individual 1 affect individual 2's valuation more than individual 1's valuation. Partha Dasgupta and Eric Maskin have shown that if each individual's type affects all valuations in a positive way, and the derivative of i's valuation with respect to i's type is at least as large as the derivative of j's valuation with respect to i's type for all i and j, then there is a variation of the Vickrey auction which results in the efficient decision and is Bayesian incentive compatible. While this is true with single dimensional types, recent work by Philippe Jéhiel and Benny Moldovanu has shown that if types take one more than one dimension then efficiency and incentive compatibility are again at odds without some strong additional assumptions.

There is still much that is not known about the existence or properties of incentive compatible mechanisms that are efficient (much less the balanced and individual rational), when there are general forms of uncertainty and interdependencies in the preferences of individuals. This is an active area of research.

5 Implementation

There is a strong caution to be added to the approach to mechanism design that uses the revelation principle as a tool. It is possible for a mechanism to have more than one Bayesian equilibrium, [22] and in fact uniqueness of equilibrium might be thought of as the exception rather than the rule. The revelation principle just relates one equilibrium

[22]The same can be said in the dominant strategy case, but dominant strategies (when they exist) are in many applications either unique or such that the set of dominant strategies results in equivalent outcomes.

of the mechanism to truth under the corresponding direct revelation mechanism. There could be other equilibria of the mechanism that are unaccounted for in the analysis and could be important. Also, any direct revelation mechanism may have untruthful equilibria that do not correspond to any equilibrium of the original mechanism.

As it is often important to keep track of all the equilibria of a mechanism, there is a loss of generality when restricting attention to direct mechanisms. This issue is the main point of departure between what is known as the *mechanism design* literature which focuses on individual equilibria and uses direct mechanisms as a common tool, and the *implementation* literature which keeps track of all equilibria and works with the space of indirect mechanisms. Finding mechanisms where all equilibria have desirable properties adds additional constraints that can be quite limiting. That problem is not discussed here, and the reader is referred to the bibliography surveys of the implementation literature.

Acknowledgments: Some of the contents of this paper are based on lectures Jackson gave at the CORE-Francqui Summer School in May and June of 2000. He thanks the organizers Claude d'Aspremont, Michel Le Breton and Heracles Polemarchakis, as well as CORE and the Francqui Foundation, and the participants of the summer school for their support and feedback. He also thanks Ulrike Ervig and Benny Moldovanu for comments on an earlier draft.

6 Bibliography

Corchon, L. (1996), *The Theory of Implementation of Socially Optimal Decisions in Economis*, McMillan: London, [A text that covers of dominant strategy mechanism design and implementation theory.].

Fudenberg, D. and J. Tirole (1993), *Game Theory*, MIT Press: Cambridge MA, [A text with a chapter devoted to Bayesian mechanism design.].

Jackson, M.O. (1997), "A Crash Course in Implementation Theory ," Forthcoming in *The Axiomatic Method: Principles and Applications to Game Theory and Resource Allocation*, edited by William Thomson, [An overview of implementation theory with a comprehensive and up-to-date bibliography.].

Mas-Colell, A., M. Whinston, and J. Green (1995), *Microeconomic Theory*, Oxford University Press: Oxford, [A text with a chapter devoted to mechanism design.].

Moore, J. (1992), "Implementation in Environments with Complete Information." In J.J. Laffont, *Advances in Economic Theory: Proceedings of the Congress of the Econometric Society*, Cambridge University Press, [A survey of implementation theory including dominant strategy mechanism design.].

Moulin, H. (1991), *Axioms of Cooperative Decision Making*, Cambridge University Press, Cambridge, [A text with chapters including aspects of dominant strategy mechanism design as well as results on public goods mechanisms.].

Palfrey, T., and S. Srivastava (1993), *Bayesian Implementation*, Harwood Academic Publishers: Switzerland, [A monograph that surveys Bayesian implementation.].

Addendum to the Working Paper Version

Bibliographic Notes:

This addendum to the working paper version of this manuscript contains some bibliographic notes as well as a more extensive bibliography.

The Gibbard-Satterthwaite Theorem reported in Section 3.3 is due in Gibbard (1973) and Satterthwaite (1975) and can be found in many sources. It extends to infinite decision spaces and a proof with continuous preferences can be found in Barberà and Peleg (1990).

The phantom voting methods described in Section 3.4 are due to Moulin (1980). A proof that such methods are strategy-proof in the direct mechanism (when individuals can report full preferences) appears in Barberà and Jackson (1994). The fixed price and proportion trading methods are characterized by Barberà and Jackson (1995). There are other domains where interesting strategy-proof rules exist. Another leading example is the allotment of a divisible good or task under single-peaked preferences, as first studied in Sprumont (1991) and further in Ching (1992, 1994) and Barberà, Jackson and Neme (1997).

The Groves mechanisms described in section 3.5 are due to Groves (1973) and the converse of the characterization theorem is from Green and Laffont (1977). The mechanisms are discussed in some detail in Green and Laffont (1979).

The pivotal mechanism in section 3.6 was first described by Clarke (1971) and Groves (1973), while the Vickrey auctions were first analyzed by Vickrey (1961). The mentioned Moulin characterization of the pivotal mechanism as providing the highest minimum utility level appears in Related auctions for multiple goods and interdependent valuations have recently been introduced and explored by Ausubel (1997), Dasgupta and Maskin (1997), and Perry and Reny (1999).

Difficulties with balance of Groves schemes have been explored in Green and Laffont (1979) and Rob (1982), among others. These authors also explore the issues of large numbers and approximate efficiency. Laffont and Maskin (1980) explore domains where there exist balanced Groves' schemes and Groves and Loeb (1975) exhibit a class of public goods problems where balance is achievable via an anonymous mechanism. The issue of reconciling efficiency with dominant strategies in economic settings going beyond quasi-linear ones appear in Hurwicz (1972), Hurwicz and Walker (1990), Zhou (1991), Moreno (1994), Barberà and Jackson (1995), Cordoba and Hammond (1997), Kovalenkov (1996), Schummer (1998), Nicolo (1999), and Serizawa (1999b). The analysis of strategy-proof mechanisms in a variety of settings has been an area of extensive recent research and many references are provided below.

The dominant strategy incentive compatible and balanced, but inefficient rules described in section 3.9 were first analyzed in Moulin (1994) with some additional axioms and the theorem quoted here is a special case of a theorem by Serizawa (1999). The full class of strategy-proof rules mentioned in footnote 11 appear in Roberts (1979).

The balanced, efficient, and Bayesian incentive compatible mechanisms described in Theorem 4 are due to d'Aspremont and Gerard-Varet (1979) and Arrow (1979), while Theorem 5 is due to d'Aspremont, Crémer and Gerard-Varet (1990). The Myerson-Satterthwaite theorem described after example 6 appears in Myerson and Satterthwaite (1983).

The question of whether large societies can help reconcile problems of finding approximately efficient, balanced, incentive compatible, still has many open facets. The mechanism described in Example 7 is due to McAfee (1993). The convergence discussed in footnote 18 has been explored by Gresik and Satterthwaite (1989), Satterthwaite and Williams (1989), and Rustichini, Satterthwaite and Williams (1995) (see also Satterthwaite (1999)). Example 8 illustrates ideas that are explored in Mailath and Postlewaite (1990) and Al-Najjar and Smorodinsky (2000ab). The possibility of finding such approximately efficient mechanisms with large numbers in settings with richer preference structures has been explored in Roberts and Postlewaite (1979), Gul and Postlewaite (1992), Jackson (1992), Jackson and Manelli (1997), and McLean and Postlewaite (1999), and answers depend on the information structure and details of modeling such as infinite versus finite type spaces in subtle ways.

The power of taking advantage of correlation in individual types was explored in Crémer and McLean (1985, 1988) (see also an example in Myerson (1981)) in specific contexts and has been studied more generally by many authors including d'Aspremont, Crémer and Gerard-Varet (1990), McAfee and Reny (1992), Shinotsuka and Wilkie (1999), and Spiegel and Wilkie (2000).

Example 10 on interdependent valuations is due to Maskin (1992). Recent studies of the efficient allocation of private goods with interdependencies in valuations include Dasgupta and Maskin (1997), Jehiel and Moldovanu (1998), Perry and Reny (1998), Pesendorfer and Swinkels (1998), Jackson (1998), Fieseler, Kittsteiner, and Moldovanu (2000).

The implementation issues mentioned in section 5 are analyzed in an extensive literature. An introduction to that literature can be found in Jackson (1997) focusing on the complete information setting, and Palfrey and Srivastava (1993) provide an overview of the Bayesian implementation literature. A few references to the papers from the implementation literature that are most closely tied to this survey are provided in the bibliography below.

There are many specific problems that have been analyzed in detail that can the thought of as mechanism design literature. This survey has touched on parts of the auctions and public goods literatures, but does not attempt to provide any survey of those vast literatures. In addition, there are large literatures on price discrimination, principal-agent problems, contract theory, adverse selection, and market design where a mechanism design approach is central. A few representative references are sprinkled through the bibliography.

Bibliography

Abdulkadiroglu, A. and T. Sönmez (1998), "Random Serial Dictatorship and the Core from Random Endowments," *Econometrica,* Vol. 66, pp. 689–701.

Abdulkadiroglu, A. and T. Sönmez (1999), "House Allocation with Existing Tenants," forthcoming in *Journal of Economic Theory*.

Al-Najjar, N. and R. Smorodinsky (2000), "Provision of a Public Good with Bounded Cost," *Economics Letters,* Vol. , pp. .

Al-Najjar, N. and R. Smorodinsky (2000), "Pivotal Players and the Characterization of Influence," *Journal of Economic Theory,* Vol. , pp. .

Armstrong, M. (1996), "Multiproduct Nonlinear Pricing," *Econometrica,* Vol. 64, pp. 51–75.

Arrow, K. (1979), "The Property Rights Doctrine and Demand Revelation under Incomplete Information." In M. Boskin, *Economies and Human Welfare*, Academic Press: NY.

Ausubel, L. (1997), "An Efficient Ascending-Bid Auction for Multiple Objects," mimeo: University of Maryland.

Bag, P.K. (1996), "Efficient Allocation of a "Pie": King Solomon's Dilemma," *Games and Economic Behavior,* Vol. 12, pp. 21–41.

Bagnoli, M. and B. Lipman (1989), "Provision of Public Goods: Fully Implementing the Core through Private Contributions," *Review of Economic Studies,* Vol. 56, pp. 583–601.

Bali, V. and M.O. Jackson (2000), "Asymptotic Revenue Equivalence in Auctions," mimeo: Caltech.

Barberà, S. (2000), "An Introduction to Strategy-Proof Social Choice Functions," mimeo: Universitat Autonoma de Barcelona.

Barberà, S. and M.O. Jackson (1994), "A Characterization of Strategy–proof Social Choice Functions for Economies with Pure Public Goods," *Social Choice and Welfare,* Vol. 11, pp. 241–252.

Barbera, S. and M.O. Jackson (1995), "Strategy–proof Exchange," *Econometrica,* Vol. 63, pp. 51–87.

Barbera, S., M.O. Jackson and A. Neme (1997), "Strategy–Proof Allotment Rules," *Games and Economic Behavior,* Vol. 18, pp. 1–21.

Barbera, S., F. Gul, and E. Stacchetti (1993), "Generalized Median Voter Schemes and Committees," *Journal of Economic Theory,* Vol. 61, pp. 262–289.

Barbera, S., J. Masso, and A. Neme (1997), "Voting Under Constraints," *Journal of Economic Theory,* Vol. 76, pp. 298–321.

Barbera, S., H. Sonnenschein, and L. Zhou (1991), "Voting by Committees," *Econometrica,* Vol. 59, pp. 595–609.

Barberà, S. and B. Peleg (1990), "Strategy–Proof Voting Schemes with Continuous Preferences," *Social Choice and Welfare,* Vol. 7, pp. 31–38.

Berga, D. (1997), "Maximal Domains and Strategy-Proofness in Public Good Economies," thesis: Universitat Autonoma de Barcelona.

Border, K. and J.S. Jordan (1983), "Straightforward Elections, Unanimity and Phantom Voters," *Review of Economic Studies,* Vol. 50, pp. 153–170.

Brusco, S. (1998), "Unique Implementation of the Full Surplus Extraction Outcome in Auctions with Correlated Types," *Journal of Economic Theory,* Vol. 80, pp. 185–200.

Brusco, S. and M. Jackson (1999), "The Optimal Design of a Market," *Journal of Economic Theory,* Vol. 88, pp. 1–39.

Clarke, E.H. (1971), "Multi-part Pricing of Public Goods," *Public Choice,* Vol. 11, pp. 17–33.

Ching, S. (1992), "A Simple Characterization of the Uniform Rule," *Economics Letters,* Vol. 40, pp. 57–60.

Ching, S. (1994), "An Alternative Characterization of the Uniform Rule," *Social Choice and Welfare,* Vol. 11, pp. 131–136.

Corchon, L. (1996), *The Theory of Implementation of Socially Optimal Decisions in Economics,* McMillan: London.

Cordoba, J. and P. Hammond (1997), "Asymptotically Strategy-proof Walrasian Exchange," *forthcoming: Mathematical Social Sciences,* Vol. , pp. .

Cramton, P., R. Gibbons and P. Klemperer (1987), "Dissolving a Partnership Efficiently," *Econometrica,* Vol. 55, pp. 615–632.

Crémer, J. and R. McLean (1985), "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica,* Vol. 53, pp. 345–361.

Crémer, J. and R. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica,* Vol. 56, pp. 1247–1257.

Dasgupta, P. and E. Maskin (1997), "Notes on Efficient Auctions," mimeo Harvard University.

d'Aspremont, C. and L.-A. Gerard-Varet (1979), "Incentives and Incomplete Information," *Journal of Public Economics,* Vol. 11, pp. 25–45.

d'Aspremont, C., J. Crémer, and L.-A. Gerard-Varet (1990), "Incentives and the Existence of Pareto-Optimal Revelation Mechanisms," *Journal of Economic Theory,* Vol. 51, pp. 233–254.

Fieseler, K., T. Kittsteiner, and B. Moldovanu (2000), "Partnerships, Lemons, and Efficient Trade," mimeo: University of Mannheim.

Fudenberg, D. and J. Tirole (1993), *Game Theory,* MIT Press: Cambridge MA, .

Gibbard, A. (1973), "Manipulation of Voting Schemes: A General Result," *Econometrica,* Vol. 41, pp. 587–601.

Glazer, J. and A. Ma (1989), "Efficient Allocation of a "Prize": King Solomon's Dilemma," *Games and Economic Behavior,* Vol. 1, pp. 222–233.

Green, J. and J.J. Laffont (1977), "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica,* Vol. 45, pp. 727–738.

Green, J. and J.J. Laffont (1979), *Incentives in Public Decision Making,* North Holland: Amsterdam.

Gresik, T. and M. Satterthwaite (1989), "The Rate at Which a Simple Market Converges to Efficiency as the Number of Traders Increases," *Journal of Economic Theory,* Vol. 48, pp. 304–332.

Groves, T. (1973), "Incentives in Teams," *Econometrica,* Vol. 41, pp. 617–663.

Groves, T. (1977), "Efficient Collective Choice with Compensation." In J.-J. Laffont, *Aggregation and Revelation of Preferences*, North-Holland, Amsterdam.

Groves, T. and J. Ledyard (1977), "Optimal Allocation of Public Goods: A Solution to the Free Rider Problem," *Econometrica,* Vol. 45, pp. 783–809.

Groves, T. and J. Ledyard (1987), "Incentive Compatibility since 1972." In T. Groves, R. Radner, and S. Reiter, *Information, Incentives and Economic Mechanisms*, University of Minnesota Press.

Groves, T. and M. Loeb (1975), "Incentives and Public Inputs," *Journal of Public Economics,* Vol. 4, pp. 211–226.

Gul, F. and A. Postlewaite (1992), "Asymptotic Efficiency in Large Exchange Economies with Asymmetric Information," *Econometrica,* Vol. 60, pp. 1273–1292.

Hagerty, K. and W. Rogerson (1987), "Robust Trading Mechanisms," *Journal of Economic Theory,* Vol. 42, pp. 94–107.

Harsanyi, J.C. (1967–68), "Games with Incomplete Information Played by 'Bayesian' Players," *Management Science,* Vol. 14, pp. 159–189, 320–334, 486–502.

Holmström, B. (1979), "Groves Scheme on Restricted Domain," *Econometrica,* Vol. 47, pp. 1137–1147.

Holmström, B. and R. Myerson (1983), "Efficient and Durable Decision Rules with Incomplete Information," *Econometrica,* Vol. 51, pp. 1799–1819.

Hurwicz, L. (1972), "On Informationally Decentralized Systems." In C.B. McGuire and R. Radner, *Decision and Organization*, North Holland, Amsterdam.

Hurwicz, L. and M. Walker (1990), "On the Generic Nonoptimality of Dominant Strategy Allocation Mechanisms: A General Theorem that Includes Pure Exchange Economies," *Econometrica,* Vol. 58, pp. 683–704.

Jackson, M.O. (1991), "Bayesian Implementation," *Econometrica,* Vol. 59, pp. 461–478.

Jackson M.O. (1992), "Incentive Compatibility and Competitive Allocations," *Economics Letters,* Vol. 40, pp. 299–302.

Jackson, M.O. (1997), "A Crash Course in Implementation Theory ," Forthcoming in *The Axiomatic Method: Principles and Applications to Game Theory and Resource Allocation*, edited by William Thomson, .

Jackson, M.O. (1998), "Efficiency and Information Aggregation in Auctions with Costly Information," mimeo: Caltech.

Jackson, M.O. and A. Manelli (1997), "Approximately Competitive Equilibria in Large Finite Economies," *Journal of Economic Theory,* Vol. 76, pp. .

Jackson, M.O. and H. Moulin (1992), "Implementing a Public Project and Distributing its Cost," *Journal of Economic Theory,* Vol. 57, pp. 125–140.

Jehiel, P. and B. Moldovanu (1999), "Efficient Design with Interdependent Valuations," forthcoming: *Econometrica*.

Jehiel, P. B. Moldovanu, and E. Stacchetti (1999), "Multidimensional Mechanism Design for Auctions with Externalities," *Journal of Economic Theory,* Vol. 85, pp. 258–293.

Klaus, B., H. Peters, and T. Storcken (1995), "Strategy–Proof Division with Single–Peaked Preferences and Initial Endowments," mimeo: Department of Quantitative Economics, University of Limburg.

Kovalenkov, A. (1996), "On a Folk Strategy-Proof Approximately Walrasian Mechanism," forthcoming:*Journal of Economic Theory*.

Krishna, V. and M. Perry (1998), "Efficient Mechanism Design," mimeo: Hebrew University in Jerusalem.

Laffont, J.J. and E. Maskin (1980), "A Differential Approach to Dominant Strategy Mechanism Design," *Econometrica,* Vol. 48, pp. 1507–1520.

Laffont, J.J. and D. Martimort (1997), "Mechanism Design with Collusion and Correlation," mimeo: IDEI.

Ledyard, J. (1979), "Incentive Compatibility and Incomplete Information." In J.-J. Laffont, *Aggregation and Revelation of Preferences*, Amsterdam: North Holland.

Ledyard, J. and T. Palfrey (1994), "Voting and Lottery Drafts as Efficient Public Goods Mechanisms," *Review of Economic Studies,* Vol. 61, pp. 327–355.

Ledyard, J. and T. Palfrey (2000), "A Characterization of Interim Efficiency with Public Goods," *Econometrica,* Vol. 67, pp. 435–448.

Mailath, G. and A. Postlewaite (1990), "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies,* Vol. 57, pp. 351–367.

Makowski, L. and C. Mezzetti (1993), "The Possibility of Efficient Mechanisms for Trading an Indivisible Object," *Journal of Economic Theory,* Vol. 59, pp. 451–465.

Makowski, L. and C. Mezzetti (1994), "Bayesian and Weakly Robust First Best Mechanisms: Characterizations," *Journal of Economic Theory,* Vol. 64, pp. 500– 519.

Makowski, L. and J. Ostroy (1987), "Vickrey-Clarke-Groves Mechanisms and Perfect Competition," *Journal of Economic Theory,* Vol. 42, pp. 244–261.

Mas-Colell, A., M. Whinston, and J. Green (1995), *Microeconomic Theory,* Oxford University Press: Oxford.

Maskin, E. (1992), "Auctions and Privatization." In H. Siebert, *Privatization*, Institut für Weltwirtschaft an der Universität Kiel .

McAfee, R.P. (1992), "A Dominant Strategy Double Auction," *Journal of Economic Theory,* Vol. 56, pp. 434–450.

McAfee, R.P. (1993), "Mechanism Design by Competing Sellers," *Econometrica,* Vol. 61, pp. 1281–1312.

McAfee, R.P. and J. McMillan (1988), "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory,* Vol. 46, pp. 335–354.

McAfee, R.P. and P. Reny (1992), "Correlated Information and Mechanism Design," *Econometrica,* Vol. 60, pp. 395–422.

McLean, R. and A. Postlewaite (1999), "Informational Size and Incentive Compatibility," mimeo: University of Pennsylvania.

Milgrom, P. and R Weber (1982), "A Theory of Auctions and Competitive Bidding," *Econometrica,* Vol. 50, pp. 1089–1122.

Mirrlees, J. (1971), "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies,* Vol. 38, pp. 175–208.

Miyagawa, E. (1999), "Mechanisms for Multilateral Trading and Fixed Prices," mimeo: Columbia University.

Moreno (1994), "Non-Manipulable Decision Mechanisms for Economic Environments," Social Choice and Welfare.11225–240

Mookherjee, D. and S. Reichelstein (1990), "Implementation via Augmented Revelation Mechanisms," *Review of Economic Studies,* Vol. 57, pp. 453–476.

Mookherjee, D. and S. Reichelstein (1992), "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules," *Journal of Economic Theory,* Vol. 56, pp. 378–399.

Moore, J. (1992), "Implementation in Environments with Complete Information." In J.J. Laffont, *Advances in Economic Theory: Proceedings of the Congress of the Econometric Society*, Cambridge University Press: Cambridge.

Moulin, H. (1980), "On Strategy-Proofness and Single-Peakedness," *Public Choice,* Vol. 35, pp. 437–455.

Moulin, H. (1986), "Characterizations of the Pivotal Mechanism," *Journal of Public Economics,* Vol. 31, pp. 53–78.

Moulin, H. (1991), *Axioms of Cooperative Decision Making,* Cambridge University Press, Cambridge.

Moulin, H. (1994), "Serial Cost Sharing of Excludable Public Goods," *Review of Economic Studies,* Vol. 61, pp. 305–325.

Moulin, H. and S. Shenker (1992), "Serial Cost Sharing," *Econometrica,* Vol. 64, pp. 178–201.

Moulin, H. (1999), "Incremental Cost Sharing: Characterization by Coalition Strategy-Proofness," *Social Choice and Welfare,* Vol. 16, pp. 279–320.

Myerson, R. (1979), "Incentive Compatibility and the Bargaining Problem," *Econometrica,* Vol. 47, pp. 61–74.

Myerson, R. (1981), "Optimal Auction Design," *Mathematics of Operations Research,* Vol. 6, pp. 58–73.

Myerson, R. and M. Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory,* Vol. 29, pp. 265–281.

Nicolo, A. (1999), "Efficient and Strategy-Proof Exchange with Leontief Preferences," mimeo: Universitat Autonoma de Barcelona.

Norman, P. (1999), "Efficient Mechanisms for Public Goods with Use Exclusions," mimeo: University of Wisconsin.

Ohseto, S. (1999), "Strategy-Proof Allocation Mechanisms for Economies with an Indivisible Good," *Social Choice and Welfare,* Vol. 16, pp. 121–136.

Palfrey, T., and S. Srivastava (1993), *Bayesian Implementation,* Harwood Academic Publishers: Switzerland.

Palfrey, T., and S. Srivastava (1989a), "Implementation with Incomplete Information in Exchange Economies," *Econometrica,* Vol. 57, pp. 115–134.

Papai, S. (2000), "Strategy-Proof Multiple-Assignment Using Quotas," *Review of Economic Design,* Vol. , pp. .

Papai, S. (1999), "Strategy-Proof Assignment by Hierarchical Exchange," forthcoming.

Perry, M. and P. Reny (1999), "A General Solution to King Solomon's Dilemma," *Games and Economic Behavior,* Vol. 26, pp. 279–285.

Perry, M. and P. Reny (1999), "An Ex-Post Efficient Auction," mimeo: University of Chicago.

Pesendorfer, W. and J. Swinkels (1995), "Efficiency and Information Aggregation in Auctions," forthcoming: American Economic Review.

Pesendorfer, W. and J. Swinkels (1997), "The Loser's Curse and Information Aggregation in Common Value Auctions," *Econometrica,* Vol. 65, pp. 1247–1282.

Postlewaite, A. and D. Schmeidler (1986), "Implementation in Differential Information Economies," *Journal of Economic Theory,* Vol. 39, pp. 14–33.

Riley, J. and W. Samuelson (1981), "Optimal Auctions," *American Economic Review,* Vol. 71, pp. 381–392.

Rob, R. (1982), "Asymptotic Efficiency of the Demand-Revealing Mechanism," *Journal of Economic Theory,* Vol. 28, pp. 208–220.

Roberts, K. (1979), "The Characterization of Implementable Choice Rules." In J.-J. Laffont, *Aggregation and Revelation of Preferences*, Amsterdam: North Holland.

Roberts, D.J. and A. Postlewaite (1976), "The Incentives for Price Taking Behavior in Large Exchange Economies," *Econometrica,* Vol. 44, pp. 115–127.

Rochet, J.C. and P. Choné (1998), "Ironing, Sweeping, and Multi- Dimensional Screening," *Econometrica,* Vol. 66, pp. 783–826.

Roth, A. and M. Sotomayor (1990), *Two Sided Matching: A Study in Game-Theoretic Modeling and Analysis,* Cambridge University Press: Cambridge.

Rustichini, A., M. Satterthwaite, and S. Williams (1995), "Convergence to Efficiency in a Simple Market with Incomplete Information," *Econometrica,* Vol. 62, pp. 1041–1063.

Samuelson, P. (1954), "The Pure Theory of Public Expenditures," *Review of Economics and Statistics,* Vol. 36, pp. 387–389.

Satterthwaite, M. (1975), "Strategy–proofness and Arrow' Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Theorems," *Journal of Economic Theory,* Vol. 10, pp. 187–217.

Satterthwaite, M. (1999), "Strategy–proofness and Markets," CMSEMS Discussion Paper 1255, Northwestern University.

Satterthwaite, M. and S. Williams (1989), "Bilateral Trade with the Sealed Bid k-Double Auction: Existence and Efficiency," *Journal of Economic Theory,* Vol. 48, pp. 107–133.

Schummer, J. (1997), "Strategy-Proofness versus Efficiency on Restricted Domains of Exchange Economies," Social Choice and Welfare.1447–56

Schummer, J. and R. Vohra (1999), "Strategy-Proof Location on a Network," mimeo: Northwestern University.

Serizawa, S. (1999), "Strategy-Proof and Symmetric Social Choice Functions for Public Good Economies," *Econometrica,* Vol. 67, pp. 121–146.

Serizawa, S. (1999b), "Inefficiency of Strategy-Proof Rules for Pure Exchange Economies," mimeo: Tohoku University.

Shinotsuka, T. and S. Wilkie (1999), "Optimal Multi-Object Auctions with Correlated Types," mimeo: Caltech.

Spiegel, Y. and S. Wilkie (2000), "Optimal Multiproduct Nonlinear Pricing with Correlated Consumer Types," mimeo: Caltech.

Sprumont, Y. (1991), "The Division Problem with Single Peaked Preferences: A Characterization of the Uniform Allocation Rules," *Econometrica,* Vol. 59, pp. 509–519.

Swinkels, J. (1999), "Asymptotic Efficiency for Discriminatory Private Value Auctions," *Review of Economic Studies,* Vol. 66, pp. 509– 528.

Swinkels, J. (1998), "Efficiency of Large Private Value Auctions," *Econometrica,* Vol. , pp. .

Tirole, J. (1999), "Incomplete Contracts: Where do We Stand?," *Econometrica,* Vol. 67, pp. 741–782.

Vickrey, W. (1961), "Counterspeculation, Auctions, and Competitive Sealed-Tenders," *Journal of Finance,* Vol. 16, pp. 8–37.

Walker, M. (1980), "On the Non-existence of a Dominant Strategy Mechanism for Making Optimal Public Decisions," *Econometrica,* Vol. 48, pp. 1521–1540.

Weymark, J. (1999), "Sprumont's Characterization of the Uniform Rule when all Single-Peaked Preferences are Admissible," *Review of Economic Design,* Vol. 4, pp. 389–393.

Wilson, R. (1992), "Design of Efficient Trading Procedures." In D. Friedman, J. Geankoplos, D. Lane and J. Rust, *The Double Auction Market*, SFI Studies in the Sciences of Complexity.

Zheng, C.Z. (1999), "A Recipe for Optimal Multidimensional Auctions," mimeo: Northwestern University.

Zhou, L. (1991), "Inefficiency of Strategy–Proof Mechanisms in Pure Exchange Economies," *Social Choice and Welfare,* Vol. 8, pp. 247–254.