An Introduction to the Theory of Mechanism Design

Tilman Börgers

August 5, 2010

Preliminary Draft

Contents

Preface

This typescript consists of lecture notes that I have used to teach courses on mechanism design for graduate students at the University of Michigan and at the Institute for Advanced Studies in Vienna.

The primary objective of these notes is to give rigorous but accessible explanations of classic results in the theory of mechanism design, such as Myerson's theorem on expected revenue-maximizing auctions, Myerson and Satterthwaite's theorem on the impossibility of ex post efficient bilateral trade with asymmetric information, and Gibbard and Satterthwaite's theorem on the nonexistence of dominant strategy voting mechanisms. A second objective of these notes is to take the reader to the frontiers of research in some selected areas.

I have chosen a somewhat unusual beginning for these notes with Chapter 2, where I explain the basic theory of screening. The theory of screening is sometimes not regarded as part of the theory of mechanism design because it constructs an incentive scheme for only one agent rather than multiple interacting agents. However, the theories that are covered in Chapter 2 are intimately linked to the theories of optimal mechanisms that are explained in later chapters, particularly in Chapter 3. My hope is that by juxtaposing the theory of screening and the theory of mechanism design I can help the reader understand which features of optimal mechanism design are due to strategic interaction, and which features of optimal mechanism design are identical to the corresponding features of optimal screening.

An organizing principle of a large proportion of these notes is that I compare mechanisms that are constructed with the anticipation that agents will play a Bayesian equilibrium in some given information structure with small type spaces, to mechanisms that provide each agent with a dominant strategy. These two approaches have been the prevailing approaches of the classic theory of mechanism design. I treat these two approaches side by side with the intention to alert the reader to the potential problems of both approaches, and to prepare the grounds for a later discussion of robust mechanism design. The modern theory of robust mechanism design has moved the discussion beyond the traditional dichotomy of Bayesian mechanism design and dominant strategy mechanism design. I hope that the motivation for some of the questions asked in modern mechanism design theory becomes clearer once readers thoroughly understand the two traditional approaches.

This typescript assumes that readers have had a prior course in game theory, and that they have basic knowledge of real analysis. Other than this, however, I have sought to keep the prerequisites minimal.

This draft is incomplete in many ways. Chapters 9-12 are still to be written. Earlier chapters need further revisions. New chapters may be inserted in the future. I also plan to include further problems at the end of each chapter. Any suggestions for the further development of these notes are always welcome. Please write to: tborgers@umich.edu.

I am very grateful to Stefan Behringer and Trevor Burnham who read earlier drafts of this manuscript and pointed out my many errors. All remaining sins of commission or omission are my own.

Tilman Börgers

Chapter 1

Introduction

Suppose you want to sell your house, and your realtor has identified several potential buyers who are willing to pay your ask price. You might then wish to conduct an auction among these buyers to obtain a higher price. There are many different auction formats that you could use: For example, each buyer could be asked to send in one binding and final bid. Alternatively, buyers could bid in several rounds, and in each round they are all informed about the highest bid of the previous round, and are then asked to revise their bids. You could also use some combination of these formats. How should you choose among different auction formats? This is one of the questions that the theory of mechanism design aims to answer.

Now imagine that you and your colleagues are considering whether to buy a new refrigerator to be kept at work, in which you can store food that you bring from home. While everyone is in favor, it is not so clear how much money people are willing to contribute. How can you find out whether the sum of the amounts that everyone is willing to contribute covers the cost of the refrigerator? You could ask everyone to submit pledges simultaneously, and then see whether the sum of the pledges covers the expense. Alternatively, you could go around, and tell each colleague how much everyone else has pledged so far. Or you could divide the cost by the number of colleagues involved, and commit to buying the refrigerator only if everyone is willing to pay their share. Which of these procedures is best? Again, this is one of the questions that the theory of mechanism design addresses.

Each of the procedures that you might consider in the two examples above creates a strategic game in the sense of non-cooperative game theory among the participants. Participants in these procedures will understand that the outcome will depend not only on their own choices but also on others' choices, and that therefore their own optimal strategy may depend on others' strategies. The theory of mechanism design therefore builds on the theory of games (Fudenberg and Tirole (1993)). Game theory takes the rules of the game as given, and makes predictions about the behavior of strategic players. The theory of mechanism design is about the optimal choice of the rules of the game.

We are more frequently involved in the design of rules for games than might be obvious at first sight. How should shareholders' votes be conducted? How should promotion procedures in companies be organized? What are optimal prenuptial agreements? All these questions are about the optimal rules of games. The theory of mechanism design seeks to study the general structure underlying all these applications, but it also considers a number of particularly prominent applications in detail.

There are at least two reasons why we study mechanism design. Firstly, the theory of mechanism design aids in practice the designers of real-world mechanisms. The theory of optimal auctions, for example, is frequently invoked in discussions about the design of government and industry auctions. Secondly, we can explain why real-world institutions are as they are by interpreting them as rational choices of those who designed them. For example, we might seek to explain the use of auctions in some house sales, and the use of posted prices in other house sales by appealing to the theory of mechanism design which indicates that posted prices are optimal in some circumstances and auctions are optimal in other circumstances.

The incentives created by the choice of rules of games are central to the theory of mechanism design. Incentives are also at the center of contract theory (Bolton and Dewatripont (2005)). At first sight the distinction between the theory of mechanism design and contract theory is simple: In contract theory, we study the optimal design of incentives for a single agent. In mechanism design, we study the optimal design of incentives for a group of agents, such as the buyers in our first example, and the colleagues in the second example. Contract theory therefore, unlike the theory of mechanism design, does not have to deal with strategic interaction.

The relation between contract theory and the theory of mechanism design is more subtle. One part of the theory of mechanism design is, in fact, a straightforward extension of insights from contract theory. This is surprising because one might have expected the element of strategic interaction, that is present in mechanism design, but absent in contract theory, to create substantial new problems. It is interesting and conceptually important to understand why this is not the case, and we shall address this issue in detail below. The close parallel between contract theory and mechanism design applies only to some parts of mechanism design. Other parts of mechanism design are, of course, unrelated to contract theory.

Contract theory has traditionally been divided into two parts: the theory of hidden information (also referred to as the theory of "adverse selection"), and the theory of hidden action (also referred to as the theory of "moral hazard"). The distinction is easily explained within the context of contracts for health insurance. Whether you have experienced severe chest pain in the past is something that you know, but that the company from which you are trying to buy health insurance does not know. It is "hidden information." Whether you exercise regularly, or take it a little more easy once you have bought complete insurance coverage for heart surgery, is a choice that you make that your insurance company does not observe unless it puts into place a surveillance operation. It is a "hidden action." Both hiddens, information and actions, matter for contract design. For example, by offering you a menu of insurance contracts and observing your choice from this menu an insurance company might be able to infer information about your health risks that you might wish to conceal from the company. By introducing deductibles, an insurance company might seek to maintain your incentives to look after your own health, and thus alleviate moral hazard problems.

Mechanism design, as traditionally understood, is about hidden information, not hidden actions, with multiple agents. In our first example, the hidden information that the seller of a house seeks to find out is the buyers' true willingness to pay for the house. In our second example in this section, the hidden information that we seek to find out is the colleagues' true willingness to pay for an office refrigerator. In voting, the hidden information that we seek to find out is individuals' true ranking of different alternatives or candidates. Of course, hidden action with many agents involved is a subject of great interest, and the theory that deals with it is concerned with the optimal choice of rules for a game, like the theory of mechanism design. For example, promotion schemes within a company set work incentives for a group of employees, and the optimal choice of such schemes is an important subject of economic theory. However, it is not the subject of mechanism design as the term has traditionally been interpreted.

When choosing the rules for the strategic interaction among agents, we might restrict ourselves to a small subset of all conceivable rules; or we might try to cast our net wide, and consider as large a set of rules as possible. For example, when considering how to auction your house, you might restrict attention to the choice of the minimum bid, and take for granted that the auction will proceed with all potential buyers submitting their bids simultaneously. You would then focus on the choice of only one parameter in a much larger set of possible choices. Alternatively, you might consider all conceivable ways of proceeding, not just auctions, but, for example, also simultaneous negotiations with all buyers that follow some predetermined format. It has been one of the accomplishments of the theory of mechanism design to develop a framework in which one can find the optimal rules of the game among all conceivable rules. Indeed, mechanism design has traditionally been understood as the field in which this grand optimization among all conceivable procedures is considered. In these notes, we shall stick to this interpretation.

Suppose you have considered all possible rules for proceeding with your house sale, and you have come to the conclusion that an auction with just one round of bidding is optimal. After the highest bid has been revealed, one of the losing bidders approaches you with a new and improved bid that is higher than the winning bid in the auction. Will you accept? This is an obvious temptation, but if you accept later bids, are you still conducting an auction with just a single round of bidding? In these notes we shall assume that the mechanism designer has full commitment power. The rules, once announced, are set in stone. The mechanism designer will not deviate from them. In our example, the mechanism designer will absolutely refuse to renegotiate after the auction results have been revealed. This is obviously a strong assumption. In contract theory, much attention has been given to the optimal design of contracts if full commitment cannot always be achieved, and this line of research has been very productive. It is also likely to be an interesting line of research in mechanism design. We do not consider this line of argument in these notes because we want to maintain a focus on the central arguments of the traditional theory of mechanism design, and these arguments have assumed full commitment by the mechanism designer.

These notes are meant for graduate students of economics who have a good understanding of game theory. Fudenberg and Tirole (1993) contains more than enough material for this course. We shall also assume a basic knowledge of real analysis that can, for example, be acquired from Royden (1988).

We proceed as follows: Chapter 2 presents some material from the theory of optimal contract design with hidden information. This theory is also referred to as the theory of "monopolistic screening." As we mentioned above, the classic Bayesian theory of optimal mechanism design can be viewed as an extension of the theory of monopolistic screening. We begin with the theory of monopolistic screening to make this connection transparent, and also because the basic results are most easily understood in the simple context of monopolistic screening.

Chapters 3 and 6 then review the classic Bayesian theory first in the context of some prominent examples, and then in general. This theory is built on several restrictive assumptions, which we shall discuss in detail. One of these assumptions is that for given rules of the mechanism, agents play a Bayesian equilibrium of the mechanism for a particular specification of agents' beliefs about each others' private information. This assumption might attribute more information to the mechanism designer than is realistic, and therefore the literature has sought to develop mechanisms that require the mechanism designer to know less about agents' beliefs. The classic approach to this problem is to seek dominant strategy mechanisms. We present this approach in Chapters 4, in the context of examples, and in Chapter 5, in general.

In Chapters 6.4, 7 and 8 we relax other assumptions of the classic model. In Chapter 9 we return to the issue of what the mechanism designer knows about agents' beliefs about each other and investigate more modern approaches to this problem which do not necessarily require the construction of a dominant strategy mechanism. Chapter 11 presents models of mechanism design that apply to dynamic contexts. Like robust mechanism design, this is an area of current research interest. Chapter 12 concludes the notes with a brief review and an outlook.

Chapter 2

Screening

2.1 Introduction

Important parts of the theory of mechanism design are the multi-agent extension of the theory of screening. We begin by explaining three examples from the theory of screening. We use these examples to introduce some topics and techniques that are also important in mechanism design. By introducing these topics and techniques in the context of screening, we explain them in the simplest possible context.

2.2 Pricing a Single Indivisible Good

A seller seeks to sell a single indivisible good. The seller herself does not attach any value to the good. Her objective is to maximize the expected revenue from selling the good. She is thus risk-neutral.

There is just one potential buyer. The buyer's von Neumann-Morgenstern utility if she purchases the good and pays a monetary transfer t to the seller is: $\theta - t$. The buyer's utility[1] if she does not purchase the good is zero. Here, $\theta > 0$ is a number that we can interpret as the buyer's valuation of the good, because our assumptions imply that the buyer is indifferent between paying θ and obtaining the good, and not obtaining the good.

Two aspects of the assumptions about the buyer's utility deserve emphasis. Firstly, we have assumed that the buyer's utility is the sum of the utility derived from the good, if it is purchased, and the disutility resulting

[1]From now on, when we refer to "utility" we shall mean "von Neumann-Morgenstern utility."

from the money payment. A more general formulation would write utility as $u(I,t)$ where I is an indicator variable that is 1 if the buyer purchases the good and 0 otherwise. Our assumption that u can be written as the sum of θ and $-t$ is usually described as the assumption that utility is "additively separable." Additive separability of the utility function implies that the buyer's utility from consuming the good is independent of the amount of money that he pays for it, and that the buyer's disutility from money payments is independent of whether or not he owns the good. One can easily think of real world contexts in which these assumptions are probably violated. But we shall stick to these assumptions for most of these notes. Much of the classical theory of mechanism design is based on these assumptions. We shall consider a more general case in Chapter 7.

The second aspect of our assumptions about the buyer's utility that deserves emphasis is that we have assumed that the buyer is risk-neutral with respect to money; that is, his utility is linear in money. Like additive separability, this is a very restrictive assumption, but much of the classical theory of mechanism design makes it. The more general model of Chapter 7 will relax this assumption, too.

We now introduce a crucial assumption about information. It is that the value of θ is known to the buyer, but it is not known to the seller. This seems plausible in many contexts. Buyers often know better than sellers how well some particular product meets their preferences. We shall refer below to θ as the buyer's "type."

As is often done in economic theory, we shall assume that the seller has a subjective probability distribution over possible values of θ. This probability distribution can be described by a cumulative distribution function F. We shall assume that F has a density f. Moreover, we shall assume that the support of F, that is, the smallest closed set that has probability 1, is an interval $[\underline{\theta}, \bar{\theta}]$, where $0 \leq \underline{\theta} < \bar{\theta}$. For technical convenience, we shall assume that the density is strictly positive on the support: $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. We can now think of θ as a random variable with the cumulative distribution function F, the realization of which is observed by the buyer but not by the seller.

Our interest is in procedures for selling the good which the seller should adopt to maximize expected profits. One obvious way would be to pick a price p and to say to the buyer that he can have the good if and only if he is willing to pay p. This is the selling procedure that we study in elementary microeconomics. Suppose the seller picks this procedure. Which price p should she choose? The probability that the buyer's value is below p is given by the value of the cumulative distribution function $F(p)$. The probability that the buyer's value is above p, and hence that he accepts a price offer p, is $1 - F(p)$. Expected revenue is therefore $p(1 - F(p))$, and the optimal strategy for the seller is to pick some price that maximizes $p(1 - F(p))$. Note that this is just the monopoly problem from elementary microeconomics with demand function $1 - F(p)$.

In this very simple context we shall now ask a straightforward question: "Is picking a price p really the best the seller can do?" What else could the seller do? The seller could, for example, negotiate with the agent. The seller could offer the agent a lottery where in return for higher or lower payments the buyer could be given a larger or smaller chance of getting the object. One can think of many other procedures that the seller might adopt to sell the good. Is setting a price really the best procedure?

To make our question more precise we have to be specific about which procedures the seller can commit to. We shall assume that the seller has unlimited powers of commitment: The seller can commit to an arbitrary extensive game tree where the players are the seller and the buyer, and where each terminal history is associated with a probability distribution over $\{0, 1\} \times \mathbb{R}$. The interpretation of such a probability distribution is that it describes the probability with which the object is transferred to the buyer together with a probability distribution over transfer payments by the buyer.

The seller will also find it to his advantage to commit to a strategy for himself, and to announce this strategy to the buyer before play begins. An example would be that the seller announces that he will bargain over the price with the seller, and that he announces in advance that he will turn down certain offers by the buyer. We shall assume that like the seller's ability to commit to a game, also the seller's ability to commit to a strategy in the game are unlimited.

Once the seller has committed to a game and a strategy, the buyer will choose his own strategy in the game. We shall assume that the buyer chooses his own strategy, knowing the value of θ, to maximize his expected utility.

Our question is now clearer: If the seller can commit to an arbitrary extensive game, and if she can also commit to a strategy for playing that game, which choice maximizes her expected revenue? In particular, is her expected revenue maximized by committing to a price, and by committing to selling the good at that price whenever the buyer is willing to pay the price?

A rather silly answer to our first question could be this: The seller should choose the game according to which the buyer has only a single choice that is such that the buyer does not get the object but nonetheless has to pay $\$x$ were x could be some arbitrarily large number. Clearly, if we allow such procedures, the seller can extract an arbitrarily large amount of money from the buyer. But the buyer wouldn't want to participate in such a mechanism. We will rule out such mechanisms by requiring that the buyer finds it in his interest to participate in the game proposed by the seller, regardless of his value of θ. In other words: For every type θ, the buyer will need to find that when he chooses his expected utility-maximizing strategy, his expected utility will be at least zero (the utility that he would obtain if he did not buy the good and did not pay anything). We shall refer to this constraint as the "individual rationality" constraint. Sometimes, it is also called "participation constraint."

Our objective is thus to study the optimization problem in which the seller's choice variables are an extensive game and a strategy in that game, in which the seller's objective function is expected revenue, and in which the constraint on the seller's choice is the individual rationality constraint.

At first sight, this looks like a hard problem, as the seller's choice set is very large. There are many extensive games that the seller could consider. However, a simple, yet crucial result enables us to get a handle on this optimization problem. The result says that we can restrict our attention to a small set of mechanisms, called "direct mechanisms."

Definition 2.1. [2] A "direct mechanism" consists of functions q and t where:

$$q : [\underline{\theta}, \bar{\theta}] \rightarrow [0, 1]$$

and

$$t : [\underline{\theta}, \bar{\theta}] \rightarrow \mathbb{R}.$$

The interpretation is that in a direct mechanism the buyer is asked to report θ. The seller commits to transferring the good to the buyer with probability $q(\theta)$ if the buyer reports that her type is θ, and the buyer has to pay the seller $t(\theta)$ if she reports that her type is θ. Note that the payment is deterministic. It is not conditional on the event that the buyer obtains the good. It would make no difference if we allowed the payment to be

[2] To be precise, this definition should include the requirement that the functions q and t are Lebesgue-measurable. We omit measurability requirements throughout this text, which will lead to a small cost in terms of rigor in some places.

random. All our analysis below would go through if we interpreted $t(\theta)$ as the buyer's expected payment conditional on θ. All selling mechanisms that are not direct mechanisms are called "indirect mechanisms."

A buyer's strategy σ in a direct mechanism is a mapping $\sigma : [\underline{\theta}, \bar{\theta}] \to [\underline{\theta}, \bar{\theta}]$ that indicates for every true type θ that the buyer might have the type $\sigma(\theta)$ that the buyer reports to the seller as his type.

The next result, a very simple version of the famous "Revelation Principle," shows that it is without loss of generality to restrict our attention to direct mechanisms.

Proposition 2.1 (Revelation Principle)**.** *For every mechanism Γ and every optimal buyer strategy σ in Γ there is a direct mechanism Γ' and an optimal buyer strategy σ' in Γ' such that:*

(i) The strategy σ' satisfies:

$$\sigma'(\theta) = \theta \text{ for every } \theta \in [\underline{\theta}, \bar{\theta}],$$

i.e. σ' prescribes telling the truth;

(ii) For every $\theta \in [\underline{\theta}, \bar{\theta}]$ the probability $q(\theta)$ and the payment $t(\theta)$ under Γ' equal the probability of purchase and the expected payment that result under Γ if the buyer plays her optimal strategy σ.

Proof. For every $\theta \in [\underline{\theta}, \bar{\theta}]$ define $q(\theta)$ and $t(\theta)$ as required by (ii) in Proposition 2.1. We prove the result by showing that for this direct mechanism the strategy $\sigma'(\theta) = \theta$, that is, truthfully reporting her type, is optimal for the buyer. Note that under this strategy, for every θ, the buyer with type θ obtains in the mechanism Γ' the same expected utility as in the mechanism Γ when choosing strategy $\sigma(\theta)$. Moreover, when pretending to be some type $\theta' \neq \theta$, the buyer obtains the same expected utility that she would have obtained had she played type θ's strategy $\sigma(\theta')$ in Γ. The optimality of truthfully reporting θ in Γ' then follows immediately from the optimality of $\sigma(\theta)$ in Γ. □

The Revelation Principle allows us to simplify our analysis greatly because it shows that without loss of generality we can restrict our search for optimal mechanisms to direct mechanisms, that is, pairs of functions q and t, where the buyer finds it always optimal to truthfully report her type.

Given a direct mechanism, we define define the buyer's expected utility $u(\theta)$ conditional on her type being θ by: $u(\theta) = \theta q(\theta) - t(\theta)$. Using this notation, we can now formally define the condition that the buyer finds it always optimal to truthfully report her type.

Definition 2.2. A direct mechanism is "incentive-compatible" if truth telling is optimal for every type, that is, if:

$$u(\theta) \geq \theta q(\theta') - t(\theta') \text{ for all } \theta, \theta' \in [\underline{\theta}, \bar{\theta}].$$

We mentioned before that it makes sense to also require that the buyer's expected utility from the mechanism is not lower than some lower bound, say zero. This requirement is captured in the following definition.

Definition 2.3. A direct mechanism is "individually rational" if the buyer, conditional on her type, is voluntarily willing to participate, that is, if:

$$u(\theta) \geq 0 \text{ for all } \theta \in [\underline{\theta}, \bar{\theta}].$$

Notice that in this definition we require voluntary participation *after agents have learned their types.* A weaker requirement would be to require voluntary participation *before agents have learned their types.*

We now consider in more detail the conditions under which a direct mechanism is incentive-compatible. Later we bring individual rationality into the picture.

Lemma 2.1. *If a direct mechanism is incentive-compatible, then q is increasing in θ.*[3]

Proof. Consider two types θ and θ' with $\theta > \theta'$. Incentive compatibility requires:

$$\theta q(\theta) - t(\theta) \geq \theta q(\theta') - t(\theta') \tag{2.1}$$
$$\theta' q(\theta) - t(\theta) \leq \theta' q(\theta') - t(\theta') \tag{2.2}$$

Subtracting these two inequalities we obtain

$$(\theta - \theta')q(\theta) \geq (\theta - \theta')q(\theta') \Leftrightarrow \tag{2.3}$$
$$q(\theta) \geq q(\theta') \tag{2.4}$$

□

[3] Throughout these notes we shall say that a function f is "increasing" if it is weakly monotonically increasing, i.e. if $x > x'$ implies $f(x) \geq f(x')$.

Lemma 2.2. *If a direct mechanism is incentive-compatible, then u is increasing. It is also convex, and hence differentiable except in at most countably many points. For all θ for which it is differentiable, it satisfies:*

$$u'(\theta) = q(\theta).$$

Note that the equation for $u'(\theta)$ is the same as the formula for the derivative of maximized utility functions in the "envelope theorem." For completeness, we provide a self-contained proof.

Proof. For all θ:

$$u(\theta) = \max_{\theta' \in [\underline{\theta}, \bar{\theta}]} (\theta q(\theta') - t(\theta')). \tag{2.5}$$

Given any value of θ, $\theta q(\theta') - t(\theta')$ is an increasing and affine (and hence convex) function. The maximum of increasing functions is increasing, and the maximum of convex functions is convex. Therefore, u is increasing and convex. Convex functions are differentiable in at most countably many points (Proposition 5.16 in Royden (1988)). Consider any θ for which u is differentiable. Let $\delta > 0$. Then:

$$\lim_{\delta \to 0} \frac{u(\theta + \delta) - u(\theta)}{\delta} \geq \lim_{\delta \to 0} \frac{((\theta + \delta)q(\theta) - t(\theta)) - (\theta q(\theta) - t(\theta))}{\delta} \tag{2.6}$$

$$= q(\theta). \tag{2.7}$$

Similarly:

$$\lim_{\delta \to 0} \frac{u(\theta) - u(\theta - \delta)}{\delta} \leq \lim_{\delta \to 0} \frac{(\theta q(\theta) - t(\theta)) - ((\theta - \delta)q(\theta) - t(\theta))}{\delta} \tag{2.8}$$

$$= q(\theta). \tag{2.9}$$

Putting the two inequalities together we obtain $u'(\theta) = q(\theta)$ whenever u is differentiable. □

The next lemma is essentially an implication of Lemma 2.2 and the fundamental theorem of calculus. In the proof, we take care of the possible lack of differentiability of u at some points.

Lemma 2.3 (Payoff Equivalence)**.** *Consider an incentive-compatible direct mechanism. Then for all $\theta \in [\underline{\theta}, \bar{\theta}]$:*

$$u(\theta) = u(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} q(x)dx.$$

Proof. The fact that u is convex implies, by Proposition 5.16 in Royden (1988), that it is absolutely continuous. By Proposition 5.13 in Royden (1988), this implies that it is the integral of its derivative. □

Lemma 2.3 shows that the expected utilities of the different types of the buyer are pinned down by the the function q and by the expected utility of the lowest type of the buyer, $u(\underline{\theta})$. Any two indirect mechanisms which, once the buyer optimizes, give rise to the same q and $u(\underline{\theta})$ therefore imply the same expected payoff for all types of the buyer. We have therefore indirectly shown a "payoff equivalence" result for classes of indirect mechanisms.

A short computation turns Lemma 2.3 into a result about the transfer payments that the buyer expects to make to the seller. This is shown in the next lemma.

Lemma 2.4 (Revenue Equivalence)**.** *Consider an incentive-compatible direct mechanism. Then for all* $\theta \in [\underline{\theta}, \bar{\theta}]$

$$t(\theta) = t(\underline{\theta}) + (\theta q(\theta) - \underline{\theta} q(\underline{\theta})) - \int_{\underline{\theta}}^{\theta} q(x)dx.$$

Proof. Recall that $u(\theta) = \theta q(\theta) - t(\theta)$. Substituting this into the formula in Lemma 2.3 and solving for $t(\theta)$ yields the result. □

Lemma 2.4 shows that the expected payments of the different buyer types are pinned down by the the function q and by the expected payment of the lowest type of the buyer, $t(\underline{\theta})$. For any given q and $t(\underline{\theta})$ there is thus one, and only one, incentive-compatible direct mechanism. Any two indirect mechanisms which, once the buyer optimizes, give rise to the same q and $t(\underline{\theta})$ therefore imply the same expected payment for all buyer types. For the seller, it follows that any two such indirect mechanisms yield the same expected revenue, because the seller's expected revenue is the expected value of the buyer's expected payments, where the seller takes expected values over the buyer's types. We have therefore indirectly shown a "revenue equivalence" result for classes of indirect mechanisms.

Lemma 2.4 is the famous "Revenue Equivalence Theorem" of auction theory adapted to our more simple setting of monopolistic screening. The full Revenue Equivalence Theorem, which will be discussed in Chapter 3, is based on essentially the same argument that we have explained here.

Lemmas 2.1 and 2.4 give necessary conditions for a direct mechanism to be incentive-compatible. It turns out that these conditions are also sufficient.

Proposition 2.2. *A direct mechanism* (q,t) *is incentive-compatible if and only if:*

(i) q *is increasing;*

(ii) For every $\theta \in [\underline{\theta}, \bar{\theta}]$*:*

$$t(\theta) = t(\underline{\theta}) + (\theta q(\theta) - \underline{\theta} q(\underline{\theta})) - \int_{\underline{\theta}}^{\theta} q(x)dx$$

Proof. To prove sufficiency, we have to show that no type θ prefers to pretend to be a type θ' if q is increasing and t is given by the formula in the Proposition:

$$\begin{aligned}
u(\theta) &\geq \theta q(\theta') - t(\theta') \Leftrightarrow && (2.10)\\
u(\theta) &\geq \theta q(\theta') - \theta' q(\theta') + \theta' q(\theta') - t(\theta') \Leftrightarrow && (2.11)\\
u(\theta) &\geq \theta q(\theta') - \theta' q(\theta') + u(\theta') \Leftrightarrow && (2.12)\\
u(\theta) - u(\theta') &\geq (\theta - \theta') q(\theta') \Leftrightarrow && (2.13)\\
\int_{\theta'}^{\theta} q(x)dx &\geq \int_{\theta'}^{\theta} q(\theta')dx && (2.14)
\end{aligned}$$

To obtain the left hand side of the last inequality we used the formula in Lemma 2.3. This formula is an implication of the formula for the payment in the proposition, as one can see by doing the calculation referred to in the proof of Lemma 2.4 in reverse order. Comparing the two integrals on the left hand side and the right hand side of the last inequality, suppose first that $\theta > \theta'$. The two integrals have the same integration limits, and the function $q(x)$ is everywhere at least as large as the constant $q(\theta')$ that is being integrated on the right hand side because q is increasing. Therefore, the integral on the left hand side is at least as large as the integral on the right hand side. If $\theta < \theta'$ the argument is analogous. □

We have now obtained a complete characterization of all incentive-compatible direct mechanisms. We now bring in individual rationality.

Proposition 2.3. *An incentive-compatible direct mechanism is individually rational if and only if* $u(\underline{\theta}) \geq 0$ *(or equivalently:* $t(\underline{\theta}) \leq \underline{\theta} q(\underline{\theta})$*).*

Proof. By Lemma 2.2 u is increasing in θ for incentive-compatible mechanisms. Therefore, $u(\theta)$ is non-negative for all θ if and only if it is non-negative for the lowest θ. □

We have now completely characterized the set of all direct mechanisms from which the seller can choose. We turn to the seller's problem of picking from this set the mechanism that maximizes expected revenue. We begin with the observation that it is optimal for the seller to set the lowest type's payment so that this type has zero expected utility.

Lemma 2.5. *If an incentive-compatible and individually rational direct mechanism maximizes the seller's expected revenue then*

$$t(\underline{\theta}) = \underline{\theta} q(\underline{\theta}).$$

Proof. By Proposition 2.3, we have to have: $t(\underline{\theta}) \leq \underline{\theta} q(\underline{\theta})$. If $t(\underline{\theta}) < \underline{\theta} q(\underline{\theta})$, then the seller could increase expected revenue by choosing a direct mechanism with the same q, but with a higher $t(\underline{\theta})$. By the formula for payments in Proposition 2.2, all types' payments would increase. □

Using Lemma 2.5, we can now simplify the seller's choice set further. The seller's choice set is the set of all increasing functions $q : [\underline{\theta}, \bar{\theta}] \to [0, 1]$. Given any such function, the seller will optimally set $t(\underline{\theta}) = \underline{\theta} q(\underline{\theta})$ so that the lowest type has zero expected utility, and all other types' payments are determined by the formula in Proposition 2.2. Substituting the lowest type's payment, this formula becomes:

$$t(\theta) = \theta q(\theta) - \int_{\underline{\theta}}^{\theta} q(x) dx, \tag{2.15}$$

i.e. type θ pays his expected utility from the good, $\theta q(\theta)$, minus a term that reflects a surplus that the seller has to grant to the buyer to provide incentives to the buyer to correctly reveal his type. This term is also called the buyer's "information rent."

To determine the optimal function q we could use equation (2.15) to obtain an explicit formula for the seller's expected revenue for any given function q. This is an approach that we shall take later in the next section. In the current context, an elegant argument based on convex analysis yields a simple and general answer. You need to be a little patient as the formal machinery that is needed for this approach is introduced.

We begin by considering in more detail the set of functions q that the seller can choose from. This set is a subset of the set of all bounded functions $f : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$. We denote this larger set by $\mathcal{F}$. We give $\mathcal{F}$ the linear structure:

$$g = \alpha f \Leftrightarrow [g(x) = \alpha f(x) \ \ \forall x \in [0,1]] \qquad \text{for all } \alpha \in \mathbb{R}, f, g \in \mathcal{F} \tag{2.16}$$

$$h = f + g \Leftrightarrow [h(x) = f(x) + g(x) \ \ \forall x \in [0,1]] \qquad \text{for all } f, g, h \in \mathcal{F} \tag{2.17}$$

This makes $\mathcal{F}$ a vector space. We also give $\mathcal{F}$ the L^∞-norm. Intuitively, this is the supremum of $|f(\theta)|$ over $\theta \in [\underline{\theta}, \bar{\theta}]$, where we neglect sets of measure zero. Formally, the L^∞-norm is defined by:[4]

$$\| f \| = \inf \{M \mid \mu(\{\theta \mid |f(\theta)| > M\}) = 0\} \qquad \text{for all } f \in \mathcal{F}. \tag{2.18}$$

where μ is the Lebesgue measure.

We denote by $\mathcal{M} \subset \mathcal{F}$ the set of all *increasing* functions in $\mathcal{F}$ such that $f(x) \in [0,1]$ $\forall x \in [0,1]$. This is the set from which the seller chooses. We now have a simple but crucial observation about $\mathcal{M}$.

Lemma 2.6. *$\mathcal{M}$ is compact and convex.*

Convexity is obvious as the convex combination of two increasing functions is increasing. Compactness is an implication of Helly's compactness theorem.[5]

Next, we have a closer look at the seller's objective function, which is the expected value of the right hand side of (2.15). The simple observation that we need is that this objective function is continuous and linear in q. From (2.15) we see that if we multiply q by a constant α, then $t(\theta)$ is multiplied by α for all θ. Therefore, the expected value of $t(\theta)$ is also multiplied by α, and expected revenue is a linear function of q.

We thus see that the seller maximizes a continuous linear function over a compact, convex set. A fundamental theorem of real analysis provides conditions under which the maximizers of a linear function over a convex set include "extreme points" of that set. This is intuitive. For example, you are probably familiar with the situation in linear programming in which the

[4]The details of this construction are in Royden (1988), p. 112. Note Royden's footnote 1 which indicates that to be rigorous we should define $\mathcal{F}$ to consist of equivalence classes of functions where two functions are in the same equivalence class if and only if they differ on a set of measure zero only. We shall not be precise about this, but neglecting this point leads to a certain lack of rigor in what follows.

[5]Doob (1994), p. 165.

set of optimal points includes at least one of the corner points. We shall apply this insight here. We first provide the definition of extreme points of a convex set. This is a generalization of the idea of corner points.

Definition 2.4. If C is a convex subset of a vector space X, then $x \in C$ is an extreme point of C if for every $y \in X, y \neq 0$ we have that either $x+y \notin C$ or $x - y \notin C$ or both.

So, for example in two-dimensional space the corners of a triangle are the extreme points of the triangle, and the circumference of a circle is the set of extreme points of the circle.

Now we can state the result that we will use to study the seller's optimal mechanism. It is called the "Extreme Point Theorem" in Ok (2007), p. 658.

Proposition 2.4. *Let X be a compact, convex subset of a normed vector space, and let $f : X \to \mathbb{R}$ be a continuous linear function. Then the set E of extreme points of X is non-empty, and there exists an $e \in E$ such that*

$$f(e) \geq f(x) \text{ for all } x \in X.$$

This result implies that a function q that is an extreme point of $\mathcal{M}$ and that maximizes expected revenue among all extreme points of $\mathcal{M}$ also maximizes expected revenue among all functions in $\mathcal{M}$. We may thus simplify the seller's problem further. Instead of considering all functions in the set $\mathcal{M}$, it is sufficient to consider only the set of all extreme points of $\mathcal{M}$. The following result characterizes the extreme points of $\mathcal{M}$.

Lemma 2.7. *A function $q \in \mathcal{M}$ is an extreme point of $\mathcal{M}$ if and only if $q(\theta) \in \{0, 1\}$ for almost all*[6] $\theta \in [\underline{\theta}, \bar{\theta}]$.

Proof. Consider any function as described in the lemma, and suppose that $\hat{q}$ is another function that satisfies: $\hat{q}(\theta) \neq 0$ for some θ.[7] If $\hat{q}(\theta) > 0$ and $q(\theta) = 0$ then $q(\theta) - \hat{q}(\theta) < 0$, and hence $q - \hat{q} \notin \mathcal{M}$. If $\hat{q}(\theta) > 0$ and $q(\theta) = 1$ then $q(\theta) + \hat{q}(\theta) > 1$, and hence $q + \hat{q} \notin \mathcal{M}$. The case $\hat{q}(\theta) < 0$ is analogous.

Now consider any function q that is not as described in the lemma, i.e. there is some θ^* such that $q(\theta^*) \in (0, 1)$.[8] Define $\hat{q}(\theta) = q(\theta)$ if $q(\theta) \leq 0.5$, and $\hat{q}(\theta) = 1 - q(\theta)$ if $q(\theta) > 0.5$. Clearly, $\hat{q} \neq 0$. Consider now first the function $q + \hat{q}$. We have that $q(\theta) + \hat{q}(\theta) = 2q(\theta)$ if $q(\theta) \leq 0.5$, and $q(\theta) + \hat{q}(\theta) = 1$ if $q(\theta) > 0.5$. Thus, evidently, $q \in \mathcal{M}$. The argument for $q - \hat{q}$ is analogous. We conclude that q is not an extreme point of M. □

[6] Where "almost all θ" means "for a set of θ with Lebesgue measure $\bar{\theta} - \underline{\theta}$."

[7] More precisely: $\hat{q}(\theta) \neq 0$ for a set of θ with strictly positive Lebesgue measure.

[8] More precisely: $q(\theta^*) \in (0, 1)$ for a set of θ^* with strictly positive Lebesgue measure.

The seller can thus restrict her attention to non-stochastic mechanisms. But a non-stochastic mechanism is monotone if and only if there is some $p \in [\underline{\theta}, \bar{\theta}]$ such that $q(\theta) = 0$ if $\theta < p^*$ and $q(\theta) = 1$ if $\theta > p^*$. This direct mechanism can be implemented by the seller simply quoting the price p^* and the buyer either accepting or rejecting p^*. Our results therefore imply that the seller cannot do better than quoting a simple price p^* to the buyer. This analysis is summarized in the following proposition.

Proposition 2.5. *A direct mechanism maximizes the seller's expected revenues among all incentive-compatible, individually rational direct mechanisms if and only if there is a* $p^* \in argmax_{p \in [\underline{\theta}, \bar{\theta}]} p(1 - F(p))$ *such that*

$$q(\theta) = \begin{cases} 1 & \text{if } \theta > p^*; \\ \\ 0 & \text{if } \theta < p^*, \end{cases}$$

and

$$t(\theta) = \begin{cases} p^* & \text{if } \theta > p^*; \\ \\ 0 & \text{if } \theta < p^*. \end{cases}$$

Proof. As argued above, we only need to consider functions q where the buyer obtains the good with probability 1 if his value is above some price p^*, and with probability 0 if his value is below this price. The optimal function q of this form is obviously the one indicated in the Proposition. The formula for t follows from Proposition 2.2. □

It may seem that we have gone to considerable length to derive a disappointing result, namely, a result that does not offer the seller any more sophisticated selling mechanisms than we are familiar with from elementary microeconomics. However, apart from introducing some technical tools that we use later in more complicated contexts, the reader should appreciate that we have uncovered a rather sophisticated rationale for a familiar everyday phenomenon. This is perhaps analogous to invoking Newton's law of gravity as an explanation of the fact that apples fall from apple trees. The fact is familiar, but the explanation is non-obvious.

2.3 Nonlinear Pricing

Now we study a model in which a monopolist offers an infinitely divisible good, say sugar, to one potential buyer. We introduce this model because it is more commonly studied than the model in the previous section, and because its analysis introduces additional elements, beyond those presented in the previous section, that will reappear in almost exactly the same form in the analysis of optimal mechanisms.

For simplicity we assume that production costs are linear; that is, producing quantity $q \geq 0$ costs cq, where $c > 0$ is a constant. The seller is risk-neutral, so that she seeks to maximize her expected revenue. The buyer's utility from buying quantity $q \geq 0$ and paying a monetary transfer t to the monopolist is $\theta\nu(q) - t$. We assume that $\nu(0) = 0$ and that ν is a twice-differentiable, strictly increasing and strictly concave function: $\nu'(q) > 0, \nu''(q) < 0$ for all $q \geq 0$.

Because $\nu(0) = 0$ the buyer's utility when buying nothing and paying nothing is zero. We can interpret $\theta\nu(q)$ as the buyer's willingness to pay for quantity q. Note that we have assumed, as in the previous section, that utility is additively separable in consumption of the good and money, and that the consumer is risk-neutral in money.

The parameter θ reflects how much the consumer values the good. More precisely, the larger the value of θ, the larger is the consumer's absolute willingness to pay $\theta\nu(q)$ and the consumer's marginal willingness to pay $\theta\nu'(q)$ for any given quantity q. The parameter θ can take any value between $\underline{\theta}$ and $\bar{\theta}$. The value of θ is known to the buyer but not to the seller. The seller's beliefs about θ are given by a cumulative distribution function F with density f on the interval $[\underline{\theta}, \bar{\theta}]$. We assume that f satisfies: $f(\theta) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$.

We assume that $\bar{\theta}\nu'(0) > c$. This means that the largest marginal willingness to pay that the buyer might possibly have is above the marginal cost of production. This assumption ensures that the seller and the buyer have an incentive to trade at least for the largest type of the buyer. A final assumption is that $\lim_{q\to\infty} \bar{\theta}\nu'(q) < c$. This means that even the highest type's marginal willingness to pay falls below c as q gets large. This assumption ensures that the quantity that the seller supplies to the buyer is finite for all possible types of the buyer.

We seek to determine optimal selling procedures for the seller. As in the previous section, the revelation principle holds and we can restrict our attention to direct mechanisms. In the current context, we use the following definition of direct mechanisms:

Definition 2.5. A "direct mechanism" consists of functions q and t where:

$$q : [\underline{\theta}, \bar{\theta}] \rightarrow \mathbb{R}_+$$

and

$$t : [\underline{\theta}, \bar{\theta}] \rightarrow \mathbb{R}.$$

The interpretation is that the buyer is asked to report θ, and the seller commits to selling quantity $q(\theta)$ to the buyer and the buyer commits to paying $t(\theta)$ to the seller. Note that we use the same notation as in the previous section, but that in this section $q(\theta)$ is a quantity, whereas in the previous section $q(\theta)$ was a probability. In this section we ignore stochastic mechanisms—that is, we assume that for each type θ the quantity sold to the buyer if he is of type θ is a non-negative number, not a probability distribution over non-negative numbers. We make this assumption for simplicity. It is non-trivial to study stochastic direct mechanisms in our context.

We do not state the revelation principle formally for our context. It is analogous to the revelation principle in the previous section. One modification is needed, however, in the statement of the result. To obtain deterministic direct mechanisms, we have to restrict our attention to general mechanisms and twice-differentiable strategies that result in a deterministic quantity for the buyer for each type of the buyer.

As in the previous section we can then study incentive compatibility and individual rationality of direct mechanisms. The analysis proceeds along exactly the same lines as in the previous section, and we shall just state the result of the analysis. One finds that a direct mechanism (q, t) is incentive-compatible if and only if

(i) q is increasing;

(ii) For every $\theta \in [\underline{\theta}, \bar{\theta}]$:

$$t(\theta) = t(\underline{\theta}) - \underline{\theta}\nu(q(\underline{\theta})) + \theta\nu(q(\theta)) - \int_{\underline{\theta}}^{\theta} \nu(q(x))dx. \tag{2.19}$$

An incentive-compatible mechanism is individual rational if and only if

$$t(\underline{\theta}) \leq \underline{\theta}\nu(q(\underline{\theta})). \tag{2.20}$$

The seller's decision problem is to pick among all direct mechanisms satisfying these two conditions the one that maximizes expected revenue. It is obvious that the seller will choose $t(\underline{\theta})$ so that the utility of type $\underline{\theta}$ is zero, that is:

$$t(\underline{\theta}) = \underline{\theta}\nu(q(\underline{\theta})). \tag{2.21}$$

Substituting this into equation (2.19) yields:

$$t(\theta) = \theta\nu(q(\theta)) - \int_{\underline{\theta}}^{\theta} \nu(q(x))dx. \tag{2.22}$$

The choice that remains to be studied is that of the function q.

At this point we depart from the line of argument that we followed in the previous section. The reason is that the seller's objective function is no longer linear in q. This is clear from equation (2.22) where q enters the non-linear function ν. Because the objective function is not linear in q, the extreme point argument of the previous section does not apply here. We shall use instead equation (2.22) to study in more detail the seller's expected profit. If the seller chooses $q(\cdot)$, then his expected profit is:

$$\begin{aligned}
&\int_{\underline{\theta}}^{\bar{\theta}} \left[\theta\nu(q(\theta)) - \int_{\underline{\theta}}^{\theta} \nu(q(x))dx - cq(\theta)\right] f(\theta)d\theta \\
= &\int_{\underline{\theta}}^{\bar{\theta}} \theta\nu(q(\theta))f(\theta)d\theta - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} \nu(q(x))dxf(\theta)d\theta - \int_{\underline{\theta}}^{\bar{\theta}} cq(\theta)f(\theta)d\theta
\end{aligned} \tag{2.23}$$

We seek to simplify the expression in (2.23). The calculation that follows, although it contains no conceptually or mathematically deep insights, appears in this or in similar form frequently in the theory of mechanism design. It is therefore worthwhile to consider it in detail. We focus initially on the double integral in the second term in (2.23).

$$\begin{aligned}
&\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} \nu(q(x))dxf(\theta)d\theta \\
= &\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} \nu(q(x))f(\theta)dxd\theta \\
= &\int_{\underline{\theta}}^{\bar{\theta}} \int_{x}^{\bar{\theta}} \nu(q(x))f(\theta)d\theta dx \\
= &\int_{\underline{\theta}}^{\bar{\theta}} \nu(q(x)) \int_{x}^{\bar{\theta}} f(\theta)d\theta dx \\
= &\int_{\underline{\theta}}^{\bar{\theta}} \nu(q(x))(1 - F(x))dx
\end{aligned}$$

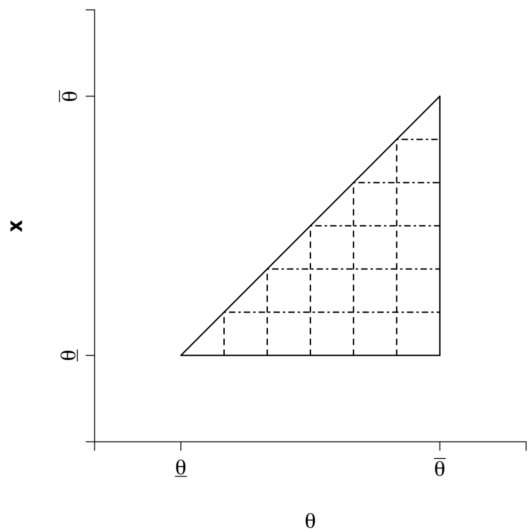

Figure 2.1: Changing the order of integration

$$= \int_{\underline{\theta}}^{\bar{\theta}} \nu(q(\theta))(1 - F(\theta))d\theta \tag{2.24}$$

When moving from the second to the third line in (2.24) we change the order of integration. By Fubini's theorem, this leaves the value of the integral unchanged. We indicate in Figure 2.1 the change in the order of integration. In the second line we first integrate along the vertical lines and then horizontally. In the third line we first integrate along the horizontal lines in Figure 2.1 and then vertically.

We now substitute the last line in (2.24) into the seller's objective function in (2.23).

$$\begin{aligned}
&\int_{\underline{\theta}}^{\bar{\theta}} \left[\theta\nu(q(\theta)) - cq(\theta)\right] f(\theta)d\theta - \int_{\underline{\theta}}^{\bar{\theta}} \nu(q(\theta))(1 - F(\theta))d\theta \\
&= \int_{\underline{\theta}}^{\bar{\theta}} \left[\theta\nu(q(\theta)) - cq(\theta)\right] f(\theta)d\theta - \int_{\underline{\theta}}^{\bar{\theta}} \nu(q(\theta))\frac{1 - F(\theta)}{f(\theta)} f(\theta)d\theta \\
&= \int_{\underline{\theta}}^{\bar{\theta}} \left[\nu(q(\theta))\left(\theta - \frac{1 - F(\theta)}{f(\theta)}\right) - cq(\theta)\right] f(\theta)d\theta.
\end{aligned} \tag{2.25}$$

The seller thus chooses q to maximize the expected value of the expression that is in large square brackets in (2.25). Expected values are taken over θ. The seller must choose an increasing function q.

Suppose we ignore for the moment the constraint that q must be increasing. Then the seller can choose $q(\theta)$ for each θ separately to maximize the expression in the large square brackets. This choice of q also maximizes the expected value of that expression. We study this approach to the choice of q first, and then impose a condition that makes sure that the function q that we find is indeed increasing. The first-order condition for maximizing the expression in square brackets for given θ is:

$$\nu'(q(\theta))\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) - c = 0 \Leftrightarrow \tag{2.26}$$

$$\nu'(q(\theta))\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) = c \tag{2.27}$$

We now investigate the existence of a solution to (2.27). If

$$\theta - \frac{1-F(\theta)}{f(\theta)} \leq 0 \tag{2.28}$$

then there is obviously no solution, and the optimal choice is

$$q(\theta) = 0. \tag{2.29}$$

Now consider

$$\theta - \frac{1-F(\theta)}{f(\theta)} > 0. \tag{2.30}$$

Recall that we have assumed that ν' is differentiable, and therefore continuous, and decreasing, and that $\bar{\theta}\nu(q)$ tends to less than c as q tends to infinity. Obviously, the left hand side of (2.27) shares all these properties. If

$$\nu'(0)\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) \leq c \tag{2.31}$$

then it is again obvious that the optimal choice is:

$$q(\theta) = 0. \tag{2.32}$$

If

$$\nu'(0)\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) > c \tag{2.33}$$

then our assumptions imply that there is a unique solution to (2.27), and that this stationary point is also the unique optimal choice of $q(\theta)$.

We have now determined for each θ the choice of $q(\theta)$ that maximizes the expression in large square brackets in (2.25). The seller seeks to maximize the expected value of this expression, and the seller is constrained to choose a function q that is increasing. If the function q that we have determined above is increasing, then it must be the optimal choice for the seller. We now introduce the following assumption, which implies that the q that we have determined is increasing.

Assumption 2.1. $\theta - \frac{1-F(\theta)}{f(\theta)}$ *is increasing in* θ*.*

To verify that Assumption 2.1 implies that q is increasing, note that Assumption 2.1 implies that the left hand side of (2.27) is increasing in θ for every q. The optimal q is the intersection point of that expression with c or zero, whatever is greater. It is then easy to see that the optimal q is increasing in θ.

A sufficient condition for Assumption 2.1 is that $\frac{f(\theta)}{1-F(\theta)}$ is increasing in θ. This sufficient condition is often referred to as the "increasing hazard rate" condition. Think of $F(\theta)$ as the probability that an individual dies before time θ. Then $1 - F(\theta)$ is the probability that the individual survives until time θ, and $\frac{f(\theta)}{1-F(\theta)}$ can be thought of as the conditional probability of dying at time θ of an individual that has survived until time θ. The sufficient condition is that this conditional probability of dying, one of the unavoidable hazards of life, is increasing in θ.

Many commonly considered distributions F satisfy Assumption 2.1. We shall refer to such distributions as "regular." The analysis of this section is summarized in the following proposition:

Proposition 2.6. *Suppose that* F *is regular. Then an expected profit maximizing choice of* q *is given by:*

(i) if $\nu'(0)\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) \leq c$*:*

$$q(\theta) = 0;$$

(ii) otherwise:

$$\nu'(q(\theta))\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) = c.$$

The profit maximizing t is given by:

$$t(\theta) = \theta\nu(q(\theta)) - \int_{\underline{\theta}}^{\theta} \nu(q(x))dx.$$

To understand the economic meaning of Proposition 2.6, note that for the highest type $\theta = \bar{\theta}$ we have: $1 - F(\theta) = 0$. Therefore, the second of the two cases in Proposition 2.6 applies to $\bar{\theta}$, and $q(\bar{\theta})$ is determined by:

$$\nu'(q(\theta))\bar{\theta} = c. \tag{2.34}$$

This equation shows that the highest type is supplied the quantity at which this type's marginal willingness to pay is exactly equal to the marginal cost of production. This is the quantity that this type would choose to produce if he owned the firm. We refer to this quantity as the "first-best" quantity. For all lower types $\theta < \bar{\theta}$, the quantity supplied to these types is determined by equation (2.27), which differs from the first-best condition (2.34) (with $\bar{\theta}$ replaced by θ) in that the left hand side is smaller for every q. Thus, the marginal costs are not equated with the marginal benefits, but with a quantity smaller than the marginal benefits. This means that all types that are lower than $\bar{\theta}$ are offered a quantity that is smaller than the "first-best" quantity.

We conclude with a numerical example.

Example 2.1. $c = 1$, $\nu(q) = \sqrt{q}$, *θ is uniformly distributed on* $[0, 1]$, *i.e.:* $F(\theta) = \theta$ *and* $f(\theta) = 1$ *for all* $\theta \in [0, 1]$. *To verify that Assumption 2.1 is satisfied, we have to check that the following expression is increasing in θ:*

$$\theta - \frac{1 - F(\theta)}{f(\theta)} = \theta - \frac{1 - \theta}{1} = 2\theta - 1, \tag{2.35}$$

which is obviously the case.

Next we determine for which values of θ the optimal quantity $q(\theta)$ equals zero:

$$\begin{aligned}
\nu'(0)\left(\theta - \frac{1 - F(\theta)}{f(\theta)}\right) &\leq c \Leftrightarrow \\
\theta - \frac{1 - F(\theta)}{f(\theta)} &\leq 0 \Leftrightarrow \\
2\theta - 1 &\leq 0 \Leftrightarrow \\
\theta &\leq 0.5
\end{aligned} \tag{2.36}$$

The first and the second line are equivalent because in our example: $\nu'(0) = +\infty$. *If* $\theta > 0.5$, *the optimal* $q(\theta)$ *is given by:*

$$\begin{aligned} \nu'(q(\theta))\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) = c \Leftrightarrow \\ \frac{1}{2\sqrt{q}}(2\theta - 1) = 1 \Leftrightarrow \\ \sqrt{q} = \theta - \frac{1}{2} \Leftrightarrow \\ q = \left(\theta - \frac{1}{2}\right)^2. \end{aligned} \tag{2.37}$$

The corresponding transfer $t(\theta)$ *is zero if* $\theta \leq 0.5$, *and if* $\theta > 0.5$ *it is given by:*

$$\begin{aligned} t(\theta) &= \theta\nu(q(\theta)) - \int_{\underline{\theta}}^{\theta} \nu(q(x))dx \\ &= \theta\left(\theta - \frac{1}{2}\right) - \int_{0.5}^{\theta} x - \frac{1}{2}dx \\ &= \theta\left(\theta - \frac{1}{2}\right) - \left[\frac{1}{2}x^2 - \frac{1}{2}x\right]_{0.5}^{\theta} \\ &= \theta\left(\theta - \frac{1}{2}\right) - \left(\frac{1}{2}\theta^2 - \frac{1}{2}\theta - \frac{1}{8} + \frac{1}{4}\right) \\ &= \frac{1}{2}\theta^2 - \frac{1}{8} \end{aligned} \tag{2.38}$$

We want to translate the solution into an optimal non-linear pricing scheme. We can express the transfer t *as a function of* q. *For this, we first determine which type* θ *purchases any given quantity* q:

$$\begin{aligned} q(\theta) = q \Leftrightarrow \\ \left(\theta - \frac{1}{2}\right)^2 = q \Leftrightarrow \\ \theta - \frac{1}{2} = \sqrt{q} \Leftrightarrow \\ \theta = \sqrt{q} + \frac{1}{2} \end{aligned} \tag{2.39}$$

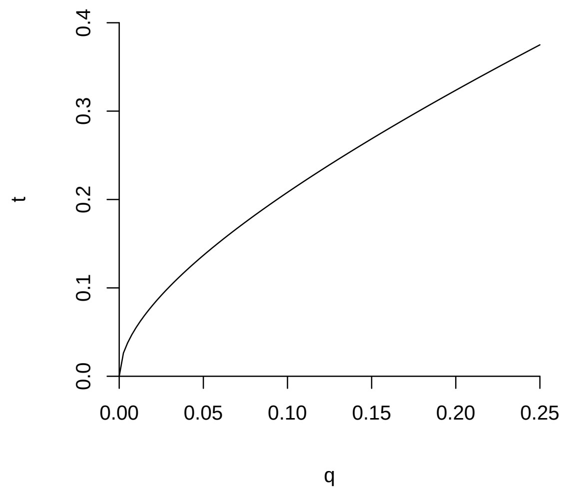

Figure 2.2: The optimal non-linear pricing scheme

The payment by type θ is:

$$\begin{aligned} t(\theta) &= \frac{1}{2}\theta^2 - \frac{1}{8} \\ &= \frac{1}{2}\left(\sqrt{q} + \frac{1}{2}\right)^2 - \frac{1}{8} \\ &= \frac{1}{2}q + \frac{1}{2}\sqrt{q} \end{aligned} \tag{2.40}$$

Essentially, the monopolist thus offers to consumers the deal that they can buy any quantity $q \in [0, \frac{1}{4}]$. Their payment is $t(q) = \frac{1}{2}q + \frac{1}{2}\sqrt{q}$. This optimal non-linear pricing scheme is shown in Figure 2.2. We note that there is a quantity discount. The per unit price:

$$\frac{t}{q} = \frac{\frac{1}{2}q + \frac{1}{2}\sqrt{q}}{q} = \frac{1}{2} + \frac{1}{2\sqrt{q}} \tag{2.41}$$

decreases in q.

2.4 Bundling

The theory of screening can be extended into many different directions. For example, it is of great interest to consider the case in which the buyer's private information is also of relevance to the seller's assessment of a possible sale. Insurance contracts are an example. A buyer's private information about his health situation affects not only his own evaluation of an insurance contract, but also the insurance seller's evaluation of the contract. We shall not pursue this direction here because it is not related to the theory of mechanism design as far as it is covered in these notes.

Another important extension is to the case in which the buyer's private information is multi-dimensional. The case of multi-dimensional private information is also relevant in the theory of mechanism design. Therefore, we give here a simple example of screening when the buyer has multi-dimensional private information.

Suppose a seller has two distinct indivisible goods, good A and good B, for sale. For simplicity, we assume that the seller values the goods at zero and is risk-neutral, so that she seeks to maximize her expected revenue. Let I_A be an indicator variable equal to 1 if the buyer obtains good A, and 0 otherwise. Define I_B similarly. Denote by t the monetary transfer from the buyer to the seller. Then the buyer's utility is:

$$I_A v_A + I_B v_B - t. \tag{2.42}$$

Note that utility is additive in the two goods and in money. We assume in addition that the marginal value of each good does not depend on whether the other good is also obtained. In this way the two goods are entirely independent. They are not like pasta and tomato sauce, but they are like pasta and a watch.

The parameters v_A and v_B indicate the buyer's willingness to pay for the two goods. These parameters are known to the buyer but not known to the seller. The seller's belief about these two parameters is given by the uniform distribution F over the unit square $[0,1]^2$. Note that we assume here that v_A and v_B are stochastically independent. Thus, we assume for a second time that there is no relation at all between the two goods.

Our interest is again in optimal selling procedures for the seller. As in the previous sections, the revelation principle holds and we could restrict our attention to direct mechanisms. We shall instead simplify our problem much more, and only consider a very small class of indirect mechanisms.

Suppose the seller considers quoting three prices: p_A, p_B and p_{AB}. The interpretation is that the buyer can buy good A at price p_A, good B at price p_B, or goods A and B at price p_{AB}. We assume that the seller cannot stop the buyer from buying goods A and B at price $p_A + p_B$, so the price p_{AB}, if it is to have any effect, has to satisfy: $p_{AB} \leq p_A + p_B$.

What is the optimal choice of p_A, p_B and p_{AB}? This is a simple calculus exercise. It turns out that the optimal prices are:

$$\begin{aligned} p_A = p_B &= \frac{2}{3} \\ p_{AB} &= \frac{1}{3}\left(4 - \sqrt{2}\right) \approx 0.862 \end{aligned} \tag{2.43}$$

Note that the optimal price P_{AB} is indeed strictly smaller than $p_A + p_B$. The seller thus offers to the buyer that he can buy the two goods separately, but that he gets a better deal if he buys the two goods together. The literature refers to the combination of goods A and B as a "bundle." The monopolist's strategy in our example is also described as "mixed bundling" because the monopolist offers the bundle to consumers, but he also offers to them the option to buy the goods individually.

The buyer's demand behavior given the optimal prices is shown in Figure 2.3. Depending on the value of v_A and v_B the buyer purchases good A only, good B only, both goods, or no good. Efficiency would, of course, require that the good is transferred to the buyer for all values of v_A and v_B except zero. We thus observe in Figure 2.3 several distortions of efficiency.

Figure 2.3 illustrates the direct mechanism implemented by the seller if he quotes the three optimal prices. We present this example mainly to illustrate that even in the simple case of screening, multi-dimensional private information may cause surprising and counterintuitive effects. It is very surprising that the seller offers the goods as a bundle at a discount even though from the consumer's point of view the goods are entirely unrelated. The literature has, in fact, spent some time seeking to understand the intuition behind this effect. As this example indicates, the general theory of screening with multiple goods and multi-dimensional private information is rather complicated.

2.5 Comments on the Literature

Our exposition in Section 2.2 is a modified version of Manelli and Vincent (2007). The theory of non-linear pricing in Section 2.3 is discussed further in

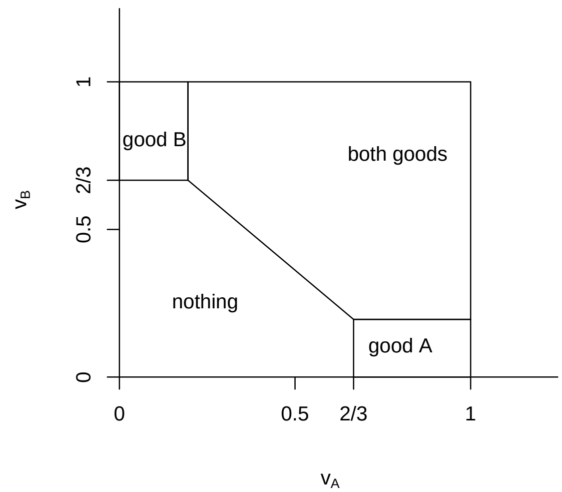

Figure 2.3: Buyer behavior given optimal prices

Bolton and Dewatripont (2005). Finally, the example in Section 2.4 is based on Chapter 6 of Hermalin (2005). The seminal paper on mixed bundling is Adams and Yellen (1976). Manelli and Vincent (2007) applies the extreme point analysis of Section 2.2 to a general screening model and illustrates the potentially complicated stochastic nature of optimal selling mechanisms.

2.6 Problems

a) Give an example in the setting of Section 2.2 in which the buyer has only two possible types, and in which the revenue equivalence principle does not hold.

b) Does Proposition 2.5 hold if the type distribution F is discrete?

c) Prove that the conditions in (2.19) are sufficient for incentive compatibility.

d) Prove that the prices in (2.43) maximize the seller's profits.

Chapter 3

Classic Bayesian Mechanism Design: Examples

3.1 Introduction

This chapter describes three classic mechanism design problems using Bayesian Nash equilibrium to predict agents' strategic behavior for any given mechanism. The next chapter will consider the same examples, but we shall use dominant strategies as our concept for predicting agents' behavior. We start with examples to illustrate the general analysis that will follow in later chapters. The examples are also interesting in their own right. We contrast optimal Bayesian mechanisms with optimal dominant-strategy mechanisms to illustrate the adjustments that need to be made to a mechanism if the weaker condition of Bayesian incentive compatibility is replaced by the more restrictive condition of dominant strategy incentive compatibility. Understanding this contrast will prepare the reader for our later discussion of robust mechanism design.

3.2 Single Unit Auctions

3.2.1 Set-Up

The model in this chapter is the same as in Chapter 2.2, except that we now have more than one potential buyer. A seller seeks to sell a single indivisible good. There are $N \geq 2$ potential buyers. We denote the set of potential buyers by $I = \{1, 2, ..., N\}$. Buyer i's utility if she purchases the good and pays a transfer t_i to the seller is: $\theta_i - t_i$. Buyer i's utility if she does not purchase the good and pays a transfer of t_i to the seller is: $0-t_i$. The seller's utility if he obtains transfers t_i from buyers $i = 1, 2, \ldots, N$ is: $\sum_{i\in I} t_i$.

We assume that buyer i knows θ_i, but that neither the seller nor any other buyer $j \neq i$ knows θ_i. We model the valuation θ_i as a random variable with cumulative distribution function F_i with density f_i. The support of θ_i is $[\underline{\theta}, \bar{\theta}]$ where $0 \leq \underline{\theta} < \bar{\theta}$. Thus, although we do not assume that the random variables θ_i have the same distribution for different i, we do assume that they have the same support. This is for convenience only. For technical convenience, we also assume that $f_i(\theta_i) > 0$ for all $i \in I$ and all $\theta_i \in [\underline{\theta}, \bar{\theta}]$.

We assume that for $i, j \in I$ with $i \neq j$, the random variables θ_i and θ_j are independent. We denote by θ the vector $(\theta_1, \theta_2, \ldots, \theta_N)$. The support of the random variable θ is $\Theta \equiv [\underline{\theta}, \bar{\theta}]^N$. The distribution of θ is denoted by F, which is the product of the distributions F_i, and the density of θ is denoted by f. Each potential buyer i observes θ_i, but neither the seller nor the other buyers $j \neq i$ observe θ_i. The distribution F, however, is common knowledge among the buyers and the seller.

The model that we have described is known in the literature as a model with "independent private values." One also refers to the "independent private values" assumption. All examples in this chapter will be built on this assumption. The assumption is very restrictive. We discuss why in this paragraph and the next. The phrase "independent" refers to the fact that we have assumed that values are independent and that they follow a commonly known prior distribution F. Note first that the distribution F describes not only the seller's beliefs, as in Chapter 2.2, but also the potential buyers' beliefs about each other. The assumption that buyers' values are independent implies that each buyer's beliefs about the other buyers' values is independent of his own value. So, for example, if buyer i has a high value, he does not attach more probability to the event that buyer $j \neq i$ has a high value than if i had had a low value. The assumption that F is a *common* prior of the seller and all buyers implies that two buyers i, i' with $i, i' \neq j$ have the same belief about buyer j's value, and that this belief is also shared by the seller. The assumption that F is common knowledge implies that it is commonly known among the sellers and the buyers that they share the same beliefs about other buyers.

We have assumed *private values* in the sense that each buyer's private information is sufficient to determine this buyer's value of the good. No buyer would change his value of the good if he knew what other buyers know. Thus, the private information that leads one buyer to value the good highly (or not) would not change any other buyer's value if it was known to that buyer. The most plausible interpretation is that the private information is about each buyer's private tastes rather than about objective features of the good.

3.2.2 Mechanisms, Direct Mechanisms, and the Revelation Principle

We will be interested in procedures that the seller can use to sell his good. For example, he could pick a price, ask each buyer to indicate whether she is willing to pay this price for the good, and then randomly pick one of the buyers, if any, who have said that they are willing to buy the good and transact with this buyer at the announced price.

We will consider a much more general class of methods for selling the good. We will allow the seller to pick an arbitrary extensive game tree where the players are the potential buyers and the seller. The seller assigns to each terminal history of the game (of finite or infinite length) an outcome, that is a probability of transferring the good, and if so to whom, and a probability distribution over vectors of transfers from the buyers to the seller. Formally an outcome is a probability distribution over $\{0, 1, 2, \ldots, N\} \times \mathbb{R}^N$. Here, 0 stands for the outcome that the good remains with the seller and is hence not sold.

The seller will find it to her advantage to commit in advance to a strategy. Therefore, we might as well eliminate the seller as a player from the game, and restrict our attention to extensive game trees where only the potential buyers are players. We shall understand by a "mechanism" such an extensive game tree together with an assignment of a probability distribution over outcomes to each terminal history. For simplicity, we shall not provide a formal definition of a "mechanism."

A mechanism, in conjunction with the assumptions about utilities, information, and the distribution of types that we made in the previous subsection defines a game of incomplete information. The standard solution concept for such games is that of a Bayesian Nash equilibrium (Fudenberg and Tirole (1993), p. 215). Note that games may have none, one, or more than one, Bayesian Nash equilibrium. We shall imagine that the seller only proposes games that do have at least one Bayesian Nash equilibrium, and that, when announcing the mechanism, the seller also proposes a Bayesian Nash equilibrium of the corresponding game. The buyers will play the equilibrium that the seller proposes. Thus, if there are multiple equilibria, the seller can in a sense "pick" which equilibrium the buyers will play. This assumption is important for the revelation principle, and we will comment further on it when discussing the revelation principle below.

We assume that the utility that buyers obtain if they walk away from the mechanism proposed by the seller is zero. Participation in the mechanism and the equilibrium that the seller proposes must be voluntary, and therefore we assume that the equilibrium that the seller proposes must offer each potential buyer an expected utility of at least zero.

We now introduce a subclass of mechanisms, "direct mechanisms," and then show that we can restrict our attention to such mechanisms without loss of generality. In the following definition Δ denotes the set of all probability distributions over the set I of buyers to whom the good might be sold, and over the possibility of not selling the good. Formally, Δ is defined by: $\Delta \equiv \{(q_1, q_2, \ldots, q_N) \mid 0 \leq q_i \leq 1 \text{ for all } i \in I \text{ and } \sum_{i \in I} q_i \leq 1\}$. Note that the probabilities $q_1, q_2, \ldots, q_N$ may add up to less than 1. In this case the remaining probability, $1 - \sum_{i \in I} q_i$, is the probability that the good is not sold.

Definition 3.1. A "direct mechanism" consists of functions q and t_i (for $i \in I$) where:

$$q : \Theta \to \Delta$$

and

$$t_i : \Theta \to \mathbb{R}$$

for $i \in I$.

The interpretation is that in a direct mechanism the buyers are asked to simultaneously and independently report their types. The function $q(\theta)$ describes the rule by which the good is allocated if the reported type vector is θ. We shall refer to q as the "allocation rule." The probability $q_i(\theta)$ is the probability that agent i obtains the good if the type vector is θ. The probability $1 - \sum_{i \in I} q_i(\theta)$ is the probability with which the seller retains the good if the type vector is θ. The functions t_i describe the transfer payment that buyer i makes to the seller. Note that we have assumed that this transfer payment is deterministic. This is without loss of generality.

We now state the "Revelation Principle" which as in Chapter 2 shows that in some sense there is no loss of generality in restricting attention to direct mechanisms.

Proposition 3.1 (Revelation Principle)**.** *For every mechanism Γ and Bayesian Nash equilibrium σ of Γ, there exists a direct mechanism Γ' and a Bayesian Nash equilibrium σ' of Γ' such that:*

(i) For every i and every θ_i, the strategy vector σ' satisfies:

$$\sigma'_i(\theta_i) = \theta_i,$$

that is, σ' prescribes telling the truth;

(ii) For every type vector θ, the distribution over outcomes that result under Γ if the agents play σ is the same as the distribution over outcomes that result under Γ' if the agents play σ', and the expected value of the transfer payments that result under Γ if the agents play σ is the same as the transfer payments that result under Γ' if the agents play σ'.

Proof. Construct Γ' by defining the functions q and t_i as required by item (ii) in Proposition 3.1. We can prove the result by showing that truth telling is a Bayesian equilibrium of the game. Suppose it were not. If type θ_i prefers to report that her type is θ'_i, then the same type θ_i prefers to deviate from σ, and to play the strategy that σ prescribes for θ'_i in Γ. Hence σ is not a Bayesian equilibrium of Γ. □

Proposition 3.1 shows that in the setup that we have described we can, without loss of generality, restrict our attention to the case in which the seller chooses a direct mechanism and proposes to agents that they report their types truthfully. Note, however, that it is crucial to this construction that we have neglected problems of multiple equilibria by assuming that agents follow the seller's proposal provided that it is an equilibrium, and provided that it gives them expected utility of at least zero. This is crucial because the equivalent direct mechanism that is constructed in the proof of Proposition 3.1 might have Bayesian Nash equilibria other than truth-telling, and there is no reason why these equilibria should be equivalent to any Bayesian Nash equilibrium of the indirect mechanism Γ. Depending on how equilibria are selected, one or the other mechanism might be strictly preferred by the seller in that case.

The revelation principle greatly simplifies our search for optimal mechanisms. We can restrict our attention to direct mechanisms in which it is a Bayesian equilibrium that everyone always reports their type truthfully, and in which every type's expected utility is at least zero. We want to define these properties of a direct mechanism formally. For this we introduce additional notation.

We denote by θ_{-i} the vector of all types except player i's type. We define $\Theta_{-i} \equiv \Theta^{N-1}$. We denote by F_{-i} the cumulative distribution of θ_{-i}, and we denote by f_{-i} the density of θ_{-i}. Given a direct mechanism, we define for each agent $i \in I$ a function $Q_i : [\underline{\theta}, \bar{\theta}] \to [0, 1]$ by setting:

$$Q_i(\theta_i) = \int\limits_{\Theta_{-i}} q_i(\theta_i, \theta_{-i}) f(\theta_{-i}) d\theta_{-i}. \tag{3.1}$$

Thus, $Q_i(\theta_i)$ is the conditional expected value of the probability that agent i obtains the good, conditioning on agent i's type being θ_i. We also define for each agent $i \in I$ a function $T_i : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$ by setting:

$$T_i(\theta_i) = \int\limits_{\Theta_{-i}} t_i(\theta_i, \theta_{-i}) f(\theta_{-i}) d\theta_{-i}. \tag{3.2}$$

Thus, $T_i(\theta_i)$ is the conditional expected value of the transfer that agent i makes to the seller, again conditioning on agent i's type being θ_i.[9] Finally, we also define agent i's expected utility $U_i(\theta_i)$ conditional on her type being θ_i. This is given by:

$$U_i(\theta_i) = \theta_i Q_i(\theta_i) - T_i(\theta_i).$$

Using this notation we can now formally define the two conditions that the seller has to respect when choosing a selling mechanism.

Definition 3.2. A direct mechanism is "incentive-compatible" if truth telling is a Bayesian Nash equilibrium; that is, if:

$$\theta_i Q_i(\theta_i) - T_i(\theta_i) \geq \theta_i Q_i(\theta_i') - T_i(\theta_i') \text{ for all } i \in I \text{ and } \theta_i, \theta_i' \in [\underline{\theta}, \bar{\theta}].$$

Definition 3.3. A direct mechanism is "individually rational" if each agent, conditional on her type, is voluntarily willing to participate, that is, if:

$$U_i(\theta_i) \geq 0 \text{ for all } i \in I \text{ and } \theta_i, \in [\underline{\theta}, \bar{\theta}].$$

To conclude this subsection, it is useful to introduce some further terminology. In the timeline of the game defined by a mechanism, the phase that follows after agents have learned their types, but before all agents' types are revealed, is often referred to as the "interim" phase. The phase before agents have learned their types is referred to as the "ex ante" phase, and

[9] We ignore questions of existence and uniqueness of the conditional expected values referred to in this paragraph.

the phase after all agents have revealed their types in a direct mechanism is then called the "ex post" phase. We shall use this terminology occasionally in these notes. For example, we shall refer to $T_i(\theta_i)$ as the interim expected transfer of agent i if she is of type θ_i, and we shall refer to $U_i(\theta_i)$ as the interim expected utility of agent i if she is of type θ_i.

3.2.3 Characterizing Incentive Compatibility and Individual Rationality

In this subsection we seek to understand better the structure of the set of all direct mechanisms that satisfy the two conditions introduced in Definitions 3.2 and 3.3. We proceed in much the same way as in Chapter 2.2 and therefore we omit most proofs. We first focus on incentive compatibility.

Lemma 3.1. *If a direct mechanism is incentive-compatible, then for every agent $i \in I$ the function Q_i is increasing.*

The proof of this is the same as the proof of Lemma 2.1, with the functions Q_i and T_i replacing the functions q and t. Similarly, we obtain analogues to Lemmas 2.2, 2.3 and 2.4:

Lemma 3.2. *If a direct mechanism is incentive-compatible, then for every agent $i \in I$ the function U_i is increasing. It is also convex, and hence differentiable except in at most countably many points. For all θ_i for which it is differentiable, it satisfies:*

$$U_i'(\theta_i) = Q_i(\theta_i).$$

Lemma 3.3 (Payoff Equivalence)**.** *Consider an incentive-compatible direct mechanism. Then for all $i \in I$ and all $\theta_i \in [\underline{\theta}, \bar{\theta}]$:*

$$U_i(\theta_i) = U_i(\underline{\theta}) + \int_{\underline{\theta}}^{\theta_i} Q_i(x)dx.$$

Lemma 3.4 (Revenue Equivalence)**.** *Consider an incentive-compatible direct mechanism. Then for all $i \in I$ and all $\theta_i \in [\underline{\theta}, \bar{\theta}]$:*

$$T_i(\theta_i) = T_i(\underline{\theta}) + (\theta_i Q_i(\theta_i) - \underline{\theta}_i Q_i(\underline{\theta}_i)) - \int_{\underline{\theta}}^{\theta_i} Q_i(x)dx.$$

Lemma 3.3 shows that the interim expected payoffs of the different buyer types are pinned down by the functions Q_i and the expected payoff of the lowest type. Lemma 3.4 shows similarly that the interim expected payments of the different buyer types are pinned down by the functions Q_i and the expected payment of the lowest type. Note that this does not mean that the ex post payment functions t_i are uniquely determined. Different functions t_i might give rise to the same interim expected payments T_i.

Consider two different indirect mechanisms, and Bayesian Nash equilibria of these mechanisms, such that they imply the same interim expected probability of obtaining the object for each type of each agent, and such that the expected payment made by the lowest type is the same in the two mechanisms. Then Lemma 3.4 implies that all types' interim expected payments are the same for these two indirect mechanisms, and therefore, of course, also the expected revenue of the seller is the same for these two mechanisms. It is for this reason that the result is called the "revenue equivalence theorem."

We wish to explain an application of the revenue equivalence theorem. Consider the symmetric case in which F_i does not depend on i. Suppose we wanted to compare auctioneer's expected revenue from the second price auction with minimum bid 0 to the expected revenue from the first price auction with minimum bid 0. In the second price auction it is a weakly dominant strategy, and hence a Bayesian Nash equilibrium, to bid one's true value. A symmetric Bayesian Nash equilibrium for the first price auction is constructed in Proposition 2.2 of Krishna (2002). This equilibrium is in strictly increasing strategies. Hence this equilibrium shares with the equilibrium of the second price auction that the expected payment of the lowest type is zero (because this type's probability of winning is zero), and that the highest type wins with probability 1. Therefore, the equilibria imply the same values for $T_i(\underline{\theta})$ and $Q_i(\theta_i)$ for all $i \in I$ and $\theta_i \in [\underline{\theta}, \bar{\theta}]$. The revenue equivalence theorem implies that the expected revenue from the equilibria of the two different auction formats is the same.

We described in the previous paragraph the most famous application of Lemma 3.4. But note that the Lemma is much more general. In shorthand expression, the Lemma says that the interim expected payments of all types only depend on the interim expected allocation rule and the interim expected payment of the lowest type.

As in Chapter 2.2 we can collect the observations made so far, and obtain conditions that are not only necessary but also sufficient for incentive compatibility. The proof is analogous to the proof of Proposition 2.2 and is therefore omitted.

Proposition 3.2. *A direct mechanism* $(q, t_1, t_2, \ldots, t_N)$ *is incentive-compatible if and only if for every* $i \in I$*:*

(i) Q_i *is increasing;*

(ii) For every $\theta_i \in [\underline{\theta}, \bar{\theta}]$*:*

$$T_i(\theta_i) = T_i(\underline{\theta}) + (\theta_i Q_i(\theta_i) - \underline{\theta} Q_i(\underline{\theta})) - \int_{\underline{\theta}}^{\theta_i} Q_i(x)dx$$

We have now obtained a complete understanding of the implications of incentive compatibility for the seller's choice. The seller can focus on two choice variables: firstly the allocation rule q, and secondly the interim expected payment by a buyer with the lowest type: $T_i(\underline{\theta})$. As long as the seller picks an allocation rule q such that the functions Q_i $(i \in I)$ are increasing, he can pick the interim expected payments by the lowest types in any arbitrary way, and be assured that there will be some transfer scheme that makes the allocation rule incentive-compatible and that implies the given interim expected payments by the lowest types. Moreover, any such transfer scheme will give him the same expected revenue, and therefore the seller does not have to worry about the details of this transfer scheme.

So far we have focused on the characterization of incentive compatibility. Now we turn to individual rationality. However, we restrict our attention to incentive-compatible direct mechanisms. Then we have the following result that is analogous to Lemma 2.3.

Proposition 3.3. *An incentive-compatible direct mechanism is individually rational if and only if for every* $i \in I$ *we have:* $T_i(\underline{\theta}_i) \leq \underline{\theta}_i Q_i(\underline{\theta}_i)$.

Thus, we have one further constraint on the seller's choice of direct mechanism. The seller has to choose a mechanism that implies an expected utility of at least zero for the lowest type agents.

3.2.4 Expected Revenue Maximization

We now study the expected revenue maximizing choice of selling mechanism. We begin with a simple observation that is analogous to Lemma 2.5.

Lemma 3.5. *If an incentive-compatible and individually direct mechanism maximizes the seller's expected revenue, then for every* $i \in I$*:*

$$T_i(\underline{\theta}) = \underline{\theta} Q_i(\underline{\theta}).$$

We can now simplify the seller's problem further. The seller has to choose a function q so that the interim probabilities Q_i are increasing for all $i \in I$. The payments are then completely determined by part (ii) of Proposition 3.2 and Lemma 3.5. Substituting the formula in Lemma 3.5 into part (ii) of Proposition 3.2 we get for every $i \in I$ and $\theta_i \in [\underline{\theta}, \bar{\theta}]$:

$$T_i(\theta_i) = \theta_i Q_i(\theta_i) - \int_{\underline{\theta}}^{\theta_i} Q_i(x)dx \tag{3.3}$$

Note that for the seller's expected revenue the details of the function q don't matter, only the interim probabilities Q_i.

We shall now focus on the optimal choice of q. We shall proceed as in Chapter 2.3 and not as in Chapter 2.2. The reason is that there are many extreme points of the seller's choice set. These were the focus of Chapter 2.2, where it was sufficient to characterize these extreme points, but in our context a characterization of these extreme points wouldn't take us very far. By doing the same calculations that led in Chapter 2.3 to equation (2.25), we can calculate the seller's expected revenue from any particular buyer i.

$$\int_{\underline{\theta}}^{\bar{\theta}} Q_i(\theta_i)\left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}\right) f_i(\theta_i)d\theta_i. \tag{3.4}$$

To obtain a formula for the total expected transfer by all agents we add the formula in equation (3.4) over all $i \in I$. We obtain:

$$\begin{aligned}&\sum_{i\in I}\left[\int_{\underline{\theta}}^{\bar{\theta}} Q_i(\theta_i)\left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}\right) f_i(\theta_i)d\theta_i\right]\\ =&\sum_{i\in I}\left[\int_{\Theta} q_i(\theta)\left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}\right) f(\theta)d\theta\right]\end{aligned} \tag{3.5}$$

where the last equality becomes obvious if one recalls the definition of $Q_i(\theta_i)$.

As in Chapter 2.3, we first ask which function q the seller would choose if he did not have to make sure that the functions Q_i are increasing. In a second step, we introduce an assumption that makes sure that the optimal q from the first step implies increasing functions Q_i. If monotonicity could be ignored, then the seller would choose for each θ the probabilities $q_i(\theta)$ so as to maximize the expression in the large round brackets in the formula for expected revenue. We define this expression to be $\psi_i(\theta_i)$:

$$\psi_i(\theta_i) \equiv \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \text{ for all } i \in I \text{ and } \theta_i \in [\underline{\theta}, \bar{\theta}]. \tag{3.6}$$

The optimal allocation rule without monotonicity is then:

$$q_i(\theta) = \begin{cases} 1 & \text{if } \psi_i(\theta_i) > 0 \text{ and } \psi_i(\theta_i) > \psi_j(\theta_j) \text{ for all } j \in I \text{ with } j \neq i; \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{for all } i \in I \text{ and } \theta \in \Theta. \tag{3.7}$$

Note that we have ignored the case that $\psi_i(\theta_i) = \psi_j(\theta_j)$ for some $j \neq i$. This is a zero probability event, and it does not affect either the buyer's incentives or the seller's revenue.

We now introduce an assumption under which this allocation rule satisfies the monotonicity constraint of the seller's maximization problem. The assumption is that for all agents $i \in I$ the distribution functions F_i are "regular" in the same sense as in Assumption 2.1.

Assumption 3.1. *For every $i \in I$, the function $\psi_i(\theta_i)$ is strictly increasing.*

For the allocation rule q described above the probability $Q_i(\theta_i)$ is the probability that $\psi_i(\theta_i)$ is larger than zero and larger than $\psi_i(\theta_j)$ for every $j \in I$ with $j \neq i$. Clearly, if ψ_i is increasing, as required by the regularity assumption, this probability is an increasing function of θ_i. Thus, Q_i is indeed increasing. We have arrived at the following result:

Proposition 3.4 (Myerson (1981))**.** *Suppose that for every agent $i \in I$ the cumulative distribution function F_i is regular. Among all incentive-compatible and individually rational direct mechanisms, those mechanisms maximize the seller's expected revenue that satisfy for all $i \in I$ and all $\theta \in \Theta$:*

(i)

$$q_i(\theta) = \begin{cases} 1 & \textit{if } \psi_i(\theta_i) > 0 \textit{ and } \psi_i(\theta_i) > \psi_j(\theta_j) \textit{ for all } i \in I \textit{ with } j \neq i; \\ 0 & \textit{otherwise.} \end{cases}$$

(ii)

$$T_i(\theta_i) = \theta_i Q_i(\theta_i) - \int_{\underline{\theta}}^{\theta_i} Q_i(x)dx.$$

We have characterized the optimal choice of the allocation rule q and of the interim expected payments. We have not described the actual transfer schemes that make these choices incentive-compatible and individually rational, although we know that such transfers can be found.

The expression $\psi_i(\theta_i)$ is sometimes referred to as seller i's "virtual type." Using this expression we can rephrase the result in Proposition 3.4 as follows: The expected revenue maximizing auction allocates the object to the buyer with the highest virtual type, provided that this type is at least zero.

If the buyers are symmetric, i.e. the distribution functions F_i are all the same, the optimal mechanism prescribes that the object is given to the buyer with the highest value, if it is sold at all. This is because then ψ_i is the same for all i, and hence $\psi_i(\theta_i) > \psi_j(\theta_j) \Leftrightarrow \theta_i > \theta_j$. Note that in the case with asymmetric buyers, the optimal mechanism may sometimes give the good to a buyer who does not have the highest value.

In the symmetric case, the optimal direct mechanism can be implemented using either a first or a second price auction with minimum bid $\psi_i^{-1}(0)$, where ψ_i^{-1} is the inverse of any one of the functions ψ_i. Thus, in the symmetric case, familiar auction processes, with appropriately chosen minimum bids, are optimal. To show this, one has to derive equilibrium bidding functions for these auctions, and verify that they imply the allocation rule and the transfer payments indicated in Proposition 3.4. An excellent reference on this is Chapter 2 of Krishna (2002).

3.2.5 Maximizing Welfare

Suppose that the seller were not maximizing expected profits but expected welfare. Let us assume that the seller uses the following utilitarian welfare definition:

$$\sum_{i \in I} q_i(\theta)\theta_i.$$

Note that this seller is no longer concerned with transfer payments. Expected welfare depends only on the allocation rule q.

We can easily analyze this seller's problem using the framework described before. The seller can choose any rule q that is such that the functions Q_i are monotonically increasing. He can choose any transfer payments such that $T_i(\underline{\theta}_i) \leq \underline{\theta}_i Q_i(\underline{\theta}_i)$ for all $i \in I$.

Which rule q should the seller choose? If types were known, maximization of the welfare function would require that the object be allocated to the potential buyer for whom θ_i is largest. Note that welfare if the object is not transferred to one of the potential buyers is assumed to be zero, and hence the welfare maximizing seller always wants to transfer the object. Can the welfare maximizing seller allocate the object to the buyer with the highest value even if he doesn't know the valuations? We know that this is possible if the implied functions Q_i are increasing. This is obviously the case. Therefore we conclude:

Proposition 3.5. *Among all incentive-compatible, individually rational direct mechanisms, a mechanism maximizes welfare if and only if for all $i \in I$ and all $\theta \in \Theta$:*

(i)

$$q_i(\theta) = \begin{cases} 1 & \textit{if } \theta_i > \theta_j \textit{ for all } j \in I \textit{ with } j \neq i; \\ \\ 0 & \textit{otherwise.} \end{cases}$$

(ii)

$$T_i(\theta_i) \leq \theta_i Q_i(\theta_i) - \int_{\underline{\theta}}^{\theta_i} Q_i(x)dx.$$

Note that this result does not rely on Assumption 3.1. In comparing welfare maximizing and revenue maximizing mechanisms in the case that Assumption 3.1 holds, we observe that there are two differences. The first is that revenue maximizing mechanism allocates the object to the highest virtual type whereas the welfare maximizing mechanism allocates the object to the highest actual type. In the symmetric case, the functions ψ_i are the same for all $i \in I$ and there is no difference between these two rules. But in the asymmetric case the revenue maximizing mechanism might allocate the object inefficiently. A second difference is that the revenue maximizing mechanism sometimes does not sell the object at all, whereas the welfare maximizing mechanism always sells the object. This is an instance of the well-known inefficiency that monopoly sellers make goods artificially scarce.

Proposition 3.5, like Proposition 3.4, does not describe the transfer scheme associated with a welfare maximizing mechanism in detail. An example of a direct mechanism that is incentive-compatible and individually rational and that maximizes welfare is the second price auction with reserve price zero.

3.2.6 Numerical Examples

We give one symmetric and one asymmetric example.

Example 3.1. *Suppose that* $\underline{\theta} = 0$, $\bar{\theta} = 1$, *and that* θ_i *is uniformly distributed so that* $F(\theta_i) = \theta_i$. *We begin by calculating for* $i = 1, 2$:

$$\begin{aligned} \psi_i(\theta_i) &= \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \\ &= \theta_i - \frac{1 - \theta_i}{1} \\ &= 2\theta_i - 1. \end{aligned} \tag{3.8}$$

Note that the regularity assumption 3.1 is satisfied.

In the expected revenue maximizing auction, the good is sold to neither bidder if:

$$\begin{aligned} \psi_i(\theta_i) < 0 &\Leftrightarrow \\ 2\theta_i - 1 < 0 &\Leftrightarrow \\ \theta_i < \frac{1}{2} \end{aligned} \tag{3.9}$$

holds for $i = 1$ *and* $i = 2$. *If the good is sold, it is sold to bidder 1 if:*

$$\begin{aligned} \psi_1(\theta_1) > \psi_2(\theta_2) &\Leftrightarrow \\ 2\theta_1 - 1 > 2\theta_2 - 1 &\Leftrightarrow \\ \theta_1 > \theta_2. \end{aligned} \tag{3.10}$$

The expected revenue maximizing auction will allocate the object to the buyer with the highest type provided that this type is larger than 0.5. *A first or second price auction with reserve bid* $\frac{1}{2}$ *will implement this mechanism. A first or second price auction with reserve bid 0 maximizes expected welfare.*

Example 3.2. *Now suppose that* $N = 2$, $\underline{\theta} = 0$, $\bar{\theta} = 1$, *and that* $F_1(\theta_1) = (\theta_1)^2$ *whereas* $F_2(\theta_2) = 2\theta_2 - (\theta_2)^2$. *Thus, player 1 is more likely to have high values than player 2.*

We begin by calculating

$$\begin{aligned} \psi_1(\theta_1) &= \theta_1 - \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} \\ &= \theta_1 - \frac{1 - (\theta_1)^2}{2\theta_1} \\ &= \frac{3}{2}\theta_1 - \frac{1}{2\theta_1} \end{aligned} \tag{3.11}$$

and

$$\begin{aligned}\psi_2(\theta_2) &= \theta_2 - \frac{1 - F_2(\theta_2)}{f_2(\theta_i)} \\ &= \theta_2 - \frac{1 - 2\theta_2 + (\theta_2)^2}{2 - 2\theta_2} \\ &= \theta_2 - \frac{1 - \theta_2}{2} \\ &= \frac{3}{2}\theta_2 - \frac{1}{2} \qquad (3.12)\end{aligned}$$

Again the regularity assumption is satisfied.

In an expected revenue maximizing auction, the good is sold to neither bidder if:

$$\begin{aligned}\psi_1(\theta_1) < 0 &\Leftrightarrow \\ \frac{3}{2}\theta_1 - \frac{1}{2\theta_1} < 0 &\Leftrightarrow \\ \theta_1 < \sqrt{\frac{1}{3}} & \qquad (3.13)\end{aligned}$$

and

$$\begin{aligned}\psi_2(\theta_2) < 0 &\Leftrightarrow \\ \frac{3}{2}\theta_2 - \frac{1}{2} < 0 &\Leftrightarrow \\ \theta_2 < \frac{1}{3} & \qquad (3.14)\end{aligned}$$

If the good is sold, it is sold to bidder 1 if:

$$\begin{aligned}\psi_1(\theta_1) > \psi_2(\theta_2) &\Leftrightarrow \\ \frac{3}{2}\theta_1 - \frac{1}{2\theta_1} > \frac{3}{2}\theta_2 - \frac{1}{2} &\Leftrightarrow \\ \theta_2 < \theta_1 - \frac{1}{3\theta_1} + \frac{1}{3} & \qquad (3.15)\end{aligned}$$

Figure 3.1 shows the optimal allocation of the good. The 45°-line is shown as a dashed line. Note that the mechanism is biased against player 1. If the good is sold, bidder 1 wins the object only in a subset of all cases where his value is higher than bidder 2's value. In the expected welfare maximizing mechanism the object is allocated to player 1 if and only if his value is higher than player 2's value. A second price auction will maximize expected welfare, although a first price auction will not necessarily.

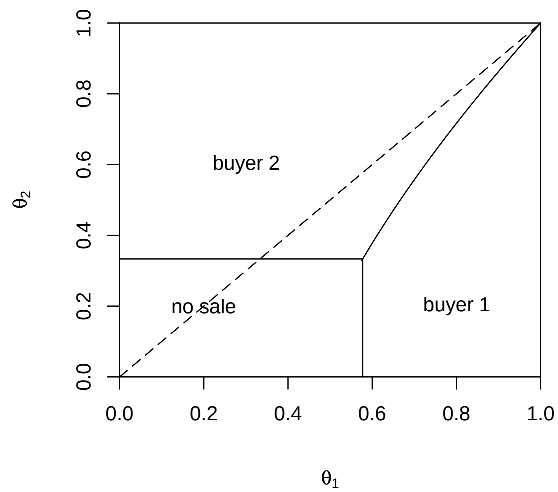

Figure 3.1: Expected revenue maximizing allocation in Example 3.2

3.3 Public Goods

3.3.1 Set-Up

Our next example is a public goods problem. The theory of mechanism design began with the theory of mechanisms for the provision of public goods. This is a central application of the theory of mechanism design. Methodologically, the example that we discuss in this section illustrates the design of optimal mechanisms subject to additional constraints beyond incentive compatibility and individual rationality constraints. The specific constraint on which we shall focus here is the government budget constraint.

We consider a community consisting of N agents: $I = \{1, 2, \ldots, N\}$ where $N \geq 2$. They have to choose whether to produce some indivisible, non-excludable public good. We denote this decision by $g \in \{0, 1\}$. If the public good is produced, then $g = 1$. If it is not produced, then $g = 0$.

Agent i's utility if the collective decision is g and if she pays a transfer t_i to the community is: $\theta_i g - t_i$. Here, θ_i is a random variable that follows a continuous distribution function F_i with density f_i. We shall refer to θ_i as agent i's type, or as agent i's valuation of the public good. The support of θ_i is $[\underline{\theta}, \bar{\theta}]$ where $0 \leq \underline{\theta} < \bar{\theta}$. We assume that $f_i(\theta_i) > 0$ for all $\theta_i \in [\underline{\theta}, \bar{\theta}]$.

We assume that for $i, j \in I$ with $i \neq j$, the random variables θ_i and θ_j are independent. We also assume that each agent i observes θ_i, but not the other agent' types θ_j where $j \neq i$. We denote by θ the vector $(\theta_1, \theta_2, \ldots, \theta_N)$. The support of the random variable θ is $\Theta = [\underline{\theta}, \bar{\theta}]^N$. The cumulative distribution function of θ will be denoted by F, and its density by f. The distribution F is common knowledge among the agents. We are thus considering an independent private values model of public goods.

The fact that the public good is non-excludable is reflected by the fact that the same variable g enters into each individual's utility function. An alternative model, that is also of interest, is a model in which individuals can be selectively excluded from consuming the public good.

The cost of producing the public good is assumed to be $c > 0$, so that a collective decision g implies cost cg. We shall consider this society from the perspective of a benevolent mechanism designer who does not observe θ, but who knows F. We attribute to the mechanism designer a utilitarian welfare function with equal welfare weights for all agents. Welfare is thus:

$$\left(\sum_{i \in I} \theta_i\right) g - \sum_{i \in I} t_i. \tag{3.16}$$

3.3.2 Incentive-Compatible and Individually Rational Direct Mechanisms

As in previous parts of these notes, we can restrict our attention, without loss of generality, to incentive-compatible direct mechanisms where agents' payments are not random. To simplify our treatment of the budget constraint we also restrict our attention to mechanisms where the decision about the public good is non-stochastic. However, stochastic mechanisms would not present conceptual problems, and the results of this section hold even if stochastic mechanisms are considered.

Definition 3.4. A "direct mechanism" consists of functions q and t_i (for $i \in I$) where:

$$q : \Theta \rightarrow \{0, 1\}$$

and

$$t_i : \Theta \rightarrow \mathbb{R}.$$

The function q assigns to each type vector θ the collective decision about the public good that is produced if the agents' types are θ. We shall refer to q as the "decision rule." For each agent i, the function t_i describes for every type vector θ the transfer that agent i makes when the types are θ.

Given a direct mechanism, we define for each agent $i \in I$ functions $Q_i : [\underline{\theta}, \bar{\theta}] \to [0,1]$ and $T_i : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$ where $Q_i(\theta_i)$ is the interim conditional probability that the public good is produced, where we condition on agent i's type being θ_i, and $T_i(\theta_i)$ is the interim conditional expected value of the transfer that agent i makes to the community, again conditioning on agent i's type being θ_i. Finally, we also define agent i's expected utility $U_i(\theta_i)$ conditional on her type being θ_i. This is given by: $U_i(\theta_i) = Q_i(\theta_i)\theta_i - T_i(\theta_i)$.

As before, we shall restrict our attention to mechanisms that are incentive-compatible and individually rational. As the notation that we have introduced in the previous paragraph parallels that of Section 3.2 we can refer to Definitions 3.2 and 3.3 for definitions and to Propositions 3.2 and 3.3 for characterizations of incentive compatibility and individual rationality.

3.3.3 Ex ante and Ex post Budget Balance

We now introduce the government budget constraint. This constraint requires that the money raised by the mechanism is at least enough to cover the costs of producing the public good. A restrictive version of the constraint requires budget balance for each realization of agents' types.

Definition 3.5. A direct mechanism is "ex post budget balanced" if for every $\theta \in [\underline{\theta}, \bar{\theta}]^N$ we have $\sum_{i \in I} t_i(\theta) \geq cq(\theta)$.

An alternative formulation requires budget balance to hold only on average, across realization of agents' types.

Definition 3.6. A direct mechanism is "ex ante budget balanced" if

$$\int_{\Theta} \sum_{i \in I} t_i(\theta) f(\theta) d\theta \geq \int_{\Theta} cq(\theta) f(\theta) d\theta.$$

Clearly, ex post budget balance implies ex ante budget balance. Ex post budget balance appears to be more restrictive. We shall show that in our context this appearance is misleading. For every ex ante budget balanced mechanism there is an *equivalent* ex post budget balanced mechanism. Here, we define "equivalent" as follows:

Definition 3.7. Two direct mechanisms are "equivalent" if for all agents $i \in I$ and for all types $\theta_i, \theta_i' \in [\underline{\theta}, \bar{\theta}]$, agent i's expected transfers, conditional on agent i's type being θ_i and agent i reporting to be type θ_i', is the same in the two mechanisms.

Notice that if two mechanisms are equivalent, and if one of them is incentive-compatible, then the same is true for the other, and if one of them is individually rational, then the same is true for the other.

Proposition 3.6. *For every direct mechanism with decision rule q that is ex ante budget balanced, there is an equivalent direct mechanism with the same decision rule q that is ex post budget balanced.*

Proof. Suppose first that we have a mechanism for which the ex ante budget balance condition holds with equality. We will consider the case of an ex ante budget surplus at the end of this proof. We show how this mechanism can be made ex post budget balanced by modifying the payment schemes of two agents. We denote the payments in the ex ante budget balanced mechanism by t_i. We denote by $T_j(\theta_i)$ the expected value of agent j's transfer, conditioning on agent i's type being θ_i. So far, we have employed this notation only in the case that $j = i$. Now we use it also in the case that $j \neq i$.

We construct the payments in the ex post budget balanced scheme by modifying the payments in the ex ante budget balanced scheme as follows. One arbitrarily selected agent, say agent 1, provides the primary coverage of the deficit. However, she does not cover that part of the ex post deficit that is predicted by her own type. In other words, we add to her original payment the ex post deficit minus the expected value of the deficit conditional on her own type. Formally, if the vector of types is θ, then agent 1's payment is in the modified mechanism:

$$t_1(\theta) + \left(cq(\theta) - \sum_{i \in I} t_i(\theta) \right) - \left(cQ_1(\theta_1) - \sum_{i \in I} T_i(\theta_1) \right). \tag{3.17}$$

We now check that agent 1's expected payoff, conditional on her type being θ_1 and her reporting that her type is θ_1', is unchanged. The expression that we have added to agent 1's payment is a random variable with expected value zero, independent of whether agent 1 reports her type truthfully or not. This is because it would have expected value zero if agent 1's true type were θ_1', and, moreover, the conditional distribution of this random variable, by the independence of private types, is the same if agent 1's true type is θ_1 rather than θ_1'.

In addition, some other arbitrarily selected agent, say agent 2, pays for the expected value of the deficit conditional on agent 1's type. Her modified payment is:

$$t_2(\theta) + cQ_1(\theta_1) - \sum_{i \in I} T_i(\theta_1). \tag{3.18}$$

Note that the random variable that we are adding to agent 2's payment is independent of agent 2's report. Moreover, it has ex ante expected value zero because the mechanism is ex ante budget balanced. Finally, because agent 2's type does not provide any information about agent 1's type, the expectation of the added term conditional on agent 2's type is the same as its ex ante expectation. Therefore, agent 2's expected payoff, if her true type is θ_2, and she reports that her type is θ_2', is unchanged.

Finally, all agents $i \neq 1, 2$ pay the same as before: $t_i(\theta)$. Adding up all agents' payments shows that the sum of the payments equals the costs $cq(\theta)$ in each state θ, and that therefore this mechanism is ex post budget balanced. This concludes the proof for the case of ex ante budget equality.

If there is an ex ante budget surplus, we subtract from some agent's payments a constant until the mechanism is exactly budget balanced. Then we conduct the transformation described above. Then we add the constant again to this agent's payments. The mechanism that we obtain has the required properties. $\square$

We shall assume that the mechanism designer in this section considers only ex post budget balanced mechanisms. The above result makes clear that this is equivalent to requiring ex ante budget balance. We shall work with either condition, whichever is more convenient.

We calculate the ex ante expected revenue from an incentive-compatible mechanism in the same way as we calculated above the expected revenue from an incentive-compatible auction. This yields the following formula, where we find it convenient to write the initial term as interim expected utility of the lowest type:

$$\sum_{i \in I} -U_i(\underline{\theta}) + \int_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta) d\theta. \tag{3.19}$$

3.3.4 Welfare Maximization

Which mechanism $(q, t_1, t_2, \ldots, t_N)$ would the designer choose if she were not constrained by incentive compatibility and individual rationality, but had to satisfy ex post budget balance? As the mechanism designer subtracts transfer payments in her welfare function (3.59) she would never raise transfers larger than what is required to cover cost. From (3.59) it is then clear that the optimal decision rule is:[10]

$$q^*(\theta) = \begin{cases} 1 & \text{if } \sum_{i \in I} \theta_i \geq c \\ \\ 0 & \text{otherwise.} \end{cases} \tag{3.20}$$

As welfare function (3.59) indicates, the utilitarian designer with equal welfare weights for all agents does not care how costs are distributed among agents. Therefore, any transfer rules that satisfy:

$$\sum_{i \in I} t_i^*(\theta) = \begin{cases} c & \text{if } q(\theta) = 1 \\ \\ 0 & \text{otherwise.} \end{cases} \tag{3.21}$$

are optimal. We call these direct mechanisms "first best."

The following impossibility result shows that in all non-trivial cases no first best mechanism is incentive-compatible and individually rational.

Proposition 3.7. *An incentive-compatible and individually rational first best mechanism exists if and only if either $N\underline{\theta} \geq c$ or $N\bar{\theta} \leq c$.*

The condition $N\underline{\theta} \geq c$ means that even if all agents have the lowest valuation the sum of the valuations is at least as high as the cost of producing the public good. Thus, for all type vectors it is efficient to produce the public good. Analogously, $N\bar{\theta} \leq c$ means that for all type vectors it is efficient no to produce the public good. These are trivial cases. For all non-trivial cases, Proposition 3.7 is an impossibility result.

Proof. If $N\underline{\theta} \geq c$ then a mechanism where the public good is always produced and all agents pay $\frac{c}{N}$ is first best and incentive-compatible and individually rational. If $N\bar{\theta} \leq c$ then a mechanism where the public good is never produced and no agent ever pays anything is first best and incentive-compatible and individually rational.

[10] In this definition, we require arbitrarily that $q^*(\theta) = 1$ if $\sum_{i \in I} \theta_i = c$. This simplifies the exposition, and could be changed without conceptual difficulty.

It remains to prove the converse. Thus we consider the case $N\underline{\theta} < c < N\bar{\theta}$, and wish to prove that there is no incentive-compatible and individually rational first best mechanism. To prove this we display a direct mechanism that makes the first best decision rule q^* incentive-compatible and individually rational. We then argue that this mechanism maximizes expected transfer payments among all incentive-compatible and individually rational first best mechanisms. Finally, we show that it yields an expected budget deficit in all non-trivial cases. The assertion then follows.

Definition 3.8. The "pivot mechanism" is the mechanism that is given by the first best decision rule q^* and by the following transfer scheme:

$$t_i(\theta) = \underline{\theta} q^*(\underline{\theta}, \theta_{-i}) + (q^*(\theta) - q^*(\underline{\theta}, \theta_{-i})) \left(c - \sum_{j \neq i} \theta_j \right)$$

for all $i \in I$ and $\theta \in [\underline{\theta}, \bar{\theta}]^N$.

To see why this mechanism is called "pivot" mechanism it is useful to ignore the first term in the sum on the right hand side of the formula for $t_i(\theta)$. This term does not depend on agent i's report θ_i. The second term equals the change to the social welfare of all other agents caused by agent i's report, and hence agent i pays only if her report is "pivotal" for the collective decision. Here we compare the actual outcome to the outcome that would have occurred had agent i reported the lowest type $\underline{\theta}$. Agent i's report changes the collective decision if $q^*(\theta) - q^*(\underline{\theta}, \theta_{-i}) = 1$. In that case agent i pays for the difference between the costs of the project, and the sum of all other agents' valuations of the project.

Lemma 3.6. *The pivot mechanism is incentive-compatible and individually rational.*

Proof. Consider an agent $i \in I$ who is of type θ_i and who contemplates reporting that she is of type $\theta_i' \neq \theta_i$. Fix the other agents' types as θ_{-i}. We are going to show that truthful reporting is optimal *whatever the other agents' types* θ_{-i} are. This obviously implies that truthful reporting is a Bayesian equilibrium. If we leave out terms that do not depend on agent i's report, then agent i's utility if reporting θ_i' is:

$$\theta_i q^*(\theta_i', \theta_{-i}) - q^*(\theta_i', \theta_{-i}) \left(c - \sum_{j \neq i} \theta_j \right)$$

$$= q^*(\theta_i', \theta_{-i}) \left(\sum_{j=1}^{N} \theta_j - c \right). \tag{3.22}$$

Thus agent i's utility is true social welfare if the collective decision is $q^*(\theta_i', \theta_{-i})$. Because q^* is first best, agent i's utility is maximized if she reports truthfully $\theta_i' = \theta_i$. This proves incentive compatibility. To verify individual rationality note that the expected utility of agent i obviously equals zero if her type is $\underline{\theta}$. By a result analogous to Lemma 3.2 this implies that all types' interim expected utility is at least zero, and the mechanism is individually rational. □

The proof reveals two important features of the formula in Definition 3.9. The first is that those parts of agent i's transfer payment that depend on agent i's report are chosen so that agent i's incentives are exactly aligned with social welfare. The second feature is that those parts of agent i's transfer payment that do not depend on agent i's report are chosen so as to equalize agent i's utility if she is of the lowest type with her reservation utility of zero. This ensure that individual rationality is satisfied for all agents.

The "pivot mechanism" is a special "Vickrey-Clarke-Groves" (VCG) mechanism. In general, a mechanism is a VCG mechanism if every agent's payment consists of two terms. Firstly, a part that depends on the agent's report, and that has the effect of aligning the agent's incentives with social welfare. Secondly, a part that does not depend on the agent's report. In a VCG mechanism in general, this second part can be arbitrary. In a pivot mechanism, the second part is chosen so as to ensure individual rationality as an equality constraint for the type whose individual rationality constraint is "most restrictive."[11] The next result indicates the special importance of the pivot mechanism.

Lemma 3.7. *No incentive-compatible and individually rational mechanism that implements the first best decision rule q^* has larger expected surplus than the pivot mechanism.*

Proof. By Corollary 3.19 the expected budget surplus of an incentive-compatible direct mechanism that implements the first best decision rule q^* equals the interim expected payments of the lowest types plus a term that is the same

[11] A type's individual rationality constraint is "most restrictive" if that type's individual rationality constraint implies all other type's individual rationality constraints. See Section 4 for an example where this type is not the lowest type in types' numerical order.

for all such rules. If a mechanism is individually rational then the interim expected payments of the lowest types can be at most such that the expected utility of the lowest types are zero. For the pivot mechanism the expected utilities of the lowest types are exactly equal to zero. Therefore, no incentive-compatible, individually rational direct mechanism can have higher expected surplus than the pivot mechanism. □

We conclude the proof by showing that the pivot mechanism has an ex ante expected deficit except in trivial cases.

Lemma 3.8. *If $N\underline{\theta} < c < N\bar{\theta}$, then the ex ante expected surplus of the pivot mechanism is negative.*

Proof. We show that the ex post surplus of the pivot mechanism is always non-positive and with positive probability negative. This implies that the ex ante expected surplus is negative. Consider first θ such that $q^*(\theta) = 0$. In this case, there are no costs and no agent pays any transfer. Hence the deficit is zero. Consider next θ such that $q^*(\theta) = 1$ and $q^*(\underline{\theta}, \theta_{-i}) = 1$ for every $i \in I$. In this case each agent pays $\underline{\theta}$. By assumption the probability of the production of the good according to the first best decision rule is not 1, and hence $N\underline{\theta} < c$. Therefore, total payments are less than c and there is a deficit.

Consider finally states θ such that $q^*(\theta) = 1$, and $q^*(\underline{\theta}, \theta_{-i}) = 0$ for some $i \in I$. Let P be the set of all $i \in I$ for which this holds, and call these agents "pivotal." Define $NP = I \setminus P$. Abusing notation slightly denote by P the number of elements of P, and by NP the number of elements of NP. The total transfers are:

$$\begin{aligned}
&\sum_{\imath\in P}\left(c-\sum_{j\neq i}\theta_j\right)+\sum_{i\in NP}\underline{\theta}\\
=&Pc-P\sum_{j\in NP}\theta_j-(P-1)\sum_{j\in P}\theta_j+\sum_{i\in NP}\underline{\theta}\\
=&Pc-(P-1)\sum_{j\in NP}\theta_j-(P-1)\sum_{j\in P}\theta_j-\sum_{i\in NP}(\theta_i-\underline{\theta})\\
=&Pc-(P-1)\sum_{j\in I}\theta_j-\sum_{i\in NP}(\theta_i-\underline{\theta}). \qquad (3.23)
\end{aligned}$$

We shall show that this is no more than c:

$$Pc-(P-1)\sum_{j\in I}\theta_j-\sum_{i\in NP}(\theta_i-\underline{\theta})\leq c \Leftrightarrow$$

$$
\begin{aligned}
(P-1)c \le (P-1)\sum_{j\in I}\theta_j + \sum_{i\in NP}(\theta_i - \underline{\theta}) &\Leftarrow \\
(P-1)c \le (P-1)\sum_{j\in I}\theta_j &\Leftrightarrow \\
c \le \sum_{j\in I}\theta_j &
\end{aligned} \tag{3.24}
$$

which is true by construction in states in which $q^*(\theta) = 1$.

States in which the public good is produced and some agents are pivotal occur with positive probability under our assumption. Moreover, conditional on such a state occurring, with probability 1 we have that $\theta_i > \underline{\theta}$ for all $i \in NP$. In this case, the above calculation shows that the surplus is strictly negative, and hence there is an expected deficit. □

This concludes the proof of Proposition 3.7. □

In the remainder of this subsection we focus on the case in which $N\underline{\theta} < c < N\bar{\theta}$, and thus it is impossible by Proposition 3.7 to implement q^* using a mechanism that is incentive-compatible, individually rational, and ex ante budget balanced. Our objective is to determine direct mechanisms that maximize expected welfare among all incentive-compatible, individually rational, and ex ante budget balanced mechanisms. We shall refer to these mechanism as "second best."

We may assume without loss of generality that the mechanism designer balances the budget exactly rather than leaving a surplus. If there is a surplus then the mechanism designer can return it to the agents. Moreover, the mechanism designer can achieve ex post budget balance when she achieves ex ante budget balance. Thus, in each state in which the public good is produced, payments will add up to c, and in other states they will be add up to zero. The designer's objective function can therefore be written as:

$$\int_{\Theta} q(\theta)\left(\sum_{i\in I}\theta_i - c\right) f(\theta)d\theta \tag{3.25}$$

It seems at first sight most natural from our discussion so far to regard q and the interim expected payments of the lowest types, $T_i(\underline{\theta})$, as the designer's choice variables. However, it is equivalent, and more convenient, to think of q and the interim expected utilities of the lowest types, $U_i(\underline{\theta})$

as the choice variables. The constraints that these variables have to satisfy are:

$$\text{for every } i \in I \text{ the function } Q_i \text{ is monotonically increasing}$$
$$(\textit{incentive constraint}); \tag{3.26}$$

$$\text{for every } i \in I : U_i(\underline{\theta}) \geq 0$$
$$(\textit{individual rationality constraint}); \tag{3.27}$$

$$-\sum_{i \in I} U_i(\underline{\theta}) + \int_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta) d\theta = 0$$
$$(\textit{budget constraint}), \tag{3.28}$$

where the budget constraint uses expression 3.19 for the expected revenue.

We now eliminate the choice variables $U_i(\underline{\theta})$ from the problem, and instead write the budget constraint as:

$$\int_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta) d\theta \geq 0 \tag{3.29}$$

with the understanding that if the left hand side is strictly positive, the interim expected utilities of the lowest types $U_i(\underline{\theta})$ will be reduced so as to satisfy exact budget balance.

We now solve the mechanism designer's problem by first considering the relaxed problem where we neglect the monotonicity constraint (3.26). Then we shall discuss conditions under which the solution to the problem without constraint (3.26) actually happens to satisfy condition (3.26) as well. Under those conditions the solution to the relaxed problem has to be a solution to the original problem.

To solve the relaxed problem we use a version of the Kuhn Tucker Theorem that applies to infinite dimensional vector spaces, such as function spaces. Theorems 1 and 2 in Luenberger (1969), p. 217 and p. 221, apply. According to these results, q solves the relaxed maximization problem if and only if there is a Lagrange multiplier $\lambda \geq 0$ such that q maximizes

$$\int_{\Theta} q(\theta) \left(\sum_{i \in I} \theta_i - c \right) f(\theta) d\theta + \lambda \int_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta) d\theta, \tag{3.30}$$

and, moreover, $\lambda = 0$ only if:

$$\int_{\Theta} q(\theta)\left[\sum_{i\in I}\left(\theta_i - \frac{1-F_i(\theta_i)}{f_i(\theta_i)}\right) - c\right] f(\theta)d\theta > 0. \tag{3.31}$$

We don't go through the details of checking the applicability of these results. However, we note that we obtain a necessary and sufficient condition because in our problem the set of admissible functions q is convex, and because the objective function that we seek to maximize is linear, hence concave.

We can write the Lagrange function (3.30) as:

$$\int_{\Theta} q(\theta)(1+\lambda)\left[\sum_{i\in I}\left(\theta_i - \frac{\lambda}{1+\lambda}\frac{1-F_i(\theta_i)}{f_i(\theta_i)}\right) - c\right] f(\theta)d\theta. \tag{3.32}$$

It is evident, from point-wise maximization, that the Lagrange function is maximized if we set $q(\theta) = 1$ whenever the expression in the square brackets is positive. This leads to the following decision rule:

$$q^*(\theta) = \begin{cases} 1 & \text{if } \sum_{i\in I}\theta_i > c + \sum_{i\in I}\left(\frac{\lambda}{1+\lambda}\frac{1-F_i(\theta_i)}{f_i(\theta_i)}\right) \\ 0 & \text{otherwise.}\end{cases} \tag{3.33}$$

Note that we must have $\lambda > 0$ if Proposition 3.7 applies, because with $\lambda = 0$ the rule (3.33) becomes the first best rule.

Now we introduce an assumption under which the rule (3.33) satisfies the monotonicity constraint (3.26) for every $\lambda > 0$. The condition is the regularity assumption that we introduced in Assumptions 2.1 and 3.1.

Assumption 3.2. *For every $i \in I$ the cumulative distribution function F_i is regular, i.e. the function $\psi_i(\theta_i) = \theta_i - \frac{1-F(\theta_i)}{f(\theta_i)}$ is strictly increasing.*

If an agent i's cumulative distribution function F_i is regular, then

$$\theta_i - \frac{\lambda}{1+\lambda}\frac{1-F_i(\theta_i)}{f_i(\theta_i)} \tag{3.34}$$

is strictly increasing for every $\lambda > 0$. This is because the potentially decreasing term, which has weight 1 in ψ_i, has weight $\lambda/(1+\lambda) < 1$ in (3.34). This implies that the second best rule in (3.33) satisfies the monotonicity condition (3.26). Therefore, we can conclude:

Proposition 3.8. *Suppose $N\underline{\theta} < c < N\bar{\theta}$, and that for every agent $i \in I$ the function F_i is regular. Then a direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is incentive-compatible, individually rational, and ex ante budget balanced, and maximizes expected welfare among all such mechanisms, if and only if:*

(i) there is some $\lambda > 0$ such that for all $\theta \in \Theta$:

$$q(\theta) = \begin{cases} 1 & \text{if } \sum_{i \in I} \theta_i > c + \sum_{i \in I} \left(\frac{\lambda}{1+\lambda} \frac{1-F_i(\theta_i)}{f_i(\theta_i)} \right) \\ 0 & \text{otherwise.} \end{cases}$$

(ii)

$$\int\limits_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta) d\theta = 0.$$

(iii) for all $i \in I$:

$$T_i(\theta_i) = \theta_i Q_i(\theta_i) - \int_{\underline{\theta}}^{\theta_i} Q_i(x) dx.$$

We can see that the second best mechanism undersupplies the public good. The public good is produced only if the sum of the valuations is larger than a lower bound that is strictly larger than c. An interesting question is whether there are simple and appealing indirect mechanisms that implements the second best.

3.3.5 Profit Maximization

We now consider briefly the problem of choosing the mechanism that maximizes the designer's expected profits among all incentive-compatible and individually rational direct mechanisms. Like the welfare maximizing mechanism designer also the profit maximizing mechanism designer has two choice variables: the allocation rule q, and the transfer payments of the lowest types, $T_i(\underline{\theta}_i)$. We do not assume that the lowest type is necessarily 0. Profit maximization requires that the transfer payments of the lowest types are set equal to those types' expected utility. This leaves q as the only choice variable. The expected profit from decision rule q can be calculated as previously in similar contexts as:

$$\int\limits_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta) d\theta \qquad (3.35)$$

The mechanism designer has to respect the constraint that for every $i \in I$ the function Q_i must be monotonically increasing.

The mechanism designer's problem can be solved using analogous reasoning as in the auction model of the previous section. We only state the result.

Proposition 3.9. *Suppose that for every agent $i \in I$ the cumulative distribution function F_i is regular. Then a direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is incentive-compatible, and individually rational, and maximizes expected profit among all such mechanisms if and only if for all $i \in I$ and all $\theta \in \Theta$:*

(i)

$$q(\theta) = \begin{cases} 1 & \text{if } \sum_{i \in I} \theta_i > c + \sum_{i \in I} \frac{1 - F_i(\theta_i)}{f_i(\theta_i)}; \\ \\ 0 & \text{otherwise.} \end{cases}$$

(ii)

$$T_i(\theta_i) = \theta_i Q_i(\theta_i) - \int_{\underline{\theta}}^{\theta_i} Q_i(x) dx.$$

Comparing Propositions 3.8 and 3.9 we find that the profit maximizing supplier for the public good supplies a lower quantity than the welfare maximizing mechanism designer.

3.3.6 A Numerical Example

We provide a very simple numerical example, yet the calculations turn out to be somewhat involved.

Example 3.3. *Suppose $N = 2$, θ_i is uniformly distributed on $[0, 1]$ for $i = 1, 2$, and $0 < c < 2$. Note that the regularity assumption is satisfied as was verified in the numerical example of the previous section. We wish to determine first the expected welfare maximizing mechanism. By Proposition 3.7 the first best cannot be achieved, as $\underline{\theta} = 0$, and $\bar{\theta} = 1$, and hence $N\underline{\theta} < c < N\bar{\theta}$. The probability with which the first best rule calls for the production of the public good is strictly between zero and one.*

By Proposition 3.8 a necessary condition for a direct incentive-compatible, individually rational, and ex ante budget balanced mechanism to maximize *expected welfare among all such mechanisms is that there exists some* $\lambda > 0$ *such that* $q(\theta) = 1$ *if and only if:*

$$\theta_1 + \theta_2 > c + \frac{\lambda}{1+\lambda}\left(\frac{1-\theta_1}{1} + \frac{1-\theta_2}{1}\right) \Leftrightarrow$$
$$\theta_1 + \theta_2 > \frac{1+\lambda}{1+2\lambda}c + \frac{\lambda}{1+2\lambda}2 \tag{3.36}$$

Denote the right hand side of this inequality by s*. Proposition 3.8 thus means that we can restrict our search for second best mechanisms to those mechanisms for which* $q(\theta) = 1$ *if and only if* $\theta_1 + \theta_2 \geq s$ *for some* $s \in (c, 2)$*. We seek to find the an appropriate value of* s*.*

We determine s *by assuming that the interim expected payments of the lowest types are zero, as required by Proposition 3.8, that the interim expected payments of all other types are as required by incentive compatibility, and that the budget surplus on the right hand side of equation (3.29) is zero, as required by the Lagrange conditions. To proceed, we calculate the total expected cost of producing the public good, denoted by* $C(s)$*, and the total revenue of a mechanism, denoted by* $R(s)$*, for incentive-compatible, individually rational threshold mechanisms. We distinguish two cases.*

CASE 1: *Suppose* $s \leq 1$*. Then the expected cost of producing the public good is:*

$$C(s) = \left(1 - \frac{1}{2}s^2\right)c. \tag{3.37}$$

Next we calculate the expected payment by the agent 1:

$$\int_\Theta q(\theta)\left(\theta_1 - \frac{1-F_1(\theta_1)}{f_1(\theta_1)}\right)f(\theta)d\theta$$
$$= \int_0^s \int_{s-\theta_1}^1 (2\theta_1 - 1)d\theta_2 d\theta_1 + \int_s^1 \int_0^1 (2\theta_1 - 1)d\theta_2 d\theta_1$$
$$= -\frac{1}{3}s^{+}\frac{1}{2}s^2 \tag{3.38}$$

Agent 2's expected payment will be the same. Therefore, the total expected revenue of the mechanism is:

$$R(s) = -\frac{2}{3}s^3 + s^2. \tag{3.39}$$

CASE 2: *Suppose $s > 1$. Then the expected costs of producing the public good are:*

$$C(s) = \frac{1}{2}(2-s)^2 c. \tag{3.40}$$

The expected payment by agent 1 is given by:

$$\begin{aligned}
&\int_\Theta q(\theta)\left(\theta_1 - \frac{1-F_1(\theta_1)}{f_1(\theta_1)}\right) f()\theta)d\theta \\
&= \int_{s-1}^{1}\int_{s-\theta_1}^{1}(2\theta_1 - 1)d\theta_2 d\theta_1 \\
&= \frac{1}{6} - \frac{1}{2}(s-1)^2 + \frac{1}{3}(s-1)^3. &(3.41)
\end{aligned}$$

Agent 2's expected payments will be the same, and therefore the total expected revenue from the mechanism is:

$$R(s) = \frac{1}{3} - (s-1)^2 + \frac{2}{3}(s-1)^3. \tag{3.42}$$

We now define $D(s) \equiv R(s) - C(s)$, so that the condition for s becomes $D(s) = 0$. To understand the set of solutions to this equation we first investigate the sign of the derivative of D with respect to s. Observe that:

$$0 < s < 1 \Rightarrow D'(s) = -2s^2 + (2+c)s = s(2(1-s)+c) > 0; \tag{3.43}$$

$$\begin{aligned}
1 < s < 2 \Rightarrow D'(s) &= -2(s-1) + 2(s-1)^2 + (2-s)c \\
&= (2-s)(c - 2(s-1)). &(3.44)
\end{aligned}$$

The term in the last line is positive if and only if:

$$\begin{aligned}
c - 2(s-1) > 0 \Leftrightarrow& \\
&s < 1 + \frac{c}{2} &(3.45)
\end{aligned}$$

Next, we investigate the sign of $D(s)$ for some important values of s. Note first that:

$$D(0) = -c < 0 \text{ and} \tag{3.46}$$

$$D(2) = 0. \tag{3.47}$$

Next, we investigate the sign of D at $s = 1 + \frac{c}{2}$. We obtain:

$$D\left(1+\frac{c}{2}\right) = \frac{1}{3} - \left(\frac{c}{2}\right)^2 + \frac{2}{3}\left(\frac{c}{2}\right)^3 - \frac{1}{2}\left(1-\frac{c}{2}\right)^2 c$$
$$= \frac{1}{3} - \frac{1}{2}c + \left(\frac{c}{2}\right)^2 - \frac{1}{3}\left(\frac{c}{2}\right)^3 \tag{3.48}$$

We want to prove that the expression in the last line is strictly positive. It clearly tends to a positive limit as $c \to 0$. If $c \to 2$, then the expression tends to zero. Thus, it is sufficient to show that the derivative of this expression is negative for $0 < c < 2$. This derivative is:

$$-\frac{1}{2} + \frac{c}{2} - \frac{1}{2}\left(\frac{c}{2}\right)^2$$
$$= \frac{c}{2}\left(1-\frac{c}{2}\right) - \frac{1}{2}$$
$$\leq \frac{1}{4} - \frac{1}{2}$$
$$= -\frac{1}{4} \tag{3.49}$$

We can conclude that $D(1 + \frac{1}{2}c) > 0$.

From our results so far we can conclude that the equation $D(s) = 0$ has exactly two solutions: one solution in the interval $(0, 1+\frac{c}{2})$, and the solution $s = 2$. We can discard the latter as we require $s \in (c, 2)$.

For the purpose of calculating the solution, it is useful to ask whether the solution of $D(s) = 0$ is larger or less than 1. For this we investigate the value of $D(1)$:

$$D(1) = \frac{1}{3} - \frac{1}{2}c > 0 \Leftrightarrow$$
$$c < \frac{2}{3} \tag{3.50}$$

Thus, the solution of $D(s) = 0$ will be between 0 and 1 if and only if $c < \frac{2}{3}$. Otherwise, it will be between 1 and $1 + \frac{c}{2}$.

If $c < \frac{2}{3}$, then we find s by solving the following equation:

$$-\frac{2}{3}s^3 + s^2 - \left(1 - \frac{1}{2}s^2\right)c = 0. \tag{3.51}$$

Unfortunately, this equation has no simple solution. If $c > \frac{2}{3}$, then we find s by solving the following equation:

$$\frac{1}{3} - (s-1)^2 + \frac{2}{3}(s-1)^3 - \frac{1}{2}(2-a)^2 c = 0. \tag{3.52}$$

It so happens that this equation has a simple analytical solution:

$$s = \frac{1}{2} + \frac{3}{4}c. \tag{3.53}$$

We can now sum up:

Proposition 3.10. *The utilitarian mechanism designer will choose a mechanism with an allocation rule q such that:*

$$q^*(\theta) = \begin{cases} 1 & \text{if } \theta_1 + \theta_2 > s; \\ 0 & \text{otherwise.} \end{cases}$$

where s is determined as follows:

(i) If $c < \frac{2}{3}$, then s is the unique solution in $[0,1]$ of the equation:

$$-\frac{2}{3}s^3 + s^2 - \left(1 - \frac{1}{2}s^2\right)c = 0.$$

(ii) If $c \geq \frac{2}{3}$, then

$$s = \frac{1}{2} + \frac{3}{4}c.$$

The expected profit maximizing mechanism follows straightforwardly from Proposition 3.9:

Proposition 3.11. *The expected profit maximizing mechanism designer will choose a mechanism with an allocation rule q such that:*

$$q^*(\theta) = \begin{cases} 1 & \text{if } \theta_1 + \theta_2 > s; \\ 0 & \text{otherwise.} \end{cases}$$

where $s = 1 + \frac{1}{2}c$.

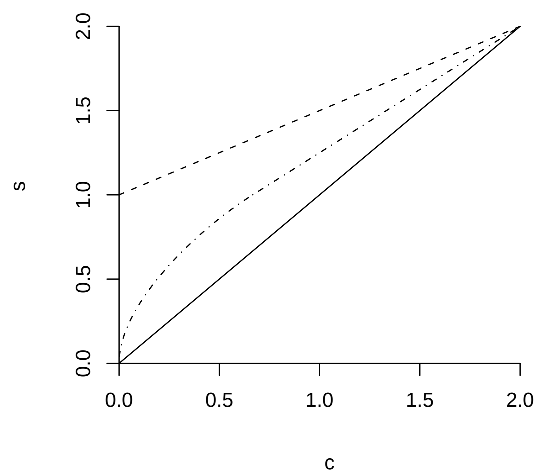

Figure 3.2: Thresholds in First-Best, Second-Best, and Profit-Maximizing Mechanisms

To illustrate our findings, we have plotted in Figure 3.2 the optimal threshold s for the utilitarian case (dashed and dotted line) and for the profit maximization case (dashed line) as a function of c. For comparison we have also plotted the 45° line which is the "first best" threshold (unbroken line).

Figure 3.2 shows that the second best threshold chosen by a welfare maximizing mechanism designer is strictly larger than the first best threshold, except if c is either 0 or 2. This reflects the fact that to provide incentives the mechanism designer must accept some inefficiencies. The inefficiencies become small as c approaches either 0 or 2. These two extreme cases correspond to cases in which the mechanism designer doesn't really have to induce truthful revelation of individuals' types, either because the production of the public good is free, and hence it should always be produced, or it is prohibitively expensive, and hence it should never be produced.

Figure 3.2 also shows that a monopoly supplier of the public good would choose a threshold that is even larger than the threshold chosen by the utilitarian mechanism designer. This corresponds to the standard textbook insight that a monopolist artificially restricts supply to raise price. The monopoly distortion exists even if production is free ($c = 0$), as in the standard textbook case. If production becomes prohibitively expensive ($c = 2$), the monopolist will not want to provide the good. Thus, the monopoly solution coincides with the first best, and the utilitarian solution. In that case, the distortion introduced by a monopolist is zero.

3.4 Bilateral Trade

3.4.1 Setup

In our next example a seller of a single indivisible good faces one buyer with unknown valuation, as in Section 2.2, but now we consider the situation from the perspective of a mechanism designer who wants to arrange a trading institution for the two parties that guarantees that they trade if and only if the buyer's value is larger than the seller's. Moreover, not only the buyer's valuation, but also the seller's valuation is unknown to the designer of this trading institution, and it may be that valuations are such that trade is not efficient. This example was first analyzed in the very well-known paper by Myerson and Satterthwaite (1983). It is the simplest example that one might analyze when seeking to build a general theory of the design of optimal trading institution, such as stock exchanges or commodity markets. Methodologically, it is interesting to see how one can treat the bilateral trade problem with similar methods as the public goods problem, although it is seemingly quite different.

A seller S owns a single indivisible good. There is one potential buyer B. Define $I = \{S, B\}$. The seller's utility if he sells the good and receives a transfer payment t is equal to t. If he does not sell the good and receives a transfer t then his utility is $\theta_S + t$ where θ_S is a random variable with cumulative distribution function F_S and density f_S. We assume that F_S has support $[\underline{\theta}_S, \bar{\theta}_S]$ and that $f_S(\theta_S) > 0$ for all $\theta_S \in [\underline{\theta}_S, \bar{\theta}_S]$. The buyer's utility if he purchases the good and pays a transfer t equals $\theta_B - t$, where θ_B is a random variable with cumulative distribution function F_B and density f_B. We assume that F_B has support $[\underline{\theta}_B, \bar{\theta}_B]$ and that $f_B(\theta_B) > 0$ for all $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$. The buyer's utility if he does not obtain the good and pays transfer t is $-t$. The random variables θ_S and θ_B are independent. We define $\theta = (\theta_S, \theta_B)$ and $\Theta = [\underline{\theta}_S, \bar{\theta}_S] \times [\underline{\theta}_B, \bar{\theta}_B]$. We denote the joint distribution of θ by F with density f. The seller only observes θ_S, and the buyer only observes θ_B.

3.4.2 Direct Mechanisms

The revelation principle implies that we can restrict our attention to direct mechanisms.

Definition 3.9. A "direct mechanism" consists of functions q, t_S and t_B where:

$$q : \Theta \to \{0, 1\}$$

and for $i = S, B$:

$$t_i : \Theta \to \mathbb{R}.$$

The function q assigns to each type vector θ an indicator variable that indicates whether trade takes place ($q(\theta) = 1$) or whether no trade takes place ($q(\theta) = 0$). We shall refer to q also as the "trading rule." For simplicity, we restrict our attention to deterministic trading rules. The function t_S indicates transfers that the seller receives, and the function t_B indicates transfers that the buyer makes. We shall mostly assume in this section[12] that $t_S = t_B$, i.e. that the seller receives what the buyer pays. Therefore, it seems redundant to introduce separate notation for the buyer's payment and the seller's receipts. However, it is much more convenient to begin with

[12] An exception is Subsection 3.4.4, where we consider profit maximizing trading mechanisms.

a more general framework in which these variables are not identical, and then to introduce an ex post budget balance condition which forces them to be equal to each other, as we shall do below. This allows us then to adopt a similar methodology as in the previous section.

Given a direct mechanism, we define for each agent $i \in \{S, B\}$ functions $Q_i : [\underline{\theta}_i, \bar{\theta}_i] \rightarrow [0, 1]$ and $T_i : [\underline{\theta}_i, \bar{\theta}_i] \rightarrow \mathbb{R}$ where $Q_i(\theta_i)$ is the conditional probability that trade takes place, where we condition on agent i's type being θ_i, and $T_i(\theta_i)$ is the conditional expected value of the transfer that agent i receives (if $i = S$) or makes (if $i = B$), again conditioning on agent i's type being θ_i. Finally, we also define agent i's expected utility $U_i(\theta_i)$ conditional on her type being θ_i. This is given by: $U_S(\theta_S) = T_S(\theta_S) + (1 - Q_S(\theta_S))\theta_S$ and $U_B(\theta_B) = Q_B(\theta_B)\theta_B - T_B(\theta_B)$.

We shall restrict our attention to direct mechanisms that are incentive-compatible, individually rational, and ex post budget balanced. Individual rationality is defined as before. Standard arguments show that a mechanism is incentive-compatible for the buyer under exactly the same conditions as before. For the seller, the standard arguments apply if types are ordered in the reverse of the numerical order, that is starting with high types rather than low types. Thus, a necessary and sufficient condition for incentive compatibility for the seller is that Q_S is decreasing, and that T_S is given by:

$$T_S(\theta_S) = T_S(\bar{\theta}_S) + (1 - Q_S(\bar{\theta}_S))\bar{\theta}_S - (1 - Q_S(\theta_S))\theta_S - \int_{\theta_S}^{\bar{\theta}_S} (1 - Q_S(x))\, dx \tag{3.54}$$

for all θ_S.

Individual rationality for the buyer is defined and characterized in the same way as before. For the seller, individual rationality means that $U_S(\theta_S) \geq \theta_S$ for all θ_S, that is, the seller trades voluntarily and obtains an expected utility that is at least as large as his utility would be if he kept the good. If a mechanism is incentive-compatible, then the seller's individual rationality condition holds if and only if it holds for the highest seller type $\bar{\theta}_S$. To prove this we show that the difference $U_S(\theta_S) - \theta_S$ is decreasing in θ_S. A seller who is of type θ_S and pretends to be of type θ'_S obtains expected utility:

$$\theta_S(1 - Q_S(\theta'_S)) + T_S(\theta'_S). \tag{3.55}$$

This exceeds θ_S by:

$$T_S(\theta'_S) - \theta_S Q_S(\theta'_S). \tag{3.56}$$

The seller maximizes over θ'_S. As θ_S increases, the function that the seller maximizes shifts downwards, and therefore the function's maximum value decreases. Thus, $U_S(\theta_S) - \theta_S$ is decreasing in θ_S.

Ex post budget balance requires that in each state θ we have: $t_S(\theta) = t_B(\theta)$. By the same argument as in the previous section it suffices to identify ex ante budget balanced mechanisms, that is, mechanisms for which the seller's ex ante expected payment is equal to the buyer's ex ante expected payment. We shall work with this condition. The by now familiar calculation shows that for an incentive-compatible direct mechanism the seller's expected payment is:

$$U_S(\bar{\theta}_S) - \int_{\Theta} (1 - q(\theta)) \left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)} \right) f(\theta) d\theta \tag{3.57}$$

and the buyer's expected payment is:

$$-U_B(\underline{\theta}_B) + \int_{\Theta} q(\theta) \left(\theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)} \right) f(\theta) d\theta. \tag{3.58}$$

3.4.3 Welfare Maximization

The mechanism designer seeks to maximize the sum of the individuals' utilities,

$$\begin{aligned} & q(\theta)\theta_B - t_B + (1 - q(\theta))\theta_S + t_S \\ = & \theta_S + q(\theta)\left(\theta_B - \theta_S\right) + t_S - t_B. \end{aligned} \tag{3.59}$$

If the mechanism designer were not constrained by incentive compatibility and individual rationality, but only had to respect ex post budget balance, then the mechanism designer would choose a "first best mechanism" where the trading rule is:

$$q^*(\theta) = \begin{cases} 1 & \text{if } \theta_B \geq \theta_S; \\ \\ 0 & \text{otherwise.} \end{cases} \tag{3.60}$$

where the decision in the case of equality of values is arbitrary. The payment rule is arbitrary as long as it is ex post budget balanced.

Our first objective is to prove that in almost all cases no first best mechanism is incentive-compatible and individually rational.

Proposition 3.12 (Myerson and Satterthwaite, 1983)**.** *An incentive-compatible, individually rational and ex-post budget balanced direct mechanism with decision rule q^* exists if and only if $\underline{\theta}_B \geq \bar{\theta}_S$ or $\underline{\theta}_S \geq \bar{\theta}_B$.*

The condition $\underline{\theta}_B \geq \bar{\theta}_S$ implies that trade is always at least weakly efficient. The condition $\underline{\theta}_S \geq \bar{\theta}_B$ implies that efficiency never requires trade. Thus, these are trivial cases. In all non-trivial cases, there is no incentive-compatible, individually rational and ex post budget balanced first best mechanism.

Proof. The "if" part of Proposition 3.12 is trivial. If $\underline{\theta}_S \geq \bar{\theta}_B$, then a mechanism under which no trade takes place and no payments are made is first best and has the required properties. If $\underline{\theta}_S \geq \bar{\theta}_B$, a mechanism where trade always takes place and the buyer always pays the seller some price $p \in [\bar{\theta}_B, \underline{\theta}_S]$ is first best and has the required properties.

To prove the "only-if" part, we proceed in a similar way as in the proof of Proposition 3.7. We first display a mechanism that is incentive-compatible and individually rational and that implements the first best trading rule. Then we argue that this mechanism maximizes expected surplus of buyer's payment over seller's receipts among all incentive-compatible and individually rational mechanisms that implement the first best trading rule. Finally, we show that the mechanism has an expected deficit if the condition of Proposition 3.7 holds. The assertion follows.

Definition 3.10. The "pivot mechanism" is the mechanism that is given by the first best trading rule and by the following transfer schemes:

$$t_S(\theta) = q^*(\bar{\theta}_S, \theta_B)\bar{\theta}_S + \left(q^*(\theta) - q^*(\bar{\theta}_S, \theta_B)\right)\theta_B$$

$$t_B(\theta) = q^*(\theta_S, \underline{\theta}_B)\underline{\theta}_B + \left(q^*(\theta) - q^*(\theta_S, \underline{\theta}_B)\right)\theta_S$$

for all $\theta \in \Theta$.

In this mechanism, the seller receives a constant plus the buyer's valuation of the good if the seller's report was pivotal for the trade happening. The buyer pays a constant plus the seller's valuation of the good if the buyer's report was pivotal for the trade happening. As in the previous section, the pivot mechanism has in common with a general Vickrey Clarke Groves mechanism that each agent's private interests are exactly aligned with social welfare. Moreover, the individual rationality constraints of the two extreme types, the lowest type of the seller, and the highest type of the buyer, are satisfied with equality. This becomes clear in the proof of the following result.

Lemma 3.9. *The pivot mechanism is incentive-compatible and individually rational.*

Proof. Suppose the seller is of type θ_S and contemplates reporting that she is of type $\theta'_S \neq \theta_S$. We fix we the buyer's type θ_B. We are going to show that truthful reporting is optimal *whatever the buyer's type.*

The seller's utility if she is of type θ_S but reports that she is of type θ'_S, leaving out all terms that do not depend on her report, is:

$$q^*(\theta'_S, \theta_B)(\theta_B - \theta_S). \tag{3.61}$$

Note that the first best decision rule q^* maximizes the social surplus $q^*(\theta)(\theta_B - \theta_S)$.[13] The seller's utility is thus equal to social surplus if the collective decision is $q^*(\theta'_S, \theta_B)$. Because the first best rule maximizes social surplus, it is optimal for the seller to report θ_S truthfully. A similar argument applies to the buyer.

To check individual rationality, it is sufficient to consider the types $\bar{\theta}_B$ and $\underline{\theta}_S$. From the payment rule, it is clear that the type $\bar{\theta}_B$ receives expected utility $\bar{\theta}_B$ whatever the seller's type, and the type $\underline{\theta}_S$ receives expected utility zero whatever the buyer's type. Individual rationality follows. □

Lemma 3.10. *The difference between the buyer's and the seller's expected payments is at least as large under the pivot mechanism as it is under any incentive-compatible and individually rational direct mechanism that implements the first best trading rule.*

Proof. Our discussion of interim expected transfers for incentive-compatible mechanisms shows that these depend only on the trading rule and on the payments of the highest buyer type and the lowest seller type. We have seen in the proof of Lemma 3.9 that the expected utility of the highest seller type in the pivot mechanism is $\bar{\theta}_S$. This implies that the interim expected payments to the seller in the pivot mechanism are the lowest payments compatible with the first best trading rule. Similarly, the expected utility of the lowest buyer type in the pivot mechanism is zero, and therefore the interim expected payments that the seller makes in the pivot mechanism are the highest payments compatible with the first best trading rule. Thus, the pivot mechanism maximizes the ex ante expected difference between buyer's

[13] We use the expression "social surplus" to distinguish this expression from social welfare as defined in equation (3.59). The first best decision rule maximizes both, social welfare and social surplus.

and seller's payments among all mechanisms that implement the first best trading rule and that are incentive-compatible. □

We show next that the pivot mechanism has an ex ante expected deficit if the condition of Proposition 3.12 holds.

Lemma 3.11. *The difference between the buyer's ex ante expected payment and the seller's ex ante expected payment is negative in the pivot mechanism if the condition of Proposition 3.12 is violated, i.e. if* $\underline{\theta}_B < \bar{\theta}_S$ *and* $\bar{\theta}_B > \underline{\theta}_S$.

Proof. We shall show that for every realized type vector θ the pivot mechanism either has a surplus of zero, or a deficit. Moreover, we shall argue that the types for which it has a deficit have positive probability. This will then imply the result.

Consider first values of θ such that $q^*(\theta) = 0$. Then both agents' transfers equal zero, and hence the deficit is zero. Now consider values of θ such that $q^*(\theta) = 1$. We calculate the ex post deficit for four different scenarios.

Case 1: $q^*(\theta) = 1, q^*(\bar{\theta}_S, \theta_B) = 1, q^*(\theta_S, \underline{\theta}_B) = 1$. Then the ex post deficit is: $\underline{\theta}_B - \bar{\theta}_S$. This is negative by the assumption of the Lemma.

Case 2: $q^*(\theta) = 1, q^*(\bar{\theta}_S, \theta_B) = 1, q^*(\theta_S, \underline{\theta}_B) = 0$. Then the ex post deficit is: $\theta_S - \bar{\theta}_S$. This is obviously not positive.

Case 3: $q^*(\theta) = 1, q^*(\bar{\theta}_S, \theta_B) = 0, q^*(\theta_S, \underline{\theta}_B) = 1$. Then the ex post deficit is: $\underline{\theta}_B - \theta_B$. This is obviously not positive.

Case 4: $q^*(\theta) = 1, q^*(\bar{\theta}_S, \theta_B) = 0, q^*(\theta_S, \underline{\theta}_B) = 0$. Then the ex post deficit is: $\theta_S - \theta_B$. This is not positive because we are considering θ such that trade is efficient.

The condition of the Lemma means that the intervals $[\underline{\theta}_B, \bar{\theta}_B]$ and $[\underline{\theta}_S, \bar{\theta}_S]$ have an intersection with non-empty interior. Consider any point in this intersection for which trade is efficient and such that $\theta_S > \theta_B$. By drawing a simple diagram one can see that the probability measure of such points is positive. Moreover, at any such point we are in Case 4, and moreover the deficit is strictly negative. Therefore, there is an expected deficit. □

This completes the proof of Proposition 3.12. □

We shall now focus on the case in which according to Proposition 3.12 no first best mechanism is incentive-compatible, individually rational, and ex post budget balanced. Our objective is to determine direct mechanisms that maximize expected welfare among all incentive-compatible, individually rational, and ex post budget balanced mechanisms. We shall refer to these mechanisms as "second best."

As in the public goods case, we may assume without loss of generality that the mechanism designer balances the ex post budget exactly rather than leaving a surplus. The mechanism designer's welfare function can then be written as $\theta_S + q(\theta)(\theta_B - \theta_S)$. Moreover, note that θ_S is not affected by the mechanism designer's decisions. Therefore, we shall assume that the mechanism designer seeks to maximize:

$$\int_\Theta q(\theta)\left(\theta_B - \theta_S\right) f(\theta)d\theta. \tag{3.62}$$

As in the previous subsection it is convenient to think of the designer's choice variables as the trading rule q and the interim expected utilities of the highest type of the seller, $U_B(\bar{\theta}_B)$, and of the lowest type of the buyer, $U_S(\underline{\theta}_S)$. The constraints are:

$$Q_S \text{ is increasing and } Q_B \text{ is decreasing}$$
$$(\textit{incentive compatibility}); \tag{3.63}$$

$$U_S(\bar{\theta}_S) \geq \bar{\theta}_S \text{ and } U_B(\underline{\theta}_B) \geq 0$$
$$(\textit{individual rationality}); \tag{3.64}$$

$$-U_B(\underline{\theta}_B) + \int_\Theta q(\theta)\left(\theta_B - \frac{1-F_B(\theta_B)}{f_B(\theta_B)}\right) f(\theta)d\theta =$$
$$U_S(\bar{\theta}_S) - \int_\Theta (1-q(\theta))\left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)d\theta$$
$$(\textit{budget balance}). \tag{3.65}$$

We simplify this optimization problem a little. First, we write the budget constraint as:

$$\int_\Theta q(\theta)\left[\left(\theta_B - \frac{1-F_B(\theta_B)}{f_B(\theta_B)}\right) - \left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right)\right] f(\theta)d\theta =$$
$$U_S(\bar{\theta}_S) + U_B(\underline{\theta}_B) - \int_\Theta \left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)d\theta \tag{3.66}$$

Next, we drop the choice variables $U_S(\bar{\theta}_S)$ and $U_B(\underline{\theta}_B)$ and re-write the budget constraint as:

$$\int_{\Theta} q(\theta)\left[\left(\theta_B - \frac{1-F_B(\theta_B)}{f_B(\theta_B)}\right) - \left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right)\right] f(\theta)d\theta \geq \bar{\theta}_S - \int_{\Theta}\left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)d\theta \tag{3.67}$$

If this inequality is strict, then the mechanism designer can allocate the difference as additional utility to the buyer and the seller. Below, we shall denote the right hand side of (3.67) by K.

When solving the mechanism designer's optimization problem we first neglect the incentive constraint. We have a concave maximization problem because the objective function and the constraint are both linear in q. Therefore, by Theorems 1 and 2 in Luenberger (1969), p. 217 and p. 221, a necessary and sufficient condition for q to be optimal is that there is a Lagrange multiplier $\lambda \geq 0$ such that q maximizes:

$$\int_{\Theta} q(\theta)\left[(1+\lambda)\theta_B - \lambda\frac{1-F_B(\theta_B)}{f_B(\theta_B)} - (1+\lambda)\theta_S - \lambda\frac{F_S(\theta_S)}{f_S(\theta_S)}\right] f(\theta)d\theta - \lambda K \tag{3.68}$$

and moreover the budget constraint (3.67) holds as a strict inequality only if $\lambda = 0$.

The Lagrange function is maximized by choosing $q(\theta)$ to be positive whenever the expression in the square brackets is positive. Brief manipulation of this condition leads to the following trading rule:

$$q(\theta) = \begin{cases} 1 & \text{if } \theta_B - \frac{\lambda}{1+\lambda}\frac{1-F_B(\theta_B)}{f_B(\theta_B)} \geq \theta_S + \frac{\lambda}{1+\lambda}\frac{F_S(\theta_S)}{f_S(\theta_S)} \\ \\ 0 & \text{otherwise.} \end{cases} \tag{3.69}$$

Can we have $\lambda = 0$? Then the optimal decision rule (3.69) would be identical to the first best trading rule, and the budget constraint would be violated by Proposition 3.12. Therefore, we must have: $\lambda > 0$.

We now make a regularity assumption that ensures that this decision rule satisfies the monotonicity constraint (3.63) of the mechanism designer's maximization problem.

Assumption 3.3. *The seller's distribution function F_S is "regular," i.e.*

$$\psi_S(\theta_S) \equiv \theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}$$

is monotonically increasing. The buyer's distribution function F_B is "regular," i.e.

$$\psi_B(\theta_B) \equiv \theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)}$$

is monotonically increasing.

If F_S and F_B are regular,

$$\theta_S + \frac{\lambda}{1+\lambda}\frac{F_S(\theta_S)}{f_S(\theta_S)} \text{ and } \theta_B - \frac{\lambda}{1+\lambda}\frac{1 - F_B(\theta_B)}{f_B(\theta_B)} \tag{3.70}$$

are increasing for every $\lambda > 0$. In this case the trading rule q derived above will imply that the function Q_S is decreasing and the function Q_B is increasing. We can then conclude:

Proposition 3.13. *Suppose $\underline{\theta}_B < \bar{\theta}_S$ and $\bar{\theta}_B > \underline{\theta}_S$, and that for every agent $i \in I$ the function F_i is regular. Then a direct mechanism (q, t_S, t_B) is incentive-compatible, individually rational, and ex ante budget balanced, and maximizes expected welfare among all such mechanisms if and only if:*

(i) there is some $\lambda > 0$ such that for all $\theta \in \Theta$:

$$q(\theta) = \begin{cases} 1 & \text{if } \theta_B - \frac{\lambda}{1+\lambda}\frac{1-F_B(\theta_B)}{f_B(\theta_B)} \geq \theta_S + \frac{\lambda}{1+\lambda}\frac{F_S(\theta_S)}{f_S(\theta_S)} \\ \\ 0 & \text{otherwise.} \end{cases} \tag{3.71}$$

(ii) exact budget balance holds:

$$\int_\Theta q(\theta)\left[\left(\theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)}\right) - \left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right)\right] f(\theta)d\theta =$$
$$\bar{\theta}_S - \int_\Theta \left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)d\theta \tag{3.72}$$

(iii) the payment rules create incentive compatibility, i.e. for all $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$ and all $\theta_S \in [\underline{\theta}_S, \bar{\theta}_S]$:

$$T_B(\theta_B) = \theta_S Q_S(\theta_S) - \int_{\underline{\theta}_S}^{\theta_S} Q_S(x)dx$$

$$T_S(\theta_S) = \bar{\theta}_S - (1 - Q_S(\theta_S))\theta_S - \int_{\theta_S}^{\bar{\theta}_S} (1 - Q_S(x))\, dx$$

Note that in the second best mechanism the good changes hands less frequently than in the first best. The buyer's value, minus some discount, still has to be larger than the seller's value, plus some discount, for trade to take place.

3.4.4 Profit Maximization

We now consider briefly the problem of choosing the mechanism that maximizes the expected profits. We assume that the mechanism designer's profit is the difference between what the buyer pays and what the seller receives. We could think of this mechanism designer as the commercial designer of a trading platform who charges a fee for transactions on the platform.

The mechanism designer's choice variables are again the expected transfers of the highest seller type and the lowest buyer type and the trading rule. It is obvious that profit maximization implies that the expected transfer of the highest seller type and the lowest buyer type are chosen so as to make these types' individual rationality constraints true as equalities. Using the formulas (3.57) and (3.58) we can thus write expected revenue as:

$$\int_\Theta q(\theta)\left(\theta_B - \frac{1-F_B(\theta_B)}{f_B(\theta_B)}\right) f(\theta)d\theta$$
$$- \left[\bar{\theta}_S - \int_\Theta (1-q(\theta))\left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)d\theta\right]. \qquad (3.73)$$

Taking out constants that are independent of q, we are left with:

$$\int_\Theta q(\theta)\left(\theta_B - \frac{1-F_B(\theta_B)}{f_B(\theta_B)} - \theta_S - \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)d\theta. \qquad (3.74)$$

Using similar reasoning as in the last subsection, we obtain:

Proposition 3.14. *Suppose that for every agent $i \in I$, the distribution function F_i is regular. Then necessary and sufficient conditions for an incentive-compatible and individually rational direct mechanism to maximize expected profits are:*

(i)

$$q(\theta) = \begin{cases} 1 & \text{if } \theta_B - \frac{1-F_B(\theta_B)}{f_B(\theta_B)} > \theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}; \\ 0 & \text{otherwise.} \end{cases}$$

(ii) for all $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$ and all $\theta_S \in [\underline{\theta}_S, \bar{\theta}_S]$:

$$T_B(\theta_B) = \theta_B Q_B(\theta_B) - \int_{\underline{\theta}_B}^{\theta_B} Q_B(x)dx$$

$$T_S(\theta_S) = \bar{\theta}_S - (1-Q_S(\theta_S))\theta_S - \int_{\theta_S}^{\bar{\theta}_S} (1-Q_S(x))\,dx$$

Comparing Propositions 3.13 and 3.14 we find that the profit maximizing mechanism designer facilitates less trade than the welfare maximizing mechanism designer.

3.4.5 A Numerical Example

We conclude with a very simple numerical example.

Example 3.4. *Suppose that θ_S as well as θ_B are uniformly distributed on the interval $[0,1]$. Note that this satisfies the regularity condition. We want to determine the welfare and the profit maximizing mechanisms.*

By Proposition 3.13 the welfare maximizing rule will be such that trade takes place if and only if:

$$\theta_B - \frac{\lambda}{1+\lambda}\frac{1-F_B(\theta_B)}{f_B(\theta_B)} > \theta_S + \frac{\lambda}{1+\lambda}\frac{F_S(\theta_S)}{f_S(\theta_S)} \Leftrightarrow \tag{3.75}$$

$$\theta_B - \frac{\lambda}{1+\lambda}(1-\theta_B) > \theta_S + \frac{\lambda}{1+\lambda}\theta_S \Leftrightarrow \tag{3.76}$$

$$\theta_B - \theta_S > \frac{\lambda}{1+2\lambda}. \tag{3.77}$$

Thus, trade will take place if and only if the difference between the buyer's and the seller's valuation is above some positive threshold s. We shall now calculate for which such thresholds s the budget balance constraint holds.

Take s as given. The buyer's willingness to pay is given by:

$$\begin{aligned}
&\int_\Theta q(\theta)\left(\theta_B - \frac{1-F_B(\theta_B)}{f_B(\theta_B)}\right) f(\theta)d(\theta) \\
&= \int_s^1 \int_0^{\theta_B - s} (2\theta_B - 1) d\theta_S d\theta_B \\
&= \frac{1}{3}s^3 - \frac{1}{2}s^2 + \frac{1}{6}. \qquad (3.78)
\end{aligned}$$

The seller's expected transfer is given by:

$$\begin{aligned}
&\bar{\theta}_S - \int_\Theta (1-q(\theta))\left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)d\theta \\
&= 1 - 2\int_\Theta (1-q(\theta))\theta_S d\theta \\
&= 1 - 2\left(\int_0^{1-s}\int_0^{\theta_S+s} \theta_s d\theta_B d\theta_S + \int_{1-s}^1 \int_0^1 \theta_s d\theta_B d\theta_S\right)
\end{aligned}$$

$$= \frac{1}{3}(1-s)^3 \tag{3.79}$$

Budget balance is achieved when:

$$\frac{1}{3}s^3 - \frac{1}{2}s^2 + \frac{1}{6} = \frac{1}{3}(1-s)^3. \tag{3.80}$$

This has two solutions:

$$s = \frac{1}{4} \text{ and } s = 1. \tag{3.81}$$

Only the solution $s = \frac{1}{4}$ *is of the required form* $\frac{\lambda}{1+2\lambda}$ *with* $\lambda > 0$. *We conclude:*

Proposition 3.15. *In the welfare maximizing incentive-compatible, individually rational and ex post budget balanced trading mechanism trade takes place if and only if*

$$\theta_B - \theta_S > \frac{1}{4}. \tag{3.82}$$

From Proposition 3.14 the following characterization of the expected profit maximizing trading mechanism is obvious.

Proposition 3.16. *In the expected profit maximizing incentive-compatible and individually rational trading mechanism, trade takes place if and only if*

$$\theta_B - \theta_S > \frac{1}{2}. \tag{3.83}$$

We can see that in both mechanisms trade takes place less frequently than in the first best, but an expected profit maximizing mechanism designer arranges for less trade than an expected welfare maximizing mechanism designer.

3.5 Comments on the Literature

The classic paper on expected revenue maximizing single unit auctions is Myerson (1981). In preparing these notes I have also used Krishna (2002), in particular: Chapter 5, and Milgrom (2004), in particular: Chapter 3. For the public goods problem a classic reference is d'Aspremont and Gerard-Varet (1979). This paper works with a more general set-up than my exposition, but it focuses only on Bayesian incentive compatibility and ex-post budget balance, neglecting individual rationality. Welfare maximization and profit maximization under individual rationality constraints are considered in Güth and Hellwig (1986). The explanation of the relation between ex ante and ex post budget balance is based on Börgers and Norman (2008). For the bilateral trade problem the classic paper is Myerson and Satterthwaite (1983). My exposition of the bilateral trade problem has also benefitted from the books by Krishna and Milgrom cited above.

3.6 Problems

a) Propositions 3.7 and 3.12 show that there are no direct mechanisms that implement the first best decision rule (in the example of the public good problem) or trade rule (in the example of the bilateral trade problem) q^*, and that have moreover three properties: (1) incentive compatibility, (2) individual rationality, (3) ex post budget balance. Prove that in both applications there are direct mechanisms that implement q^* and that have any two of these three properties.

b) Propositions 3.10 and 3.15 display second best mechanisms for numerical examples that illustrate the public goods problem and the bilateral trade problem. However, in each we have not specified the payment rules t_i for $i \in I$. Determine for each case payment rules that guarantee incentive compatibility, individual rationality, and ex post budget balance of the second best mechanism.

c) For each of the numerical examples of this Chapter investigate whether there are intuitive and simple indirect implementations of the expected profit maximizing, or expected welfare maximizing, mechanisms.

Chapter 4

Dominant Strategy Mechanisms: Examples

4.1 Introduction

The models discussed in Chapter 3 were all independent private value models. In Section 3.2.1 we have discussed the limitations of such models. Among the points that we mentioned was that the assumption of a common prior and independent types means that each agent's beliefs about the other agents' types is independent of the agent's type, and moreover that these beliefs are commonly known among the mechanism designer and the agents. This assumption seems peculiar in the context of mechanism design, where the focus is on asymmetric information. In the independent private values model, agents and the mechanism designer don't know other agents' values, but they do know those agent's beliefs. Imperfect information about others' beliefs seems at least as pervasive as imperfect information about others' preferences.

The mechanisms that we constructed in the previous chapter work well if the mechanism designer is right in his assumptions about the agents, and if agents indeed play a Bayesian equilibrium of the game that the mechanism designer creates for them. However, if the mechanism designer has made incorrect assumptions about the agents' beliefs, or if these beliefs are not common knowledge, then the mechanism designer's expectations about how the mechanism will be played might not be fulfilled.

In this chapter we revisit the examples of the previous section, but we make much weaker assumptions about the agents' and the mechanism designer's beliefs. We assume that the mechanism designer does not want to rely on *any* assumption regarding the agents' beliefs about each other. The mechanism designer in this chapter wishes to implement a mechanism of which he is sure that it produces the desired results independent of what agents think about each other. In a sense, we shall move from one extreme along a continuum of possible models to the other extreme, skipping whatever might be in the middle. While in the previous section the mechanism designer was willing to trust a single hypothesis about agents' beliefs, in this section the mechanism designer is assumed to be uncertain about agents' beliefs, and, moreover, it is assumed that the mechanism designer is not at all willing to risk making incorrect assumptions about agents' beliefs.

We shall translate this into formal theory by requiring that the strategy combination that the mechanism designer proposes to the agents when presenting the mechanism to them prescribes choices that are optimal for each type, independent of what the other types do. In short: each type is required to have a dominant strategy.[14] We shall refer to this requirement as "dominant strategy incentive compatibility." This requirement restricts the set of mechanisms and strategy combinations that the mechanism designer can choose from. The available mechanisms and strategy combinations that we consider in this section are a strict subset of those considered in the previous section. This is because, obviously, a vector of strategies that prescribes a dominant strategy to each type always constitutes a Bayesian equilibrium of a game of incomplete information, but not vice versa.

We shall apply a similar logic to the "outside option." We shall assume that whenever the mechanism designer proposes a mechanism it is assumed that in addition to the choices that he offers to the agents in the mechanism the agents can also choose to opt out of the mechanism and receive an outside option utility, such as the utility zero in the auction and public goods models of the previous chapter, or the utility of privately consuming the good in the seller's case in the bilateral trade model of the previous chapter. We will require the strategy that the principal recommends to an agent to dominate the strategy of not participating. We shall call this constraint "ex post individual rationality constraint" because it requires that ex post, after the mechanism has been played, no agent has an incentive to opt out. Recall that in the previous chapter we had interpreted the individual rationality constraint as an "interim individual rationality constraint." In that chapter agents evaluated the outside option once they knew their type using a belief

[14] This use of the phrase "dominant" is slightly sloppy. We shall comment on terminology at the end of this section.

about the other agents' types that was known to the mechanism designer. The mechanism designer had to offer a mechanism that ensured that agents when conducting this evaluation didn't have an incentive to opt out. Ex post individual rationality restricts the set of mechanisms that the mechanism designer can recommend further in comparison to the previous parts of these notes. This is because, obviously, whenever participation dominates non-participation at the ex post level it will also be optimal at the interim level, but not vice versa.

We thus consider in this chapter a mechanism designer whose choice set is smaller than the choice set available to the mechanism designer in the previous part. Our interest will be in characterizing this choice set. We shall also investigate whether the mechanisms that turned out to be optimal in Chapter 3 remain in the choice set. We shall investigate these issues for each of the three models introduced in Chapter 3.

As we shall not attribute explicit beliefs to the mechanism designer in this Chapter, it is not obvious how we should define revenue maximizing or welfare maximizing mechanisms. We shall therefore not study optimal mechanisms in this section, but focus on the characterization of classes of mechanisms that satisfy incentive compatibility, individual rationality, and budget balance requirements. Developing a coherent decision theoretic account of the mechanism designer's view of the agents' payoff types and their beliefs about each other is a project that goes beyond what we do in these notes. [15]

We conclude the introduction with one formal clarification. We shall use the expression "dominance" somewhat loosely in these notes. We shall say that the strategy s_i of player i "dominates" strategy s_i' if $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$ $\forall s_{-i} \in S_{-i}$, and we shall say that a strategy is "dominant" if it dominates all other strategies. In game theoretic terms a better expression than "dominant" would be that strategy s_i is in "always optimal." This is because the term dominance is set apart in game theory for two other concepts: "strict dominance" (where $u_i(s_i, s_{-i}) > u_i(s_i', s_{-i})$ $\forall s_{-i} \in S_{-i}$), and "weak dominance" (where $u_i(s_i, s_{-i}) \geq u_i(s_i', s_{-i})$ $\forall s_{-i} \in S_{-i}$ with $u_i(s_i, s_{-i}) > u_i(s_i', s_{-i})$ $\exists s_{-i} \in S_{-i}$). For the purposes of this note, we shall be sloppy, and define "dominant" to mean "always optimal". In many of the examples that we investigate, our conclusions would remain true if we had defined "dominant" to mean "weakly dominant" in the traditional game-

[15] The decision theoretic foundations of the dominant strategy approach to mechanism design are investigated further in Chung and Ely (2007).

theoretic sense. Our conclusions would, however, *not* remain true if we had defined "dominant" as the traditional game-theoretic concept of "strict dominance."

4.2 Single Unit Auctions

4.2.1 Set-Up

We return to the single unit auction setting of Section 3.2. We briefly recapitulate the aspects of the model that are relevant here. A seller seeks to sell a single indivisible good. There are N potential buyers: $I = \{1, 2, ..., N\}$. Buyer i's type is $\theta_i \in [\underline{\theta}, \bar{\theta}]$, where $0 \leq \underline{\theta} < \bar{\theta}$. We denote by θ the vector $(\theta_1, \theta_2, \ldots, \theta_N)$. θ is an element of $\Theta \equiv [\underline{\theta}, \bar{\theta}]^N$. We denote by θ_{-i} the vector θ when we leave out θ_i. We denote by Θ_{-i} the set $[\underline{\theta}, \bar{\theta}]^{N-1}$, so that $\theta_{-i} \in \Theta_{-i}$. Each buyer i knows her own type θ_i but not θ_{-i}.

We assume that buyer i's utility if she is of type θ_i equals $\theta_i - t_i$ if she obtains the good and pays a transfer $t_i \in \mathbb{R}$ to the seller, and that it is $-t_i$ if she does not obtain the good and pays a transfer t_i to the seller. The seller's utility if he obtains transfers t_i from the N agents is $\sum_{i=1}^{N} t_i$.

4.2.2 Mechanisms, Direct Mechanisms, and the Revelation Principle

Recall that a general mechanism is a game tree together with an assignment of a probability distribution over outcomes, that is a probability distribution over $\{\emptyset, 1, 2, \ldots, N\} \times \mathbb{R}^N$, to each terminal history of the game. We shall imagine, as in the previous Chapter, that the seller proposes a mechanism together with one strategy for each buyer, where a strategy assigns to every type of the buyer a complete behavior plan, possibly randomized, for the game tree.

The incentive compatibility requirement that the seller's proposed strategy combination has to satisfy in this Chapter is more restrictive than it was in the previous Chapter. We now require that each type of each buyer finds it optimal to choose the proposed strategy, for all possible strategy combinations that the other buyers might play. Let us say that in this case the strategy prescribed for each type of each buyer is a "dominant" strategy. We shall also require that each type of each buyer finds that her expected utility, if she follows the seller's recommendation, is non-negative, for each possible realization of the other buyers' types, and for each strategy of the other buyers. We shall say that in this case participation is for each type of each buyer a "dominant" strategy.

As before, we do not formalize the general definition of a mechanism, nor do we formalize the notions of incentive compatibility and of individual rationality that we have just described. The reason is that such a formalization would take long to write down, and we would need it only for a very short while. We wouldn't have to use this formalization much further in these notes because the "revelation principle" extends to our current setting, and we will be able to restrict attention to direct mechanisms. The definition of direct mechanisms is as before, see Definition 3.1. The following result is the revelation principle for our setting.

Proposition 4.1 (Revelation Principle for Dominant Strategy Mechanisms)**.** *Suppose a mechanism* Γ *and a strategy combination* σ *for* Γ *are such that for each type* θ_i *of each buyer* i*, the strategy* $\sigma_i(\theta_i)$ *is a dominant strategy in* Γ*. Then there exists a direct mechanism* Γ' *and a strategy combination* σ' *for* Γ' *such that for every type* θ_i *of each buyer* i *the strategy* $\sigma_i'(\theta_i)$ *is a dominant strategy in* Γ'*, and:*

(i) The strategy vector σ' *satisfies for every* i *and every* θ_i*:*

$$\sigma_i'(\theta_i) = \theta_i,$$

that is, σ' *prescribes telling the truth;*

(ii) For every vector θ *of types, the distribution over allocations and the expected payments that result under* Γ *if the agents play* σ *is the same as the distribution over allocations and the expected payments that result under* Γ' *if the agents play* σ'*.*

Proof. Construct Γ' as required by part (ii) of the proposition. We can prove the result by showing that truth telling will be a dominant strategy in this direct mechanism. To see this, suppose it were not. If type θ_i prefers to report that her type is θ_i' for some type vector of the other agents θ_{-i}, then the same type θ_i would have preferred to deviate from σ_i, and to play the strategy that σ_i prescribes for θ_i' in Γ, for the strategy combination that the types θ_{-i} play in Γ'. Hence σ_i would not be a dominant strategy in Γ. □

Proposition 4.1 shows that for dominant strategy implementation, as for Bayesian equilibrium implementation, we can restrict our attention without loss of generality to direct mechanisms in which truth-telling is a dominant strategy. Moreover, if the indirect mechanism that the mechanism designer contemplates is individually rational, then the same is true for the corresponding direct mechanism. Hence, we can restrict attention to direct mechanisms that are dominant strategy incentive compatible and ex post individually rational as defined below.

Definition 4.1. A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is "dominant strategy incentive compatible" if truth telling is a dominant strategy for each type of each buyer, that is, if for all $i \in I$, all $\theta_i, \theta_i' \in [\underline{\theta}, \bar{\theta}]$ and for all $\theta_{-i} \in \Theta_{-i}$:

$$\theta_i q_i(\theta_i, \theta_{-i}) - t_i(\theta_i, \theta_{-i}) \geq \theta_i q_i(\theta_i', \theta_{-i}) - t_i(\theta_i', \theta_{-i}).$$

Definition 4.2. A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is "ex post individually rational" if for each type of each buyer participation is a dominant strategy, that is, if for all $i \in I$, all $\theta_i \in [\underline{\theta}, \bar{\theta}]$ and all $\theta_{-i} \in \Theta_{-i}$:

$$\theta_i q_i(\theta_i, \theta_{-i}) - t_i(\theta_i, \theta_{-i}) \geq 0.$$

4.2.3 Characterizing Dominant Strategy Incentive Compatibility and Ex Post Individual Rationality

In this subsection we develop a better understanding of the structure of the set of all direct mechanisms that satisfy the two conditions introduced in Definitions 4.1 and 4.2. The characterization of incentive compatible is actually exactly as in Propositions 2.2 and 3.2 except that the result now applies for every realization θ_{-i} of types of the other agents.

Proposition 4.2. *A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is dominant strategy incentive compatible if and only if for every $i \in I$ and every $\theta_{-i} \in \Theta_{-i}$:*

(i)

$$q_i(\theta_i, \theta_{-i}) \text{ is increasing in } \theta_i;$$

(ii) for every $\theta_i \in [\underline{\theta}, \bar{\theta}]$:

$$t_i(\theta_i, \theta_{-i}) = t_i(\underline{\theta}, \theta_{-i}) + (\theta_i q_i(\theta_i, \theta_{-i}) - \underline{\theta} q_i(\underline{\theta}, \theta_{-i})) - \int_{\underline{\theta}}^{\theta} q_i(x, \theta_{-i}) dx$$

We omit the proof of this result as it is the same as the proof of Propositions 2.2, applied to each agent i, and to each possible vector θ_{-i} of the other agents. An interesting and important point is that Proposition 4.2 implies an ex post revenue equivalence result. Whereas our previous revenue equivalence results said that the allocation rule and the interim expected payment of the lowest type imply the interim expected payments for all types, part (ii) of Proposition 4.2 says that the allocation rule and the ex post payments of the lowest types imply the ex post payments of all types.

In analogy to our earlier results, we also obtain a simple characterization of ex post individual rationality for direct mechanisms that are dominant strategy incentive compatible. We omit the proof.

Proposition 4.3. *A dominant strategy incentive compatible direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is ex post individually rational if and only if for every $i \in I$ and every $\theta_{-i} \in \Theta_{-i}$ we have:*

$$t_i(\underline{\theta}, \theta_{-i}) \leq \underline{\theta} q_i(\underline{\theta}, \theta_{-i}).$$

4.2.4 Canonical Auctions

We now display a class of dominant strategy incentive compatible and ex post individually rational direct mechanisms. We call the direct mechanisms in this class "canonical auctions."[16] We don't claim that canonical auctions are the only dominant strategy incentive compatible and ex post individually rational direct mechanisms, but we show that this class is rich enough to include the mechanisms that we identified in Sections 3.2.4 and 3.2.5 as expected revenue maximizing and as expected welfare maximizing.

Recall from the discussion of the expected revenue maximizing auction in Section 3.2.4 the definition of the functions $\psi_i : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$ for $i \in I$:

$$\psi_i(\theta_i) \equiv \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \text{ for all } \theta_i \in [\underline{\theta}, \bar{\theta}]. \tag{4.1}$$

The regularity assumption under which we derived the expected revenue maximizing auction was that all functions ψ_i were increasing. Our interpretation of these functions was that they assigned to each type a "virtual type." Under the regularity assumption, the expected revenue maximizing auction assigned the object to the buyer with the highest virtual type.

Instead of working with these specific function ψ_i, we now consider arbitrary strictly increasing functions ψ_i. We show that for any such functions, an allocation rule that is constructed as in the expected revenue maximizing auction in Section 3.2.4 can be supplemented with transfer rules that

[16]This expression is not commonly used. I made it up for these notes.

make the mechanism dominant strategy incentive compatible and ex post individually rational. The advantage of generalizing the result in this way is that we can use it to also show the implementability of allocation rules other than the expected revenue maximizing. To simplify the exposition, we assume that the functions ψ_i are continuous.

Definition 4.3. A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is called a "canonical auction" if there are strictly increasing and continuous functions $\psi_i : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$ for $i \in I$ such that for all $\theta \in \Theta$ and $i \in I$:

$$q_i(\theta) = \begin{cases} \frac{1}{n} & \text{if } \psi_i(\theta_i) \geq 0 \text{ and } \psi_i(\theta_i) \geq \psi_j(\theta_j) \text{ for all } j \in I \text{ with } j \neq i \\ \\ 0 & \text{otherwise,} \end{cases}$$

where n is the number of agents $k \in I$ such that $\psi_k(\theta_k) = \psi_i(\theta_i)$, and:

$$t_i(\theta) = \begin{cases} \frac{1}{n} \min\{\widetilde{\theta}_i \in [\underline{\theta}, \bar{\theta}] \mid q(\widetilde{\theta}_i, \theta_{-i}) = 1/n\} & \text{if } q_i(\theta) > 0 \\ \\ 0 & \text{if } q_i(\theta) = 0, \end{cases}$$

for all $\theta \in \Theta$.

It is worth considering the transfer rule in detail. If bidder i does not win the auction, then bidder i does not have to pay anything. If bidder i does win the auction,[17] then bidder i's payment equals the lowest type that she might have had that would have allowed her to win the auction. The assumed continuity of ψ_i guarantees that this minimum exists.

Proposition 4.4. *Every canonical auction $(q, t_1, t_2, \ldots, t_N)$ is dominant strategy incentive compatible and ex post individually rational. Moreover, for every $i \in I$: $u_i(\underline{\theta}, \theta_{-i}) = 0$ for all $\theta_{-i} \in \Theta_{-i}$.*

Proof. We first show dominant strategy incentive compatibility. Suppose that buyer i is of type θ_i, that the other buyers have types θ_{-i}, and that $q_i(\theta_i, \theta_{-i}) = 0$. Does buyer i have an incentive to report a different type θ_i'? If $q_i(\theta_i', \theta_{-i}) = 0$, then her utility doesn't change. If $q_i(\theta_i', \theta_{-i}) = \frac{1}{n} > 0$, it will have to be the case that $\theta_i' > \theta_i$. Moreover, buyer i's payment will be larger than $\frac{1}{n}\theta_i$, as her payment will be the lowest type of buyer i that wins against θ_{-i}, and by assumption θ_i is not large enough. Thus, buyer i can win the auction, but only by paying more than the object is worth to her. Thus, she has no incentive to change her strategy.

[17] For simplicity, we ignore ties.

Consider next the case that $q_i(\theta_i, \theta_{-i}) = 1$. If buyer i changes her report to another type θ_i' for which $q_i(\theta_i', \theta_{-i}) = 1$, then her utility doesn't change, as her payment does not depend on her report. If she changes her report to a type θ_i' for which $q_i(\theta_i', \theta_{-i}) = 0$, her utility decreases. This is because if she reports θ_i truthfully, her transfer payment, by the definition of the transfer rule, will be below θ_i. Hence a truthful report gives her a positive surplus.

Consider finally the case that $q_i(\theta_i, \theta_{-i}) = \frac{1}{n}$ where $n \geq 2$. Then buyer i's transfer payment is $\frac{1}{n}\theta_i$, and her expected utility will be zero. If buyer i changes her report to another type $\theta_i' > \theta_i$ so that $q_i(\theta_i', \theta_{-i}) = 1$, then her expected utility doesn't change, as her payment will be θ_i. If she changes her report to a type θ_i' for which $q_i(\theta_i', \theta_{-i}) = 0$, her expected utility is again zero. Hence, a truthful report is optimal. This concludes the proof of dominant strategy incentive compatibility.

The proof of dominant strategy incentive compatibility also shows that a buyer's utility is always non-negative if she wins the auction. If she loses the auction, her utility is zero. Therefore, the mechanism also satisfies ex post individual rationality. The lowest type, $\underline{\theta}$, either loses the auction, and has utility zero, or wins the auction and has to pay $\underline{\theta}$, in which case utility is also zero. This proves the last sentence of the Proposition. □

We now provide two applications of Proposition 4.4. Firstly, suppose that the seller had subjective beliefs F on Θ that reflected his view regarding the likelihood of different type vectors. Suppose moreover that these beliefs satisfied the assumptions of Section 3.2.1 as well as the regularity Assumption 3.1. Then one can interpret the optimal auction identified in Section 3.2.4 as the expected revenue maximizing auction if the mechanism designer assumes that his belief about Θ is also the prior distribution from which the buyers' beliefs about the other buyers' valuations are derived. This is a very strong assumption about the seller's view of his environment. In this part we have relaxed this assumption, and have assumed that the seller is highly uncertain about the buyers' beliefs about each other. Proposition 4.4 shows that even if the seller's uncertainty is so large that he wishes to use a dominant strategy mechanism, he does not lose any expected revenue. He can find a transfer rule that implements the optimal allocation rule, and that gives the lowest types expected utility zero. Thus, it is one of the optimal auctions of Section 3.2.4. In other words, we have shown that among the optimal auctions that we found in Section 3.2.4, there are some that make truth telling a dominant strategy.

A second application arises from Proposition 4.4 if we set $\psi_i(\theta_i) = \theta_i$ for every $i \in I$ and every $\theta_i \in [\underline{\theta}, \bar{\theta}]$. In this case the auction described in Proposition 4.4 is just the Vickrey auction, and it is thus welfare maximizing. Thus, the mechanism designer can also achieve the objective of expected welfare maximization through a dominant strategy mechanism. Of course, Proposition 4.4 describes many other dominant strategy incentive compatible and ex post individually rational auctions.

4.3 Public Goods

4.3.1 Set-Up

We shall now show that in the two examples with budget constraints that we considered in Sections 3.3 and 3.4, the restriction to dominant strategy incentive compatible mechanisms may severely limit what the mechanism designer can achieve. We begin by briefly recapitulating the public goods model.

We consider a community consisting of N agents: $I = \{1, 2, \ldots, N\}$. They have to choose whether to produce some indivisible, non-excludable public good. We denote this decision by $g \in \{0, 1\}$. If the public good is produced, then $g = 1$. If it is not produced, then $g = 0$. Agent i's utility if the collective decision is g and if she pays a transfer t_i to the community is: $\theta_i g - t_i$. The cost of producing the public good is cg, where $c > 0$. Here, θ_i is agent i's type, and it is an element of $\theta_i \in [\underline{\theta}, \bar{\theta}]$ where $0 \leq \underline{\theta} < \bar{\theta}$. We denote by θ the vector $(\theta_1, \theta_2, \ldots, \theta_N)$. θ is an element of $\Theta \equiv [\underline{\theta}, \bar{\theta}]^N$. We denote by θ_{-i} the vector θ when we leave out θ_i. We denote by Θ_{-i} the set $[\underline{\theta}, \bar{\theta}]^{N-1}$, so that $\theta_{-i} \in \Theta_{-i}$. Each agent i observes her own type θ_i, but not necessarily the other agents' types θ_{-i}. The mechanism designer knows none of the θ_i's. The mechanism designer seeks to maximize expected welfare, where welfare is defined to be $\left(\sum_{i=1}^N \theta_i\right) g - \sum_{i=1}^N t_i$.

4.3.2 Direct Mechanisms

By the revelation principle we can restrict attention to direct mechanisms. We shall also, as in Section 3.3, restrict attention to deterministic mechanisms, that is, to mechanisms where the decision about the public good and also the agents' transfers are not stochastic once the agents' true types are known. We introduced this assumption in Section 3.3 because it simplified the formulation of the budget constraint and because it appeared innocuous.

In the current section, it is not clear that the assumption is innocuous. However, for simplicity, we continue to work with this assumption. Thus, direct mechanisms are defined as in Definition 3.4. The definitions of dominant strategy incentive compatibility and ex post individual rationality are as in the previous subsection. Moreover, we shall impose ex post budget balance, which is defined as in Definition 3.5. Observe that ex ante budget balance has no meaning in our current context, because there is no prior probability measure over Θ.

Note that in Definition 3.5 the budget constraint is an equality rather than an inequality. An inequality would require that the sum of the contributions is at least as large as the costs of the public good's production rather than that it is exactly equal. In the previous chapter, the distinction between an equality budget constraint and an inequality budget constraint was not of much further importance, whereas in the current chapter the equality requirement is an important restriction. We might motivate it by arguing that we consider an environment in which the agents participating in the public goods mechanism have no way of committing to a way of disposing of excess funds if the collected transfers exceed what is required for the public good production. But our main reason for restricting attention to an equality budget constraint is that this simplifies our arguments below.

4.3.3 Characterizing Dominant Strategy Incentive Compatibility and Ex Post Individual Rationality

In this subsection we neglect the budget constraint, and seek a characterization of dominant strategy incentive compatible mechanisms, and a condition under which such mechanisms are in addition ex post individually rational. We could use the same arguments for this as we used in the previous section. However, a simpler proof can be given if one restricts attention to deterministic mechanisms, as we have done here.

Proposition 4.5. *A direct mechanism is dominant strategy incentive compatible if and only for every $i \in I$ and for every $\theta_{-i} \in \Theta_{-i}$, there are a type*[18] *$\hat{\theta}_i \in \mathbb{R}$ and two payments τ_i and $\hat{\tau}_i \in \mathbb{R}$ such that:*

$$\theta_i < \hat{\theta}_i \Rightarrow q(\theta_i, \theta_{-i}) = 0 \text{ and } t_i(\theta_i, \theta_{-i}) = \tau_i;$$
$$\theta_i > \hat{\theta}_i \Rightarrow q(\theta_i, \theta_{-i}) = 1 \text{ and } t_i(\theta_i, \theta_{-i}) = \hat{\tau}_i;$$

[18] The word "type" is used in a loose sense here, as the value is not necessarily within the type range.

$$\theta_i = \hat{\theta}_i \Rightarrow q(\theta_i, \theta_{-i}) = 0 \text{ and } t_i(\theta_i, \theta_{-i}) = \tau_i \text{ or}$$
$$q(\theta_i, \theta_{-i}) = 1 \text{ and } t_i(\theta_i, \theta_{-i}) = \hat{\tau}_i;$$
$$\hat{\tau}_i - \tau_i = \hat{\theta}_i.$$

Note that in this Proposition, $\hat{\theta}_i$, τ_i and $\hat{\tau}_i$ are allowed to depend on θ_{-i}. Observe also that we have not ruled out that $\hat{\theta}_i < \underline{\theta}$ (in which case we would have $q_i(\theta_i, \theta_{-i}) = 1$ for all $\theta_i \in [\underline{\theta}, \bar{\theta}]$), nor have we ruled out $\hat{\theta}_i > \bar{\theta}$ (in which case we would have $q_i(\theta_i, \theta_{-i}) = 0$ for all $\theta_i \in [\underline{\theta}, \bar{\theta}]$).

Proof. First we show sufficiency. Consider an agent i with type θ_i. This agent can either report a type $\theta_i' > \hat{\theta}_i$, in which case the public good is produced and the agent has to pay $\hat{\theta}_i$, or the agent can report a type $\theta_i' < \hat{\theta}_i$, in which case the public good is not produced, and the agent does not have to pay anything. For the outcome, it only matters whether the reported type is above or below $\hat{\theta}_i$. Beyond that, the value of the reported type does not matter. If in truth $\theta_i \geq \hat{\theta}_i$, then reporting a type above $\hat{\theta}_i$, such as the true type, is obviously optimal, because the agent obtains more than he has to pay. A symmetric argument applies if $\theta_i \leq \hat{\theta}_i$. This proves that the conditions provided are sufficient for dominant strategy incentive compatibility.

For the converse, fix $\theta_{-i} \in \Theta_{-i}$. Consider all θ_i such that $q_i(\theta_i, \theta_{-i}) = 0$. Observe that $t_i(\theta_i, \theta_{-i})$ has to be the same for all these θ_i, because otherwise buyer i would pretend to be the type for which the transfer payment is lowest. By the same argument the transfer payment $t_i(\theta_i, \theta_{-i})$ has to be the same for all θ_i for which $q_i(\theta_i, \theta_{-i}) = 0$.

Consider first the case that $q_i(\theta_i, \theta_{-i}) = 0$ for all $\theta_i \in [\underline{\theta}, \bar{\theta}]$. Denote by τ_i the (constant) payment of agent i. If we set $\hat{\theta}_i > \bar{\theta}$ and $\tau_i = \tau_i + \hat{\theta}_i$, all conditions of Proposition 4.5 hold. The case that $q_i(\theta_i, \theta_{-i}) = 1$ for all $\theta_i \in [\underline{\theta}, \bar{\theta}]$ can be dealt with symmetrically.

Now suppose that $q_i(\theta_i, \theta_{-i}) = 0$ for some θ_i and $q_i(\theta_i, \theta_{-i}) = 1$ for some other θ_i. Denote the payments corresponding to the former case by τ_i, and denote agent i's payment in the latter case by $\hat{\tau}_i$. Define $\hat{\theta}_i = \hat{\tau}_i - \tau_i$. Then types $\theta_i > \hat{\theta}_i$ will report a type such that the public good is produced, and types $\theta_i < \hat{\theta}_i$ will report a type such that the public good is not produced. Truthful reporting is therefore optimal only if the conditions of Proposition 4.5 hold. □

Next we characterize ex post individual rationality. We need not give any proof of the following simple result.

Proposition 4.6. *A dominant strategy incentive compatible direct mechanism is ex post individually rational if and only for every* $i \in I$ *and for every* $\theta_{-i} \in \Theta_{-i}$*:*

$$t_i(\underline{\theta}, \theta_{-i}) \leq \underline{\theta} q(\underline{\theta}, \theta_{-i}).$$

4.3.4 Canonical Mechanisms

We now introduce a class of mechanisms, "canonical mechanisms," that are closely related to the mechanisms that we identified in Section 3.3.4 as welfare maximizing or profit maximizing mechanisms under the budget constraint. The definition of these mechanisms involved functions ψ_i, one for each agent $i \in I$. The regularity assumption was that these functions were strictly increasing. Here, we let ψ_i be some arbitrary strictly increasing function. To simplify the exposition we assume that the functions ψ_i are continuous. The decision rules in the mechanisms defined below is of the same form as the decision rule in the expected welfare or profit maximizing mechanisms in part 1. We show below that we can combine these decision rules with specific transfer rules to make them dominant strategy incentive compatible and ex post individually rational. However, in general, the mechanisms that we describe below will not balance the budget. In Section 3.3.4, by contrast, we obtained mechanisms that were Bayesian equilibrium incentive compatible, individually rational, and moreover the rules could be made ex post budget balanced.

Definition 4.4. A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is called "canonical" if for every agent i there is a strictly increasing and continuous function $\psi_i : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$ such that:

$$q(\theta) = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} \psi_i(\theta_i) \geq c, \\ \\ 0 & \text{otherwise,} \end{cases}$$

for all $\theta \in \Theta$ and, for every $i \in I$:

$$t_i(\theta) = \begin{cases} \min\{\widetilde{\theta}_i \in [\underline{\theta}, \bar{\theta}] \mid \psi_i(\widetilde{\theta}_i) + \sum_{j \neq i} \psi_j(\theta_j) \geq c\} & \text{if } q(\theta) = 1 \\ \\ 0 & \text{if } q(\theta) = 0, \end{cases}$$

for all $\theta \in \Theta$.

We now state the following simple result.

Proposition 4.7. *Every canonical mechanism* $(q, t_1, t_2, \ldots, t_N)$ *is dominant strategy incentive compatible and ex post individually rational. Moreover, for every* $i \in I$*:* $u_i(\underline{\theta}, \theta_{-i}) = 0$ *for all* $\theta_{-i} \in \Theta_{-i}$.

The proof of this result is left to the reader. Note that the mechanisms obtained in Propositions 3.8 and 3.9 include canonical mechanisms.

4.3.5 Ex Post Budget Balance

Now we ask which direct mechanisms are dominant strategy incentive compatible, ex post individually rational, and satisfy ex post budget balance. In these notes, we only provide a partial answer to this question. To my understanding, there is no paper in the literature that answers this question at a general level. The answer that we give in these notes is for the case that there are only 2 agents. For this case, it is easy to obtain the following characterization. To simplify the exposition, we assume in this characterization that the set of type vectors for which the public good is produced is a closed set.

Proposition 4.8. *Suppose* $N = 2$*. Suppose also that the set* $\{\theta \mid q(\theta) = 1\}$ *is closed. Then a direct mechanism is dominant strategy incentive compatible, ex post individually rational, and ex post budget balanced if and only if there are payments* $\tau_1, \tau_2 \in \mathbb{R}$ *with* $\tau_1 + \tau_2 = c$ *such that:*

$$q(\theta) = 1 \text{ and } t_i(\theta) = \tau_i \text{ for all } i \in \{1, 2\} \text{ if } \theta_1 \geq \tau_1 \text{ and } \theta_2 \geq \tau_2$$
$$q(\theta) = 0 \text{ and } t_i(\theta) = 0 \text{ for all } i \in \{1, 2\} \text{ otherwise.}$$

A simple indirect implementation of the mechanism described in Proposition 4.8 is this: Each of the two agents is allocated a share of the cost τ_i, so that $\tau_1 + \tau_2 = c$. Then each agent i is asked whether he or she is willing to contribute τ_i whereby the contribution has to be made only if the other agent also pledges his or her contribution. If both agents make a pledge, then the public good is produced, and each agent i pays τ_i. Otherwise, it is not produced and no agent pays anything. Note that, conditional on producing the public good, agents' contributions do not depend on their own or on the other agent's valuation.

Proof. It is easy to show that the mechanisms displayed in Proposition 4.8 are dominant strategy incentive compatible, ex post individually rational, and ex post budget balanced. We therefore only show the sufficiency of these conditions.

Suppose $q(\theta) = 0$ for all $\theta \in \Theta$. Then obviously the mechanism can be dominant strategy incentive compatible, ex post individually rational, and ex post budget balanced only if all payments are zero. We can describe this mechanism in the form described in Proposition 4.8 by setting $\tau_i > \bar{\theta}$ for some agent i, and defining $\tau_j \equiv c - \tau_i$ for $j \neq i$. Note that τ_j may be negative. This is not ruled out by the Proposition.

Let us now assume that $q(\theta) = 1$ for at least one $\theta \in \Theta$. For $i = 1, 2$ define $\hat{\theta}_i \equiv \min\{\theta_i \in [\underline{\theta}, \bar{\theta}] \mid q(\theta_i, \theta_j) = 1 \text{ for some } \theta_j \in [\underline{\theta}, \bar{\theta}]\}$ where $j \neq i$. Our assumption that the set $\{\theta \mid q(\theta) = 1\}$ is closed guarantees that these minima, and similar minima referred to below, are well defined. We seek to show that $\theta_i \geq \hat{\theta}_i$ implies $q(\theta_i, \hat{\theta}_j) = 1$. To show this we define $\widetilde{\theta}_1 \equiv \min\{\theta_1 \in [\underline{\theta}, \bar{\theta}] \mid q(\theta_1, \hat{\theta}_2) = 1\}$ and $\widetilde{\theta}_2 \equiv \min\{\theta_2 \in [\underline{\theta}, \bar{\theta}] \mid q(\hat{\theta}_1, \theta_2) = 1\}$. By definition $\widetilde{\theta}_i \geq \hat{\theta}_i$ for $i \in \{1, 2\}$. We claim that $\widetilde{\theta}_i = \hat{\theta}_i$ for $i \in \{1, 2\}$. Together with Proposition 4.5 this implies what we want to show.

Suppose that $\widetilde{\theta}_1 > \hat{\theta}_1$. By Proposition 4.5 $t_1(\widetilde{\theta}_1, \hat{\theta}_2) = \widetilde{\theta}_1$ and $t_2(\widetilde{\theta}_1, \hat{\theta}_2) = \hat{\theta}_2$. Budget balance requires $\widetilde{\theta}_1 + \hat{\theta}_2 = c$. By Proposition 4.5 we have that $q(\widetilde{\theta}_1, \widetilde{\theta}_2) = 1$, $t_1(\widetilde{\theta}_1, \widetilde{\theta}_2) = \hat{\theta}_1$ and $t_2(\widetilde{\theta}_1, \widetilde{\theta}_2) = \hat{\theta}_2$. Thus the sum of contributions if types are $(\widetilde{\theta}_1, \widetilde{\theta}_2)$ is $\hat{\theta}_1 + \hat{\theta}_2$. If $\widetilde{\theta}_1 < \hat{\theta}_1$, then the sum of these contributions is strictly less than $\widetilde{\theta}_1 + \hat{\theta}_2$, of which we had just shown that it equals c. Therefore, budget balance is violated. (The same reasoning applies if $\widetilde{\theta}_2 > \hat{\theta}_2$.) We conclude that $\widetilde{\theta}_i = \hat{\theta}_i$ for $i \in \{1, 2\}$. Hence $\theta_i \geq \hat{\theta}_i$ implies $q(\theta_i, \hat{\theta}_j) = 1$, where $i \neq j$.

From what we have shown in the previous paragraph we can infer that $\theta_i \geq \hat{\theta}_i$ for $i = 1, 2$ implies $q(\theta_1, \theta_2) = 1$. The reason is that first, as shown above, we can infer $q(\theta_1, \hat{\theta}_2) = 1$, and then, using Proposition 4.5, $q(\theta_1, \theta_2) = 1$. Also note that $\theta_1 < \hat{\theta}_1$ or $\theta_2 < \hat{\theta}_2$ implies $q(\theta) = 0$. This is a consequence of the definitions of $\hat{\theta}_1$ and $\hat{\theta}_2$.

Now suppose that $\hat{\theta}_i > 0$ for $i = 1, 2$. Then, by Proposition 4.5 the payment of each agent i must be $\hat{\theta}_i$ whenever the public good is produced. If we define $\tau_i \equiv \theta_i$, the mechanism is of the form described in the Proposition.

The remaining case is that $\hat{\theta}_i = 0$ for both agents i. This is the case when the public good is produced with probability 1. Incentive compatibility requires that no agent's payment depends on their own report. In principle, it may depend on the other agents' reports, though. We write agent 1's payment as $\tau_1(\theta_2)$, and agent 2's payment as $\tau_2(\theta_1)$. We want to show that τ_1 does not depend on θ_2, i.e. that $\tau_1(\theta_2) = \tau_1(\theta_2')$ for all $\theta_2, \theta_2' \in [\underline{\theta}, \bar{\theta}]$. Suppose $\tau_1(\theta_2) < \tau_1(\theta_2')$ for some $\theta_2, \theta_2' \in [\underline{\theta}, \bar{\theta}]$. Fix any θ_1. Then, if agent 2's type is θ_2, the sum of contributions is: $\tau_1(\theta_2) + \tau_2(\theta_1)$. If agent 2's type is θ_2', the sum of contributions is: $\tau_1(\theta_2') + \tau_2(\theta_1)$. Only one, but not both of these sums can equal c, thus contradicting budget balance. Therefore, $\tau_1(\theta_2)$ is a constant. We write τ_1 for this constant. Budget balance then requires τ_2 to be constant too. Again we find that the mechanism is of the form described in the Proposition. □

4.4 Bilateral Trade

4.4.1 Set-Up

We recapitulate briefly the set-up from Section 3.4. A seller S owns a single indivisible good. There is one potential buyer B. The seller's utility if he sells the good and receives a transfer payment t_S is equal to t_S. If he does not sell the good and receives a transfer t_S then his utility is $\theta_S + t_S$, where $\theta_S \in [\underline{\theta}_S, \bar{\theta}_S]$ is the seller's type. The buyer's utility if he purchases the good and pays a transfer t_B equals $\theta_B - t_B$, where $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$ is the buyer's privately observed type. The buyer's utility if he does not obtain the good and pays transfer t_B is $-t_B$. We define $\theta \equiv (\theta_S, \theta_B)$. Each agent knows his own type but not the other agent's type. We consider the situation from a mechanism designer's perspective who does not know the values of θ_S and θ_B.

4.4.2 Dominant Strategy Incentive Compatible and Ex Post Individually Rational Direct Mechanisms

By the revelation principle we restrict ourselves to direct mechanisms with deterministic trading rules, as defined as in Definition 3.9. We require these to satisfy dominant strategy incentive compatibility, ex post individual rationality, and ex post budget balance.

To characterize dominant strategy incentive compatibility we can proceed as in the previous section. The problem is analogous to that of the previous section because we have restricted attention to deterministic mechanisms. For briefness, we state the result without proof.

Proposition 4.9. *A direct mechanism is dominant strategy incentive compatible if and only if for every $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$ there exist a type $\hat{\theta}_S \in \mathbb{R}$ and payments $\tau_S, \hat{\tau}_S \in \mathbb{R}$ such that:*

$$\theta_S < \hat{\theta}_S \Rightarrow q(\theta) = 1 \text{ and } t_S(\theta) = \hat{\tau}_S;$$
$$\theta_S > \hat{\theta}_S \Rightarrow q(\theta) = 0 \text{ and } t_S(\theta) = \tau_S;$$

$$\begin{aligned}
\theta_S = \hat{\theta}_S &\Rightarrow q(\theta) = 0 \textit{ and } t_S(\theta) = \tau_S \textit{ or} \\
&\qquad q(\theta) = 1 \textit{ and } t_S(\theta) = \hat{\tau}_S; \\
\hat{\tau}_S - \tau_S &= \hat{\theta}_S;
\end{aligned}$$

and for every $\theta_S \in [\underline{\theta}_S, \bar{\theta}_S]$ *there exist a type* $\hat{\theta}_B \in \mathbb{R}$ *and payments* $\tau_B, \hat{\tau}_B \in \mathbb{R}$ *such that:*

$$\begin{aligned}
\theta_B < \hat{\theta}_B &\Rightarrow q(\theta) = 0 \textit{ and } t_B(\theta) = \tau_B; \\
\theta_B > \hat{\theta}_B &\Rightarrow q(\theta) = 1 \textit{ and } t_B(\theta) = \hat{\tau}_B; \\
\theta_B = \hat{\theta}_B &\Rightarrow q(\theta) = 0 \textit{ and } t_B(\theta) = \tau_B \textit{ or} \\
&\qquad q(\theta) = 1 \textit{ and } t_B(\theta) = \hat{\tau}_B; \\
\hat{\tau}_B - \tau_B &= \hat{\theta}_B.
\end{aligned}$$

We continue with a standard characterization of individual rationality the proof of which we also omit.

Proposition 4.10. *A dominant strategy incentive compatible direct mechanism is ex post individually rational if and only if for every* $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$:

$$t_S(\bar{\theta}_S, \theta_B) \geq \bar{\theta}_S q(\bar{\theta}_S, \theta_B)$$

and for every $\theta_S \in [\underline{\theta}_S, \bar{\theta}_S]$:

$$t_B(\theta_S, \underline{\theta}_B) \leq \underline{\theta}_B q(\theta_S, \underline{\theta}_B).$$

4.4.3 Canonical Mechanisms

We introduce again a class of mechanisms that encompasses some of those that we introduced earlier as expected welfare or expected profit maximizing (under a budget constraint), and that are dominant strategy incentive compatible and ex post individually rational.

Definition 4.5. A direct mechanism (q, t_S, t_B) is called "canonical" if for every agent $i \in \{S, B\}$ there is a strictly increasing and continuous function $\psi_i : [\underline{\theta}_i, \bar{\theta}_i] \to \mathbb{R}$ such that:

$$q(\theta) = \begin{cases} 1 & \text{if } \psi_B(\theta_B) \geq \psi_S(\theta_S), \\ \\ 0 & \text{otherwise,} \end{cases}$$

and

$$t_S(\theta) = \begin{cases} \max\{\widetilde{\theta}_S \in [\underline{\theta}_S, \bar{\theta}_S] \mid \psi_B(\theta_B) \geq \psi_S(\widetilde{\theta}_S)\} & \text{if } q(\theta) = 1 \\ 0 & \text{if } q(\theta) = 0, \end{cases}$$

and

$$t_B(\theta) = \begin{cases} \min\{\widetilde{\theta}_B \in [\underline{\theta}_B, \bar{\theta}_B] \mid \psi_B(\theta_B) \geq \psi_S(\widetilde{\theta}_S)\} & \text{if } q(\theta) = 1 \\ 0 & \text{if } q(\theta) = 0, \end{cases}$$

Thus, if trade takes place, the seller receives the largest value that he could have had and trade would still have taken place. Similarly, the buyer receives the lowest value that he could have had and trade would still have taken place.

Proposition 4.11. *Every canonical mechanism is dominant strategy incentive compatible and ex post individually rational. Moreover,* $u_S(\bar{\theta}_S, \theta_B) = \bar{\theta}_S$ *for every* $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$ *and* $u_B(\theta_S, \underline{\theta}_B) = 0$ *for every* $\theta_S \in [\underline{\theta}_S, \bar{\theta}_S]$*.*

The proof of this result is analogous to the proof of Proposition 4.7, and we omit it.

4.4.4 Ex Post Budget Balance

Now we characterize direct mechanisms that are dominant strategy incentive compatible, ex post individually rational, and satisfy ex post budget balance. The structure of our result is similar to the structure of Proposition 4.8.

Proposition 4.12. *Suppose that the set* $\{\theta \mid q(\theta) = 1\}$ *is closed. A direct mechanism is dominant strategy incentive compatible, ex post individually rational, and ex post budget balanced if and only if either:*

$$q(\theta) = 0 \text{ and } t_i(\theta) = 0 \qquad \text{for all } i \in \{S, B\} \text{ and all } \theta \in \Theta,$$

or there is a $\hat{\theta}$ *such that:*

$$\begin{aligned} &q(\theta) = 1 \text{ and } t_i(\theta) = \hat{\theta} \text{ for all } i \in \{S, B\} \quad \text{if} \quad \theta_S \leq \hat{\theta} \text{ and } \theta_B \geq \hat{\theta}; \\ &q(\theta) = 0 \text{ and } t_i(\theta) = 0 \text{ for all } i \in \{S, B\} \quad \text{if} \quad \theta_S > \hat{\theta} \text{ or } \theta_B < \hat{\theta}. \end{aligned}$$

The proof of this result is analogous to the proof of Proposition 4.8, and we omit it. The class of mechanisms described in Proposition 4.12 can best be understood as fixed price mechanisms. There is a fixed price $\hat{\theta}$, and trade takes place at this price if and only if the buyer reports a value above this price and the seller reports a value below this price. This mechanism is restrictive because the price does not depend at all on the agents' reported valuations.

4.5 Comments on the Literature

The results on dominant strategy mechanisms in the single unit auction setting are special cases of a more general result in Mookherjee and Reichelstein (1992) who investigate conditions under which one can construct for any Bayesian incentive compatible mechanisms an equivalent dominant strategy mechanism. I have also used Section 5.2 of Vijay Krishna's *Auction theory* for this section of the notes. The section on public goods mechanisms is partially based on discussions with Arunava Sen from the Indian Statistical Institute, New Delhi. Dominant strategy mechanisms for bilateral trade are discussed in Hagerty and Rogerson (1987).

4.6 Problems

a) Assume that stochastic decision rules are allowed in the public goods problem, and state a characterization of dominant strategy incentive compatible direct mechanisms that is analogous to Proposition 4.2. (You don't have to prove this characterization.) Show that Proposition 4.5 is a special case of the more general result that you have stated.

Chapter 5

Dominant Strategy Mechanisms: General Theory

5.1 Introduction

In the first three chapters of these notes we have considered examples that illustrate the general theory of mechanism design. In this and the next Chapter we are now going to consider the extent to which we can generalize what we have seen in these examples. We shall investigate which of our earlier results continue to be true in the more general model. We also introduce some new results. We begin by considering dominant strategy mechanisms. The reason for starting with dominant strategy mechanisms is that these are a subset of the Bayesian incentive compatible mechanisms. Therefore, some of the results in this chapter will also be useful in the next chapter.

5.2 Set-Up

There are N agents. The set of agents is denoted by $I = \{1, 2, \ldots, N\}$. They have to choose an alternative a out of some set A of mutually exclusive alternatives. Agent i's utility if alternative a is chosen, and agent i pays transfer t_i is:

$$u_i(a, \theta_i) - t_i.$$

Here, θ_i is agent i's type, and t_i is agent i's transfer. Note that the set-up does include the case in which A is the set of lotteries over some set of non-random alternatives, and u_i is an expected utility function.

We shall employ similar notation as before: The set of possible types of agent i is Θ_i, which we now take to be some abstract set. We denote by θ the vector of types: $(\theta_1, \theta_2, \ldots, \theta_N)$. The set of all possible type vectors is $\Theta \equiv \Theta_1 \times \Theta_2 \times \ldots \times \Theta_N$. Finally, we write θ_{-i} for a vector θ of types if we leave out agent i's type. The set of all θ_{-i} is Θ_{-i}, which is the cartesian product of the sets Θ_j such that $j \neq i$.

Our interest is in dominant strategy incentive compatible mechanisms. The revelation principle holds in this setting, and we can restrict attention to direct mechanisms.

Definition 5.1. A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ consists of a mapping

$$q : \Theta \rightarrow A,$$

that maps every type vector into a collective decision, and mappings

$$t_i : \Theta \rightarrow \mathbb{R},$$

one for each player $i \in I$, that indicate for each type vector the transfer that agent i needs to make.

We call q the "decision rule." Note that we restrict ourselves to mechanisms where the payment is deterministic. If we assume that agents are risk-neutral, then this is without loss of generality, as long as we are only concerned with characterizations of incentive compatibility and individual rationality. When we consider budget balance, this point is more delicate. For the moment, however, we are only concerned with incentive compatibility.

Definition 5.2. A direct mechanism is "dominant strategy incentive compatible" if for all $\theta \in \Theta$, all $i \in I$ and all $\theta_i' \in \Theta_i'$ we have:

$$u_i(q(\theta), \theta_i) - t_i(\theta) \geq u_i(q(\theta_i', \theta_{-i}), \theta_i) - t_i(\theta_i', \theta_{-i}).$$

We introduce other requirements, such as individual rationality and budget balance, later when needed. The next sections will only be concerned with incentive compatibility.

5.3 Ex Post Revenue Equivalence

We begin by providing a result that is analogous to the ex post revenue equivalence results that we obtained in Chapter 4, such as Proposition 4.2. The proofs of those results used the fact that in all the examples the type spaces were connected sets. With finite type spaces, these results are not true. Therefore, in the abstract setting in this part of the lecture notes, we need to make an additional assumption regarding the type spaces Θ_i if we want to obtain a payoff equivalence result. The assumption that we shall make below is that for every $i \in I$ the set Θ_i is a convex subset of some finite dimensional Euclidean space.

An assumption for the type spaces alone is, however, not enough. We shall also make use of an assumption about how the agents' utility functions depend on their types. We shall assume that every agent i's utility function is a convex function of the agent's own type. This assumption will ensure that an agent's utility under an incentive compatible mechanism depends sufficiently smoothly on their type that we can employ an envelope theorem. Convexity of the utility function in a player's own type is satisfied, for example, if the space of alternatives A is the set of all lotteries over a finite outcome space, and if the type of a player is just the vector of this player's Bernoulli utilities of the outcomes. In this case, the player's utility is a linear function of the type vector, and hence convex.

Proposition 5.1. *Suppose that for every $i \in I$ the set Θ_i is a convex subset of a finite dimensional Euclidean space. Moreover, assume that for every $i \in I$ the function $u_i(a, \theta_i)$ is a convex function of θ_i. Suppose that $(q, t_1, t_2, \ldots, t_N)$ is a dominant strategy incentive compatible mechanism. Then a direct mechanism with the same decision rule, $(q, t'_1, t'_2, \ldots, t'_N)$, is incentive compatible if and only if for every $i \in I$ and every $\theta_{-i} \in \Theta_{-i}$ there is a number $\tau_i(\theta_{-i}) \in \mathbb{R}$ such that*

$$t'_i(\theta) = t_i(\theta) + \tau_i(\theta_{-i}) \text{ for all } \theta \in \Theta.$$

As our earlier revenue equivalence results, this result says that for given decision rule q the set of all transfer rules that implement q consists of one function and all its parallel translations. Because we are dealing with dominant strategy implementation, our result concerns ex post payoffs. The proof of this result is given in Krishna and Maenner (2001). The proof is analogous to the proof of similar results that we provided earlier, but it is more technical, because the appropriate differentiability properties of the functions involved needs to be established. We omit the proof.

5.4 Implementing Efficient Decision Rules

We now begin an investigation of the question for which decision rules q there are transfer rules that make the decision rule q dominant strategy incentive compatible. We begin by showing that decision rules that maximize ex ante expected welfare, where welfare is defined in a utilitarian way, are always dominant strategy incentive compatible. We shall call such decision rules "efficient," although a much more careful investigation would be needed if we wanted to clarify the relation between maximizing utilitarian welfare and "Pareto efficiency."

Definition 5.3. An allocation rule q^* is called "efficient" if for every $\theta \in \Theta$ we have:

$$\sum_{i=1}^{N} u_i(q^*(\theta), \theta_i) \geq \sum_{i=1}^{N} u_i(a, \theta_i) \text{ for all } a \in A.$$

We shall now introduce a class of mechanisms that make efficient decision rules dominant strategy incentive compatible. We encountered special cases of these mechanisms already in earlier parts of these notes.

Definition 5.4. A direct mechanism $(q, t_1, t_2, \ldots t_N)$ is called a "Vickrey-Clarke-Groves" (VCG) mechanism if q is an efficient decision rule, and if for every $i \in I$ there is a function

$$\tau_i : \Theta_{-i} \to \mathbb{R}$$

such that

$$t_i(\theta) = -\sum_{j \neq i} u_j(q(\theta), \theta_j) + \tau_i(\theta_{-i}) \text{ for all } \theta \in \Theta.$$

Proposition 5.2. *VCG mechanisms are dominant strategy incentive compatible.*

Proof. Consider any agent i, and take θ_{-i} as given. If agent i is of type θ_i and reports that she is of type θ_i', then her utility is:

$$\begin{aligned} & u_i(q(\theta_i', \theta_{-i}), \theta_i) + \sum_{j \neq i} u_j(q(\theta_i', \theta_{-i}), \theta_j) - \tau_i(\theta_{-i}) \\ = \quad & \sum_{j=1}^{N} u_j(q(\theta_i', \theta_{-i}), \theta_j) - \tau_i(\theta_{-i}). \end{aligned}$$

Note that $\tau_i(\theta_{-i})$ is not changed by agent i's report. Only the first expression matters for i's incentives. But this expression is social welfare at type vector θ if the decision is $q(\theta_i', \theta_{-i})$. As $q(\theta)$ maximizes social welfare for type vector θ, it is optimal for agent i to report her true type: $\theta_i' = \theta_i$. □

We can combine this result with the payoff equivalence result in Proposition 5.1 to obtain conditions under which VCG mechanisms are the only mechanisms that make efficient decision rules dominant strategy incentive compatible.

Corollary 5.1. *Suppose that for every $i \in I$, the set Θ_i is a convex subset of a finite dimensional Euclidean space. Moreover, assume that for every $i \in I$ the function $u_i(a, \theta_i)$ is a convex function of θ_i. Suppose that $(q, t_1, t_2, \dots t_N)$ is a dominant strategy incentive compatible mechanism, and suppose that q is efficient. Then $(q, t_1, t_2, \dots t_N)$ is a VCG mechanism.*

Proof. By Proposition 1 every dominant strategy incentive compatible mechanism that implements an efficient decision rule q must involve the same transfers as the VCG mechanism up to additive constants $\tau_i(\theta_{-i})$ that may be added to any agent i's transfers. But adding such constants to a VCG mechanism yields by the definition of VCG mechanisms another VCG mechanism. □

5.5 Characterizing All Incentive Compatible Decision Rules

In this section our objective is to characterize the set of *all* dominant strategy incentive compatible mechanisms. One might wonder why this is of interest, given that we showed in the previous section that efficient mechanisms are dominant strategy incentive compatible. There are two reasons. Firstly, the mechanism designer might have objectives other than efficiency, e.g. profit maximization. Secondly, the mechanism designer might face constraints other than dominant strategy incentive compatibility which make the implementation of an efficient decision rule impossible.

Our focus will be on the decision rules q that may form part of an incentive compatible direct mechanism, rather than on the transfer rules. We discuss a variety of closely related properties of decision rules which are necessary, and under certain conditions also sufficient for decision rules to be dominant strategy implementable.

In the three examples of Chapters 3 and 4, we found that a necessary and sufficient condition for a decision rule to be implementable by a direct incentive compatible mechanism was that q was monotone. Here, we start with a result that can be proved by an appropriate adaptation of the argument that led to our earlier results.

Definition 5.5. A decision rule q is "weakly monotone" if for every $i \in I$, $\theta_{-i} \in \Theta_{-i}$, and all $\theta_i^1, \theta_i^2 \in \Theta_i$, if $q(\theta_i^1, \theta_{-i}) = a^1$ and $q(\theta_i^2, \theta_{-i}) = a^2$, then

$$u_i(a^1, \theta_i^1) - u_i(a^2, \theta_i^1) \geq u_i(a^1, \theta_i^2) - u_i(a^2, \theta_i^2).$$

This definition considers a situation where the collective choice for a type vector $(\theta_i^1, \theta_{-i})$ is alternative a^1, but if we change agent i's type from θ_i^1 to θ_i^2, then the collective choice becomes some alternative a^2. The definition says that in this case the utility difference for agent i between a^1 and a^2 must have decreased as we switched from θ_i^1 to θ_i^2. We can interpret the utility difference as agent i's "willingness to pay for alternative a^1 as opposed to alternative a^2." The result says that agent i's willingness to pay for a^1 as opposed to a^2 must be at least as large when agent i is of type θ_i^1 as it is when agent i is of type θ_i^2.

Proposition 5.3. *Suppose $(q, t_1, t_2, \ldots, t_N)$ is a dominant strategy incentive compatible mechanism. Then q is weakly monotone.*

Proof. Dominant strategy incentive compatibility implies for all $i \in I$, $\theta_{-i} \in \Theta_{-i}$, and all $\theta_i^1, \theta_i^2 \in \Theta_i$, if $q(\theta_i^1, \theta_{-i}) = a^1$ and $q(\theta_i^2, \theta_{-i}) = a^2$, then:

$$\begin{aligned} u_i(a^1, \theta_i^1) - t_i(\theta_i^1, \theta_{-i}) \geq u_i(a^2, \theta_i^1) - t_i(\theta_i^2, \theta_{-i}) \Leftrightarrow \\ u_i(a^1, \theta_i^1) - u_i(a^2, \theta_i^1) \geq t_i(\theta_i^1, \theta_{-i}) - t_i(\theta_i^2, \theta_{-i}) \end{aligned} \tag{5.1}$$

and

$$\begin{aligned} u_i(a^1, \theta_i^2) - t_i(\theta_i^1, \theta_{-i}) \leq u_i(a^2, \theta_i^2) - t_i(\theta_i^2, \theta_{-i}) \Leftrightarrow \\ u_i(a^1, \theta_i^2) - u_i(a^2, \theta_i^2) \leq t_i(\theta_i^1, \theta_{-i}) - t_i(\theta_i^2, \theta_{-i}). \end{aligned} \tag{5.2}$$

These two inequalities imply:

$$u_i(a^1, \theta_i^1) - u_i(a^2, \theta_i^1) \geq u_i(a^1, \theta_i^2) - u_i(a^2, \theta_i^2). \tag{5.3}$$

which means that q is weakly monotone. □

We shall now show that a stronger condition than weak monotonicity, called "cyclical monotonicity," is necessary and sufficient for implementability. Note that we can re-write the inequality that defines weak monotonicity as:

$$\left(u_i(a^1, \theta_i^2) - u_i(a^1, \theta_i^1)\right) + \left(u_i(a^2, \theta_i^1) - u_i(a^2, \theta_i^2)\right) \leq 0. \tag{5.4}$$

This inequality considers on the left hand side the sum of the changes to agent i's utility that occur in the following two thought experiments. First, we start at utility profile $(\theta_i^1, \theta_{-i})$ with implemented alternative a^1, and we switch agent i's type to θ_i^2, leaving the chosen alternative hypothetically unchanged as a^1. Next, we start at utility profile $(\theta_i^2, \theta_{-i})$ with implemented alternative a^2, and we switch agent i's type to θ_i^1, leaving again the chosen alternative hypothetically unchanged. Both switches may cause an increase or a decrease in agent i's utility. The condition says that the sum of these two utility changes is not positive.

Now suppose we consider a sequence of length $k \in \mathbb{N}, k \geq 2$ of types of agent i: $\theta_i^1, \theta_i^2, \ldots, \theta_i^k$, and assume that the last element of the sequence is the same as the first: $\theta_i^1 = \theta_i^k$. We can consider the same sequence of switches as we considered in the previous paragraph, and sum up the changes to agent i's utility caused by these switches. "Cyclical monotonicity" requires the sum to be not positive.

Definition 5.6. A decision rule q is "cyclically monotone" if for every $i \in I$, every $\theta_{-i} \in \Theta_{-i}$ and every sequence of length $k \in \mathbb{N}$ of types of agent i, $(\theta_i^1, \theta_i^2, \ldots, \theta_i^k) \in \Theta_i^k$, with $\theta_i^k = \theta_i^1$, we have:

$$\sum_{\kappa=1}^{k-1} (u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^\kappa, \theta_i^\kappa)) \leq 0$$

where for all $\kappa = 1, 2, \ldots, k$ we define: $a^\kappa \equiv q(\theta_i^\kappa, \theta_{-i})$.

Thus, cyclical monotonicity is the same as weak monotonicity if we restrict attention to the case that $k = 3$.

The following proposition is due to Rochet (1987). A remarkable aspect of this theorem is that it does not require any structure at all for the sets of alternatives or the sets of types. Nevertheless, an elementary proof shows that cyclical monotonicity is equivalent to dominant strategy implementability.

Proposition 5.4. *A decision rule q is part of a dominant strategy incentive compatible direct mechanism $(q, t_1, t_2, \ldots, t_N)$ if and only if q is cyclically monotone.*

Proof. We first show that if q is part of a dominant strategy incentive compatible direct mechanism then it is cyclically monotone. Dominant strategy incentive compatibility implies for every $\kappa = 1, 2, \ldots, k-1$ that type $\theta_i^{\kappa+1}$ has no incentive to pretend to be type κ, given that all other agents have reported type θ_{-i}:

$$\begin{aligned} u_i(a^\kappa, \theta_i^{\kappa+1}) - t_i(\theta_i^\kappa, \theta_{-i}) \leq u_i(a^{\kappa+1}, \theta_i^{\kappa+1}) - t_i(\theta_i^{\kappa+1}, \theta_{-i}) \Leftrightarrow \\ u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^{\kappa+1}, \theta_i^{\kappa+1}) \leq t_i(\theta_i^\kappa, \theta_{-i}) - t_i(\theta_i^{\kappa+1}, \theta_{-i}). \end{aligned} \tag{5.5}$$

We now sum this inequality over all $\kappa = 1, 2, \ldots, k-1$. The right hand side then becomes zero, and we can deduce:

$$\begin{aligned} \sum_{\kappa=1}^{k-1} \left(u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^{\kappa+1}, \theta_i^{\kappa+1}) \right) \leq 0 \Leftrightarrow \\ \sum_{\kappa=1}^{k-1} u_i(a^\kappa, \theta_i^{\kappa+1}) - \sum_{\kappa=1}^{k-1} u_i(a^{\kappa+1}, \theta_i^{\kappa+1}) \leq 0 \end{aligned} \tag{5.6}$$

Because $\theta_i^k = \theta_i^1$, we can write the subtracted sum as follows:

$$\begin{aligned} \sum_{\kappa=1}^{k-1} u_i(a^\kappa, \theta_i^{\kappa+1}) - \sum_{\kappa=1}^{k-1} u_i(a^\kappa, \theta_i^\kappa) \leq 0 \Leftrightarrow \\ \sum_{\kappa=1}^{k-1} \left(u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^\kappa, \theta_i^\kappa) \right) \leq 0, \end{aligned} \tag{5.7}$$

which is what we wanted to show.

Next, we shall show that cyclical monotonicity implies that the rule q can be combined with transfer payments so that the direct mechanism that results is dominant strategy incentive compatible. To construct the transfer payments, we shall need some further definitions. We fix an arbitrary type $\widetilde{\theta}_i \in \Theta_i$. Then, for every $\theta_i \in \Theta_i$ we define $\mathcal{S}(\theta_i)$ to be the set of all finite sequences $(\theta_i^1, \theta_i^2, \ldots, \theta_i^k)$ of elements of Θ_i that satisfy: $\theta_i^1 = \widetilde{\theta}_i$ and $\theta_i^k = \theta_i$. Here, k can be any element of $\mathbb{N}$ with $k \geq 2$. Then define the function $V : \Theta \to \mathbb{R}$ by:

$$V(\theta_i, \theta_{-i}) \equiv \sup_{(\theta_i^1, \theta_i^2, \ldots, \theta_i^k) \in \mathcal{S}(\theta_i)} \sum_{\kappa=1}^{k-1} \left(u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^\kappa, \theta_i^\kappa) \right) \tag{5.8}$$

for all $(\theta_i, \theta_{-i}) \in \Theta$, where a^κ is defined to be $q(\theta_i^\kappa, \theta_{-i})$ for $\kappa = 1, 2, \ldots, k-1$.

Before we proceed, we verify that the function V is well-defined. This is the case if the expression the supremum of which is the right hand side of equation (5.8) is bounded from above for given θ_i and θ_{-i} if q is cyclically monotone. This is what we show.

First we note that the condition of cyclical monotonicity means that for $\theta_i = \bar{\theta}_i$ that all sums on the right hand side of (5.8) are non-positive. Therefore, for $\theta_i = \widetilde{\theta}_i$, the supremum is well-defined. Indeed, because the trivial sequence $(\widetilde{\theta}_i, \widetilde{\theta}_i)$ is contained in $\mathcal{S}(\widetilde{\theta}_i)$, and because for this sequence the sum over which we take the supremum in the definition of V is zero, we can conclude that $V(\widetilde{\theta}_i, \theta_{-i}) = 0$ for all $\theta_{-i} \in \Theta_{-i}$.

Now consider some $\theta_i \neq \bar{\theta}_i$, fix a sequence $(\theta_i^1, \theta_i^2, \ldots, \theta_i^k) \in \mathcal{S}(\theta_i)$ and define a^κ as in the definition of V. Then:

$$\begin{aligned} V(\widetilde{\theta}_i, \theta_{-i}) &= \sup_{(\theta_i^1, \theta_i^2, \ldots, \theta_i^k) \in \mathcal{S}(\widetilde{\theta}_i)} \sum_{\kappa=1}^{k-1} \left(u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^\kappa, \theta_i^\kappa)\right) \\ &\geq \sum_{\kappa=1}^{k-1} \left(u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^\kappa, \theta_i^\kappa)\right) + \left(u_i(a^k, \widetilde{\theta}_i) - u_i(a^k, \theta_i)\right) \end{aligned} \tag{5.9}$$

Here, the inequality follows from the fact that $\mathcal{S}(\widetilde{\theta}_i)$ includes the set of all finite sequences that have θ_i as their penultimate element, and that have $\widetilde{\theta}_i$ as their last element. The inequality that we have obtained is equivalent to:

$$\sum_{\kappa=1}^{k-1} \left(u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^\kappa, \theta_i^\kappa)\right) \leq V(\widetilde{\theta}_i, \theta_{-i}) + \left(u_i(a^k, \theta_i) - u_i(a^k, \widetilde{\theta}_i)\right). \tag{5.10}$$

Using $V(\widetilde{\theta}_i, \theta_{-i}) = 0$, which we showed earlier, we can infer:

$$\sum_{\kappa=1}^{k-1} \left(u_i(a^\kappa, \theta_i^{\kappa+1}) - u_i(a^\kappa, \theta_i^\kappa)\right) \leq u_i(a^k, \theta_i) - u_i(a^k, \widetilde{\theta}_i), \tag{5.11}$$

which shows that the sums that appear on the right hand side of (5.8) are bounded from above, and therefore that V is well-defined.

We now use the function V to construct the payment schemes that make q dominant strategy incentive compatible. Indeed, we shall construct the payment schemes so that the utility $u_i(q(\theta), \theta_i)$ is exactly equal to $V(\theta)$ for all $\theta \in \Theta$. That is, we set:

$$\begin{aligned} u_i(q(\theta), \theta_i) - t_i(\theta) = V(\theta) \Leftrightarrow \\ t_i(\theta) = u_i(q(\theta), \theta_i)) - V(\theta). \end{aligned} \tag{5.12}$$

To prove that this implies that the mechanism is dominant strategy incentive compatible, we need to show:

$$\begin{aligned} u_i(q(\theta_i, \theta_{-i}), \theta_i) - t_i(\theta_i, \theta_{-i}) \geq u_i(q(\theta'_i, \theta_{-i}), \theta_i) - t_i(\theta'_i, \theta_{-i}) \Leftrightarrow \\ u_i(q(\theta_i, \theta_{-i}), \theta_i) - (u_i(q(\theta_i, \theta_{-i}), \theta_i)) - V(\theta_i, \theta_{-i})) \geq \\ u_i(q(\theta'_i, \theta_{-i}), \theta_i) - \big(u_i(q(\theta'_i, \theta_{-i}), \theta'_i)) - V(\theta'_i, \theta_{-i})\big) \Leftrightarrow \\ V(\theta_i, \theta_{-i}) \geq V(\theta'_i, \theta_{-i}) + u_i(q(\theta'_i, \theta_{-i}), \theta_i) - u_i(q(\theta'_i, \theta_{-i}), \theta'_i)) \end{aligned} \tag{5.13}$$

This is true by the definition of V because the set of all finite sequences that start with $\bar{\theta}_i$ and end with θ_i includes the set of all finite sequences that start with $\bar{\theta}_i$, have θ'_i as their penultimate element, and end with θ_i. □

5.6 All Incentive Compatible Decision Rules When Outcomes are Lotteries

We now develop a condition for implementability that is closely related to cyclical monotonicity, but that applies only to the special case in which the set of alternatives A is the set of probability distributions over some finite set of outcomes. Denote the number of such outcomes by M. Each element of A is of the form $(p_1, p_2, \ldots, p_M)$, where $p_\ell \geq 0$ for $\ell = 1, 2, \ldots, M$ and $\sum_{\ell=1}^{M} p_\ell = 1$. The set of types of any player i, Θ_i, consists of vectors $\theta_i = (\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,M})$ of Bernoulli utilities of the M outcomes. We assume that the set Θ_i is convex for every $i \in I$.

Consider a given direct mechanism. For every $i \in I$ define $U_i(\theta) \equiv q(\theta) \cdot \theta - t_i(\theta)$. Here, "$\cdot$" stands for the scalar vector product. Hence, U_i is the expected utility of player i if all players report their types truthfully. Now suppose that the mechanism is dominant strategy incentive compatible, and fix some value of θ_{-i}. Then U_i as a function of θ_i alone satisfies:

$$U_i(\theta_i, \theta_{-i}) = \max_{\theta'_i \in \Theta_i} \big(q(\theta'_i, \theta_{-i}) \cdot \theta_i - t_i(\theta'_i, \theta_{-i})\big). \tag{5.14}$$

Here is one way of parsing the right hand side of this equation: Consider the function that describes for each θ_i the expected utility that an agent of type

θ_i obtains when pretending to be type θ_i'. For every θ_i' we can define such a function. On the right hand side of the above equation we have for each θ_i the maximum of all the functions just described, where the maximum is taken over all θ_i'. Note that the function on the right hand side is linear in θ_i for given θ_i'. Thus, we have on the right hand side the maximum of a family of linear functions. Therefore, we can conclude that U_i is the maximum of functions that are linear in θ_i, and therefore it is convex in θ_i.

Now we review the following definition from Rockafellar (1970), p. 214: A vector x^* is a subgradient of a convex function f at a point x if:

$$f(z) \geq f(x) + x^* \cdot (z - x)$$

for all z in the domain of f. In our case, $q(\theta_i, \theta_{-i})$ is a subgradient of $U(\theta_i, \theta_{-i})$ (as a function of θ_i) at θ_i for every θ_i. This is because:

$$\begin{aligned}
&U_i(\theta_i', \theta_{-i}) \geq U(\theta_i, \theta_{-i}) + q(\theta_i) \cdot (\theta_i' - \theta_i) \Leftrightarrow \\
&U_i(\theta_i', \theta_{-i}) \geq q(\theta_i) \cdot \theta_i - t_i(\theta_i, \theta_{-i}) + q(\theta_i) \cdot (\theta_i' - \theta_i) \Leftrightarrow \\
&U_i(\theta_i', \theta_{-i}) \geq q(\theta_i) \cdot \theta_i' - t_i(\theta_i, \theta_{-i}) \qquad (5.15)
\end{aligned}$$

which holds by dominant strategy incentive compatibility for every $\theta_i' \in \Theta_i$. The following result shows that being a subgradient of a convex function is not only a necessary but also a sufficient condition for q to be implementable.

Proposition 5.5. *Suppose that A is the set of all probability distributions over some finite set of outcomes. Suppose that for every $i \in I$ the set Θ_i is a convex set of vectors of Bernoulli utility functions of player i. Then a decision rule q is part of a dominant strategy incentive compatible direct mechanism $(q, t_1, t_2, ..., t_N)$ if and only if for every $i \in I$ and for every $\theta_{-i} \in \Theta_{-i}$ there is a convex function $U_i : \Theta_i \to \mathbb{R}$ such that $q_i(\theta_i, \theta_{-i})$, regarded as a function of θ_i only, is a subgradient of U_i.*

Proof. The argument preceding Proposition 5.5 has shown the necessity of the condition in Proposition 5.5. To see that it is also sufficient, fix θ_{-i} and define for every $\theta_i \in \Theta_i$ player i's transfer $t_i(\theta_i, \theta_{-i})$ so that the function U_i of which q is the subgradient is exactly player i's expected utility. The equivalencies that precede Proposition 5.5 establish that dominant strategy incentive compatibility holds. □

Theorem 24.8 in Rockafellar (1970) establishes that q is cyclically monotone if and only it is subgradient of a convex function. Therefore, Rochet's result that was cited in the previous section is a generalization of Rockafellar's result to the more general setting in which the sets of alternatives and types are arbitrary.

5.7 Single Dimensional Type Spaces

We want to relate weak monotonicity in the sense of Definition 5.5 to monotonicity as we have seen it in earlier parts of these notes. A characteristic of these earlier parts was that agents' type spaces were subsets of $\mathbb{R}$, and were therefore single-dimensional. Single-dimensionality is at first sight purely mathematical concept. In economic terms, however, it means that type spaces can be completely ordered so that "higher types" have a larger marginal willingness to pay for "higher alternatives." Our purpose in this section is to study the implications of the single dimensionality condition for dominant strategy incentive compatibility. Before we can do so, we need to formalize the notions of "higher alternatives" and "higher types."

We begin by describing what "higher alternatives" are. Let R_i be an order of A. By this we mean that it is a complete, reflexive, and transitive binary relation. The strict order derived from R_i is denoted by P_i: $aP_ib \Leftrightarrow [aR_ib$ and not $bR_ia]$. The indifference relation derived from R_i is denoted by I_i: $aI_ib \Leftrightarrow [aR_ib$ and $bR_ia]$.

It is important not to mistake R_i for agent i's preference relation. To understand this, consider the single unit auction example. In that example we might take R_i to be given by: $aR_ib \Leftrightarrow [i$ obtains the good with at least as high probability in a as in $b]$. The single crossing condition will not rule out that some types prefer not obtaining the good over obtaining the good, i.e. have a negative marginal willingness to pay for the good. Single crossing will require that the larger types are, the larger is the marginal willingness to pay for the good.

Given an ordering of alternatives, we can now order types.

Definition 5.7. Consider any $i \in I$, and let R_i be an order of A. For any pair of types $\theta_i, \theta_i' \in \Theta_i$, we shall say that $\theta_i \succ_{R_i} \theta_i'$ ("θ_i is a higher type than θ_i' relative to R_i") if

$$u_i(a', \theta_i) - u_i(a, \theta_i) > u_i(a', \theta_i') - u_i(a, \theta_i') \text{ for all } a, a' \in A \text{ with } a'P_ia$$

and

$$u_i(a', \theta_i) - u_i(a, \theta_i) = u_i(a', \theta_i') - u_i(a, \theta_i') = 0 \text{ for all } a, a' \in A \text{ with } a'I_ia.$$

Intuitively, $\theta_i \succ_{R_i} \theta_i'$ means that θ_i attaches larger marginal value to higher alternatives than θ_i' for any two ordered alternatives that we compare. In other words, θ_i unambiguously has a stronger preference for higher alternatives than θ'_i. What do we mean here by "higher alternatives"? This is defined by the ordering R_i on A. Thus, the order $\succ_{R_i}$ is conditional on R_i.

We are now going to use the order $\succ_{R_i}$ to define the monotonicity property that will allow us to relate weak monotonicity as in Definition 5.5 to the monotonicity properties of earlier parts of these notes. Note that the following definition applies to any arbitrary decision rules, irrespective of whether the type spaces whose cross product makes up the domain of the decision rule satisfies any "single dimensionality condition."

Definition 5.8. Consider any agent i, and let R_i be an order of A. A decision rule q is called "monotone with respect to R_i" if:

$$\theta_i \succ_{R_i} \theta'_i \Rightarrow q(\theta_i, \theta_{-i}) R_i q(\theta'_i, \theta_{-i}) \text{ for all } \theta_{-i} \in \Theta_{-i}.$$

In words, this definition says that a decision rule is "monotone with respect to R_i" if an unambiguous increase in an agent's type leads to a "higher" alternative being chosen. The next result shows that monotonicity in this sense is implied by weak monotonicity.

Proposition 5.6. *Consider any agent i, and let R_i be an order of A. If a decision rule q is weakly monotone then it is monotone with respect to R_i.*

A remarkable feature of Proposition 5.6 is that it is true for *any* order R_i of A.

Proof. Suppose q is weakly monotone. Consider some $\theta_{-i} \in \Theta_{-i}$, and suppose $\theta_i \succ_{R_i} \theta'_i$. Define $a \equiv q(\theta_i, \theta_{-i})$ and $a' \equiv q(\theta'_i, \theta_{-i})$. By the definition of weak monotonicity, we must have: $u_i(a, \theta_i) - u_i(a', \theta_i) \geq u_i(a, \theta'_i) - u_i(a', \theta'_i)$. We now prove indirectly that $\theta_i \succ_{R_i} \theta'_i$ implies aR_ia'. Suppose $a'P_ia$. Then $\theta_i \succ_{R_i} \theta'_i$ implies

$$\begin{aligned} & u_i(a', \theta_i) - u_i(a, \theta_i) > u_i(a', \theta'_i) - u_i(a, \theta'_i) \Leftrightarrow \\ & u_i(a, \theta_i) - u_i(a', \theta_i) < u_i(a, \theta'_i) - u_i(a', \theta'_i), \end{aligned} \tag{5.16}$$

which contradicts weak monotonicity. □

The reason that we can't prove in general the converse of this result is that weak monotonicity imposes a restriction on collective decisions only in the case that types are comparable in the order $\succ_{R_i}$, which is in general incomplete. By contrast, monotonicity places restrictions on choices even if types are not ordered in $\succ_{R_i}$. But we can strengthen Proposition 5.6 so that it becomes an equivalence if we consider completely ordered type spaces.

Definition 5.9. Consider any agent i, and let R_i be an order of A. The type space Θ_i is "single dimensional with respect to R_i" if for any $\theta_i, \theta_i' \in \Theta_i$ with $\theta_i \neq \theta_i'$ we have either $\theta_i \succ_{R_i} \theta_i'$ or $\theta_i' \succ_{R_i} \theta_i$ or both.

Note that if a type is in this sense single dimensional we can assign to any type a real number, namely the difference in utility for some alternatives, say the difference $u_i(a', \theta_i) - u_i(a, \theta_i)$ where $a' P_i a$, and this mapping from types to real numbers will be invertible, that is this marginal utility unambiguously identifies the type. Moreover, the larger this number, the larger *all* marginal utilities of the type.

We can now show that for completely ordered, i.e. single-dimensional, type spaces, monotonicity is sufficient for weak monotonicity.

Proposition 5.7. *For every $i \in I$ let R_i be an ordering of A such that Θ_i is single dimensional with respect to R_i. Then a decision rule q is weakly monotone if and only if it is monotone with respect to R_i for every $i \in I$.*

Proof. In light of Proposition 5.6, we only have to show that monotonicity implies weak monotonicity. So suppose q is monotone, and consider $\theta \in \Theta$ and $\theta_i' \in \Theta_i$ such that $q(\theta) = a$ but $q(\theta_i', \theta_{-i}) = a'$. We may restrict our attention to the case $a \neq a'$. Because θ satisfies the single crossing condition we must have: $\theta_i \succ_{R_i} \theta_i'$ or $\theta_i' \succ_{R_i} \theta_i$. Without loss of generality, assume the former. Then, because q is monotone, $a R_i a'$. Moreover, because $\theta_i \succ_{R_i} \theta_i'$:

$$u_i(a, \theta_i) - u_i(a', \theta_i) > u_i(a, \theta_i') - u_i(a', \theta_i'). \tag{5.17}$$

if $a P_i a'$, or

$$u_i(a, \theta_i) - u_i(a', \theta_i) = u_i(a, \theta_i') - u_i(a', \theta_i'). \tag{5.18}$$

if $a I_i a'$. This shows that q is weakly monotone. □

Having established that weak monotonicity and monotonicity are equivalent on single dimensional domains, we next show that monotonicity is sufficient for an allocation rule to be implementable by appropriate transfer schemes. This, together with Proposition 5.3, then implies that on single-dimensional domains, monotonicity is necessary and sufficient for implementability.

For simplicity, we restrict our attention in the next result to the case in which A is finite. The case in which A is infinite is harder only in terms of notation. We also assume that the type spaces are bounded in the sense of the following definition:

Definition 5.10. For any agent i, the type space Θ_i is "bounded" if there is a constant $c > 0$ such that for all $a, a' \in A$ and all $\theta_i \in \Theta_i$ we have:

$$-c < u_i(a', \theta_i) - u_i(a, \theta_i) < c. \tag{5.19}$$

Intuitively, the type space of agent i is "bounded" if there is a uniform upper bound for agent i's willingness to pay for a change in the decision.

Proposition 5.8. *Suppose that A is finite. For every $i \in I$ let R_i be an order of A, and suppose that for every $i \in I$ the type space Θ_i is bounded and is single dimensional with respect to R_i. Let q be a decision rule that is monotone with respect to R_i for every $i \in I$. Then there are transfer rules $t_1, t_2, \ldots, t_N$ such that $(q, t_1, t_2, \ldots, t_N)$ is dominant strategy incentive compatible.*

Proof. Fix $i \in I$ and $\theta_{-i} \in \Theta_{-i}$. Denote the range of q over Θ_i, that is the set $\{q(\theta_i, \theta_{-i}) \mid \theta_i \in \Theta_i\}$, by $\{a^1, a^2, \ldots, a^n\}$ where $a^n R_i a^{n-1} R_i \ldots R_i a^1$. For simplicity we assume that all these preference relations are strict: $a^n P_i a^{n-1} P_i \ldots P_i a^1$. If some are not, then the proof below should be modified to treat alternatives between which the decision maker is indifferent as identical alternatives.

If $n = 1$, we set $t_i(\theta_i, \theta_{-i}) = 0$ for all θ_i. This will obviously make it optimal for player i to report her type truthfully if the other agents' types is θ_{-i} because the same outcome will occur regardless of player i's report.

If $n \geq 2$, define for every $k = 1, 2, \ldots, n$ the set

$$\Theta_i^k \equiv \{\theta_i \in \Theta_i \mid q(\theta_i, \theta_{-i}) = a^k\}. \tag{5.20}$$

Monotonicity and the single crossing condition imply that the sets Θ_i^k are ordered in the following sense:

$$k' > k, \theta_i^k \in \Theta_i^k, \theta_i^{k'} \in \Theta_i^{k'} \Rightarrow \theta_i^{k'} \succ_{P_i} \theta_i^k. \tag{5.21}$$

Define for every $k = 2, \ldots, n$:

$$\tau^k \equiv \inf\{u_i(a^k, \theta_i) - u_i(a^{k-1}, \theta_i) \mid \theta_i \in \Theta_i^k\}. \tag{5.22}$$

The infimum here is well-defined because Θ_i^k is bounded. Note that the ordering of types implies:

$$\begin{aligned} k' < k, \theta_i' \in \Theta_i^{k'} &\Rightarrow u_i(a^k, \theta_i') - u_i(a^{k-1}, \theta_i') \leq \tau^k \\ k' > k, \theta_i' \in \Theta_i^{k'} &\Rightarrow u_i(a^k, \theta_i') - u_i(a^{k-1}, \theta_i') \geq \tau^k. \end{aligned} \tag{5.23}$$

We define agent i's transfer payment as follows:

$$t_i(\theta) = \begin{cases} 0 & \text{if } \theta_i \in \Theta_i^1; \\ \sum_{\kappa=2}^{k} \tau^k & \text{if } \theta_i \in \Theta_i^k \text{ where } k \geq 2. \end{cases} \quad (5.24)$$

We verify that this transfer scheme makes truthful reporting of θ_i is an optimal strategy for agent i given any θ_{-i}. If agent i's true type is $\theta_i \in \Theta_i^k$, and if she reports any type in Θ_i^k, her utility will be independent of her report. If she reports a type in $\Theta_i^{k'}$ where $k' > k$, then the change in her utility in comparison to truthful reporting will be:

$$\begin{aligned} & u_i(a^{k'}, \theta_i) - u_i(a^k, \theta_i) - \sum_{\kappa=k+1}^{k'} \tau^\kappa \\ = & \sum_{\kappa=k+1}^{k'} \Big(u_i(a^\kappa, \theta_i) - u_i(a^{\kappa-1}, \theta_i) \Big) - \sum_{\kappa=k+1}^{k'} \tau^\kappa \\ \leq & \sum_{\kappa=k+1}^{k'} \big(u_i(a^\kappa, \theta_i) - u_i(a^{\kappa-1}, \theta_i) \big) - \sum_{\kappa=k+1}^{k'} \big(u_i(a^\kappa, \theta_i) - u_i(a^{\kappa-1}, \theta_i) \big) = 0 \end{aligned} \quad (5.25)$$

The first equality is a simple re-writing. The inequality follows from the first inequality in (5.23). A symmetric argument proves there is no incentive for agent i to report a type in $\Theta_i^{k'}$ where $k' < k$. □

Proposition 8 is analogous to the result that we obtained in earlier sections that monotonicity of q is necessary and sufficient for implementability of q. In all those models, the domain of preferences satisfied a single dimensionality condition (although we allowed infinite outcome sets).

5.8 Sufficiency of Weak Monotonicity

The single crossing domains described in Section 5.7 are not the only domains on which one can show that weak monotonicity is sufficient for implementability, and is thus equivalent to cyclical monotonicity. Bikhchandani et al. (2006) have found several other conditions under which this result is true. We explain here the simplest one. Bikhchandani et. al. say that the set Θ_i of types of agent i is "rich" if there is some reflexive and transitive, but possibly incomplete, binary relation R_i on A such that all utility functions that represent R_i are possible utility functions of agent i. Formally:

Definition 5.11. Consider any $i \in I$. The type space Θ_i is called "rich" if there is a possibly incomplete preference relation R_i on A such that for every $u : A \to \mathbb{R}$ that represents R_i, i.e. that satisfies: $aR_ib \Rightarrow u(a) \geq u(b)$, there is a $\theta_i \in \Theta_i$ such that

$$u_i(a, \theta_i) = u(a) \text{ for all } a \in A.$$

Note that this condition becomes more restrictive as R_i becomes less complete, i.e. as comparisons are dropped from R_i. The reason is that the less complete R_i is, the more utility functions may represent R_i.

Bikhchandani et. al.'s Theorem 1 is:

Proposition 5.9. *Suppose that A is finite. Suppose that for every $i \in I$ the type space Θ_i is rich. Let q be a weakly monotone decision rule. Then there are transfer schemes $t_1, t_2, \ldots, t_N$ such that $(q, t_1, t_2, \ldots, t_N)$ is dominant strategy incentive compatible.*

Bikhchandani et. al. emphasize that their result does not apply to the case that A is a set of lotteries over outcomes because this would make the set A infinite.

5.9 Positive Association of Differences

We now consider another condition that is weaker than monotonicity, called "positive association of differences" (PAD). We are considering this relatively weak condition because it is sufficient to obtain a surprisingly strong result.

Definition 5.12. A decision rule q satisfies "positive association of differences" (PAD) if $\theta, \theta' \in \Theta$, $q(\theta) = a$, and

$$u_i(a, \theta'_i) - u_i(b, \theta'_i) > u_i(a, \theta_i) - u_i(b, \theta_i)$$

for all $i \in I$ and $b \in A$ with $b \neq a$, implies $q(\theta') = a$.

In words, q satisfies PAD if whenever an alternative that is chosen at some type vector will also be chosen at any other type vector where the alternative's marginal utilities in comparison to other alternatives are larger.

Proposition 5.10. *If q is weakly monotone then it satisfies PAD.*

Proof. Suppose that $\theta, \theta' \in \Theta$, $q(\theta) = a$ and that θ' satisfies the condition in the definition of PAD. We shall prove that then $q(\theta'_i, \theta_{-i}) = a$ for all $i \in I$. The assertion that we have to prove, $q(\theta') = a$, then follows from the N-fold application of the same argument. Suppose $q(\theta'_i, \theta_{-i}) = b \neq a$. By the conditions on θ and θ' we have:

$$\begin{aligned} u_i(a, \theta'_i) - u_i(b, \theta'_i) &> u_i(a, \theta_i) - u_i(b, \theta_i) \Leftrightarrow \\ u_i(b, \theta'_i) - u_i(a, \theta'_i) &< u_i(b, \theta_i) - u_i(a, \theta_i). \end{aligned}$$

Thus, we have a contradiction with weak monotonicity. □

PAD is weaker than monotonicity because it only puts restrictions on type profiles which result in the same collective decision. By contrast, monotonicity also refers to type profiles which result in different collective decisions. However, Roberts (1979) showed that if the domain of a decision rule consists of all possible utility functions, and if some other conditions hold, then "Positive Association of Differences" is sufficient for dominant strategy incentive compatibility.

One of the further conditions needed to state Roberts' result is that the decision rule q is flexible:

Definition 5.13. A decision rule q is called "flexible" if its range, $q(\Theta)$, has at least three elements.

Observe that a decision rule can only be flexible if $\#A \geq 3$. We shall also assume that A is finite.

Proposition 5.11. *Suppose A is finite, and suppose for every $i \in I$ and $\nu \in \mathbb{R}^{\#A}$ there is a $\theta_i \in \Theta_i$ such that $(u_i(a, \theta_i))_{a \in A} = \nu$. For every flexible decision rule q that satisfies PAD there are transfer rules $(t_1, t_2, \ldots, t_N)$ such that $(q, t_1, t_2, \ldots, t_N)$ is dominant strategy incentive compatible.*

Note that this result, by assuming A to be finite, again rules out the case in which A is the set of all lotteries over some finite set of outcomes. Also, the condition for Θ_i on which this result relies is very restrictive. It rules out that the direct mechanism that we are constructing embodies any prior knowledge about the agents' preferences.

Roberts proved his result by obtaining an interesting characterization of all decision rules that satisfy the conditions of Proposition 5.11. Although we shall not provide a proof of this characterization, we mention it, and show how it implies Proposition 5.11.

Proposition 5.12. *Suppose A is finite, and suppose for every $i \in I$ and $\nu \in \mathbb{R}^{\#A}$ there is a $\theta_i \in \Theta_i$ such that $(u_i(a,\theta_i))_{a\in A} = \nu$. Then a flexible decision rule q satisfies PAD if and only if for every $i \in I$ there is a real number $k_i > 0$, and there is a function $F : A \to \mathbb{R}$ such that for every $\theta \in \Theta$:*

$$\sum_{i=1}^{N} k_i u_i(q(\theta),\theta_i) + F(q(\theta)) \geq \sum_{i=1}^{N} k_i u_i(a,\theta_i) + F(a)$$

for all $a \in A$.

In words, this result says that under the assumptions of the result a decision rule satisfies PAD if and only if it maximizes a weighted utilitarian welfare criterion with exogenous bias. The weight of agent i's utility under the utilitarian welfare criterion is k_i. The bias is described by the function F. This function assigns to each alternative a measure of welfare that is independent of agents' types. For example, F could pick out some particular alternative $\bar{a} \in A$ as the "status quo", and set $F(\bar{a}) = z > 0$ and $F(a) = 0$ for all $a \in A$ with $a \neq z$. Then any alternative other than $\bar{a}$ would have to imply social welfare that exceeds that of the status quo by at least z if it is to be preferred over the status quo.

We now show that Proposition 5.12 implies Proposition 5.11. It is obvious from Propositions 5.3 and 5.10 that PAD is necessary for implementability. All that is needed to derive Proposition 5.11 from Proposition 5.12 is therefore to show the following:

Proposition 5.13. *Suppose q satisfies the characterization in Proposition 5.12. Then there are transfer rules $(t_1, t_2, \ldots, t_N)$ such that $(q, t_1, t_2, \ldots, t_N)$ is dominant strategy incentive compatible.*

Proof. This follows from a generalization of the VCG construction. We can define agent i's transfer payment as follows:

$$t_i(\theta) = -\frac{1}{k_i}\left(\sum_{j\neq i} k_j u_j(q(\theta),\theta_j) + F(q(\theta))\right). \tag{5.26}$$

Here, in comparison to the VCG formula, we have omitted all terms not depending on agent i's report. Those terms don't affect the argument. Agent i's utility when she is type θ_i, she reports being type θ_i', and all other agents report type vector θ_{-i} is:

$$u_i(q(\theta_i',\theta_{-i}),\theta_i) + \frac{1}{k_i}\left(\sum_{j\neq i} k_j u_j(q(\theta_i',\theta_{-i}),\theta_j) + F(q(\theta_i',\theta_{-i}))\right). \tag{5.27}$$

Maximizing this expression is equivalent to maximizing the product of this expression and k_i. This product is:

$$\begin{aligned} & k_i u_i(q(\theta_i', \theta_{-i}), \theta_i) + \sum_{j \neq i} k_j u_j(q(\theta_i', \theta_{-i}), \theta_j) + F(q(\theta_i', \theta_{-i})) \\ = & \sum_{j=1}^{n} k_j u_j(q(\theta_i', \theta_{-i}), \theta_j) + F(q(\theta_i', \theta_{-i})) \end{aligned} \tag{5.28}$$

and hence agent i chooses his report to maximize the same function that q maximizes. Therefore, reporting θ_i truthfully is optimal. □

5.10 Individual Rationality and Budget Balance

We now enrich our framework to bring in individual rationality and budget balance. We begin with individual rationality. In the examples that we have seen in earlier parts of these notes, individual rationality required that if agents report their types truthfully they are at least as well off as they would be if they obtained some particular alternative in A and had to pay nothing. In the auction example, the alternative in A that corresponded to individual rationality was "not obtaining the good." In the public goods example it was "no production of the public good." In the bilateral trade example it was "no trade." To generalize, we shall assume that for every agent i some alternative $a_i \in A$ is given that is as good as agent i's outside option, and thus that determines his individual rationality constraint.

Definition 5.14. For every agent i, let $a_i \in A$. A direct mechanism is "ex post individually rational with respect to $(a_1, a_2, \ldots, a_N)$" if for all $i \in I$ and all $\theta \in \Theta$ we have:

$$u_i(q(\theta), \theta_i) - t_i(\theta) \geq u_i(a_i, \theta_i).$$

In the examples in previous parts of these notes, for incentive compatible mechanisms, individual rationality was true for all types if and only if it was true for the lowest (in the case of the seller in the bilateral trade model: the highest) type. To obtain a result of this kind more generally, we impose the single crossing condition, and we assume that a_i is lowest ranked in A in the ordering P_i that reflects agent i's order of the elements of A.

Proposition 5.14. *Consider any agent $i \in I$, and let R_i be an order of A. Suppose that the type set Θ_i is single-dimensional with respect to R_i. Assume that there is a type $\underline{\theta}_i$ that is the lowest type in the order $\succ_{R_i}$ in*

Θ_i, i.e. $\theta_i \succ_{R_i} \underline{\theta}_i$ for all $\theta_i \in \Theta_i$ such that $\theta_i \neq \underline{\theta}_i$. Finally, assume that a_i is the lowest element of A in the order R_i, i.e. bR_ia_i for every $b \in A$ with $b \neq a$. Then a dominant strategy incentive compatible mechanism satisfies the ex post individual rationality constraint for agent i if and only if for every $\theta_{-i} \in \Theta_{-i}$:

$$u_i(q(\underline{\theta}_i, \theta_{-i}), \underline{\theta}_i) - t_i(\underline{\theta}_i, \theta_{-i}) \geq u_i(a_i, \underline{\theta}_i).$$

Proof. The condition in Proposition 5.14 is the individual rationality constraint for the lowest type. It is therefore necessary. To show that it is also sufficient, we show that it implies individual rationality for any type $\theta_i \neq \underline{\theta}_i$. We first note that:

$$\begin{aligned} u_i(q(\underline{\theta}_i, \theta_{-i}), \underline{\theta}_i) - t_i(\underline{\theta}_i, \theta_{-i}) \geq u_i(a_i, \underline{\theta}_i) \Leftrightarrow \\ u_i(q(\underline{\theta}_i, \theta_{-i}), \underline{\theta}_i) - u_i(a_i, \underline{\theta}_i) \geq t_i(\underline{\theta}_i, \theta_{-i}). \end{aligned} \tag{5.29}$$

Now if $\theta_i \neq \underline{\theta}_i$, then by the single crossing condition: $\theta_i \succ_{R_i} \underline{\theta}_i$. Because, moreover, the alternative $q(\underline{\theta}_i, \theta_{-i})$ either satisfies: $q(\underline{\theta}_i, \theta_{-i})P_ia_i$, or: $q(\underline{\theta}_i, \theta_{-i})I_ia_i$ we can infer:

$$\begin{aligned} u_i(q(\underline{\theta}_i, \theta_{-i}), \theta_i) - u_i(a_i, \theta_i) \geq t_i(\underline{\theta}_i, \theta_{-i}) \Leftrightarrow \\ u_i(q(\underline{\theta}_i, \theta_{-i}), \theta_i) - t_i(\underline{\theta}_i, \theta_{-i}) \geq u_i(a_i, \theta_i). \end{aligned} \tag{5.30}$$

By dominant strategy incentive compatibility:

$$u_i(q(\theta_i, \theta_{-i}), \theta_i) - t_i(\theta_i, \theta_{-i}) \geq u_i(q(\underline{\theta}_i, \theta_{-i}), \theta_i) - t_i(\underline{\theta}_i, \theta_{-i}). \tag{5.31}$$

Combining the last two inequalities we obtain:

$$u_i(q(\theta_i, \theta_{-i}), \theta_i) - t_i(\theta_i, \theta_{-i}) \geq u_i(a_i, \theta_i), \tag{5.32}$$

which is ex post individual rationality for type θ_i. □

We turn next to budget balance. Without loss of generality we define budget balance as the requirement that all transfer payments add up to zero. If alternatives in A have costs associated with them, we can redefine outcomes so that some division, say equal division, of these costs is already included in outcomes.

Definition 5.15. A direct mechanism is "ex post budget balanced" if for all $\theta \in \Theta$ we have:

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

We now investigate conditions under which efficient decision rules can be implemented with a balanced budget. Because under certain conditions VCG mechanisms are the only mechanisms that implement efficient decision rules, it is relevant to ask when VCG mechanisms are budget balanced. The following Proposition provides the answer. It shows that a necessary and sufficient condition is a restriction on the functional form of the welfare generated by an efficient decision rule if this welfare is regarded as a function of the type vector.

Proposition 5.15. *Let q be an efficient decision rule. Then a budget-balanced VCG mechanism that implements q exists if and only if for every $i \in I$ there is a function $f_i : \Theta_{-i} \to \mathbb{R}$ such that:*

$$\sum_{i=1}^{N} u_i(q(\theta), \theta_i) = \sum_{i=1}^{N} f_i(\theta_{-i}) \text{ for all } \theta \in \Theta.$$

Proof. We first prove the necessity of this condition. Consider a VCG mechanism, and let $\tau_i : \Theta_i \to \mathbb{R}$ be the functions referred to in Definition 5.4. The mechanism is budget balanced if for all $\theta \in \Theta$:

$$\sum_{i=1}^{N} \left(-\sum_{j \neq i} u_j(q(\theta), \theta_j) + \tau_i(\theta_{-i}) \right) = 0 \Leftrightarrow$$

$$(N-1) \sum_{i=1}^{N} u_i(q(\theta), \theta_i) = \sum_{i=1}^{N} \tau_i(\theta_{-i}) \Leftrightarrow$$

$$\sum_{i=1}^{N} u_i(q(\theta), \theta_i) = \sum_{i=1}^{N} \frac{\tau_i(\theta_{-i})}{N-1} \qquad (5.33)$$

Hence, if we set for every $i \in I$ and $\theta_{-i} \in \Theta_{-i}$

$$f_i(\theta_{-i}) \equiv \frac{\tau_i(\theta_{-i})}{N-1} \qquad (5.34)$$

we have obtained the desired form for the function $\sum_{i=1}^{N} u_i(q(\theta), \theta_i)$.

Next we prove sufficiency of the condition. Suppose that $\sum_{i=1}^{N} u_i(q(\theta), \theta_i)$ has the form described in the Proposition. For every $i \in I$ and every $\theta_{-i} \in \Theta_{-i}$ we consider the VCG mechanism with

$$\tau_i(\theta_{-i}) \equiv (N-1) f_i(\theta_{-i}). \qquad (5.35)$$

Then for every $\theta \in \Theta$ the sum of agents' payments is:

$$\sum_{i=1}^{N}\left(-\sum_{j\neq i} u_j(q(\theta),\theta_j)+(N-1)f(\theta_{-i})\right)$$
$$= -(N-1)\sum_{i=1}^{N} u_i(q(\theta),\theta_i)+(N-1)\sum_{i=1}^{N} f(\theta_{-i}) \qquad (5.36)$$

which is zero by the assumption of the Proposition. □

As an application of this result we consider the bilateral trade model. In that model the maximized welfare is given by: $\max\{\theta_S, \theta_B\}$. The condition of the above Proposition is that there are functions f_B and f_S such that $\max\{\theta_S, \theta_B\} = f_B(\theta_S) + f_S(\theta_B)$. Now suppose that the intersection of the intervals $[\underline{\theta}_S, \bar{\theta}_S]$ and $[\underline{\theta}_B, \bar{\theta}_B]$ has a non-empty interior, and suppose that θ, θ' are two different elements of that interior with $\theta < \theta'$. Then the necessary and sufficient condition in Proposition 5.15 requires:

$$\begin{aligned}
\theta &= \max\{\theta,\theta\} = f_B(\theta)+f_S(\theta)\\
\theta' &= \max\{\theta,\theta'\} = f_B(\theta)+f_S(\theta')\\
\theta' &= \max\{\theta',\theta\} = f_B(\theta')+f_S(\theta)\\
\theta' &= \max\{\theta',\theta'\} = f_B(\theta')+f_S(\theta')
\end{aligned} \qquad (5.37)$$

Subtracting the third from the fourth equality in (5.37) we find that $f_S(\theta') = f_S(\theta)$, and subtracting the second from the fourth equality in (5.37) we find that $f_B(\theta') = f_B(\theta)$. But then the first right hand side in (5.37) has to be the same as the three other right hand sides, which contradicts the assumption that the first right hand side equals θ whereas all later right hand sides equal θ'. There is therefore no budget balanced VCG mechanism in this example.

Whenever Corollary 5.1 holds, then Proposition 5.15 gives not only necessary and sufficient conditions for the existence of a budget balanced VCG mechanism, but necessary and sufficient conditions for the existence of any budget balanced dominant strategy incentive compatible mechanism that implements efficient decisions. This is because any such mechanism has to be a VCG mechanism by Corollary 5.1. This observation applies to the bilateral trade model.

Note that the impossibility result that follows from Proposition 5.15 for the bilateral trade model does not rely on an individual rationality constraint. It is therefore not an implication of Proposition 4.12 which shows that only fixed price mechanisms are dominant strategy incentive compatible, ex post budget balanced and individually rational.

In some of the proofs, we have worked with Groves mechanisms and have argued that all such mechanisms, if they are individually rational, have a zero budget balance, or a negative budget balance, in all states of the world. I am not aware of any result that yields this conclusion for a more general class of models.

5.11 Remarks on the Literature

The payoff equivalence result that we presented is taken from Krishna and Maenner (2001). The classic papers on Vickrey-Clarke-Groves mechanisms are Clarke (1971), Groves (1973) and Vickrey (1961). The uniqueness of VCG mechanisms in the sense of Corollary 1 was shown in Green and Laffont (1977) and Holmström (1979). The result on cyclical monotonicity is from Rochet (1987). However, the concept of cyclical monotonicity is originally due to Rockafellar (1970). The discussion of the case that outcomes are lotteries is also taken from Rochet's 1987 paper. The result on the sufficiency of weak monotonicity is due to Bikhchandani et al. (2006). The concept of "positive association of differences", and its characterization are in Roberts (1979). Finally, I have taken Proposition 5.15 from Milgrom (2004), p. 54, where this result is attributed to Bengt Holmström's 1977 Stanford PhD thesis.

5.12 Problems

Chapter 6

Bayesian Mechanism Design: General Theory

.

6.1 Introduction

To generalize our earlier results on Bayesian mechanism design, we need to take two steps. First, as in the last chapter, we need to consider more general sets of alternatives and a more general specification of types and preferences. Second, we need to consider more general distributions of types. This chapter will correspondingly have two parts. The first part, dealing with more general sets of alternatives, types, and preferences, will build on the previous chapter, and will therefore be short and in parts informal. The second part, dealing with correlated types, will be longer, and the nature of the analysis will be quite different from what we have seen before.

In practice, it seems plausible that types are often not independent. For example, if one agent has private information that makes the agent value the object sold in an auction particularly highly, then this agent might think that it is likely that other agents also have information that makes them value the object highly. This motivates the study of a model of Bayesian mechanism design in which types are not independent.

The results in the second part of this chapter will indicate that with dependent types almost all combinations of decision rules and transfer rules can be implemented by a Bayesian incentive compatible direct mechanism. This result is surprising, and it does not seem plausible in practice. The result is best viewed as a paradox. It clearly indicates that some ingredient is missing in the model. In Section 6.4.6 below, we shall begin a discussion of what this missing ingredient might be.

6.2 Set-Up

There are N agents. The set of agents is denoted by $I = \{1, 2, \ldots, N\}$. They have to choose an alternative a out of some set A of mutually exclusive alternatives. Agent i's utility if alternative a is chosen, and agent i pays transfer t_i is:

$$u_i(a, \theta_i) - t_i.$$

Here, θ_i is agent i's type.

We shall employ similar notation as before: The set of possible types of agent i is Θ_i. We denote by θ the vector of types: $(\theta_1, \theta_2, \ldots, \theta_N)$. The set of all possible type vectors is $\Theta \equiv \Theta_1 \times \Theta_2 \times \ldots \Theta_N$. Finally, we write θ_{-i} for a vector θ of types if we leave out agent i's type. The set of all θ_{-i} is Θ_{-i}, which is the cartesian product of the sets Θ_j, leaving out Θ_i.

The sets Θ_i and Θ can, in principle, be abstract sets, but we want to be able to define a probability measure on them. For this, we can think of them as abstract measurable spaces, or, for concreteness, as products of intervals in finite dimensional Euclidean space, or as finite sets. We assume that there is a common prior distribution μ on Θ that is shared by the agents and the principal. We write μ_i for the marginal probability on Θ_i, and we write $\mu(\cdot \mid \theta_i)$ for the conditional probability distribution of θ_{-i} given θ_i. Note that μ is not only the common prior of the agents but it is also the mechanism designer's belief.

We assume that the mechanism designer proposes to the agents a game and a Bayesian equilibrium of that game. The revelation principle applies, and we can restrict attention to direct mechanisms.

Definition 6.1. A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ consists of a mapping

$$q : \Theta \to A,$$

that maps every type vector into a collective decision, and mappings

$$t_i : \Theta \to \mathbb{R},$$

one for each player $i \in I$, that indicate for each type vector the transfer that agent i needs to make.

We call q the "decision rule" and the functions t_i the "payment rules."

For a given direct mechanism we shall write $Q_i(\theta_i)$ for agent i's interim probability distribution on A conditional on agent i's type being θ_i. That is, if $\mathcal{A}$ is a measurable subset of A, then $Q_i(\theta_i)$ assigns to $\mathcal{A}$ the probability

$$Q_i(\mathcal{A} \mid \theta_i) \equiv \int\limits_{\Theta_{-i}} \mathcal{I}_{\mathcal{A}} d\mu(\theta_{-i} \mid \theta_i) \tag{6.1}$$

where $\mathcal{I}_{\mathcal{A}} : \Theta_{-i} \to \{0, 1\}$ is the indicator function that assigns 1 to a type vector θ_{-i} if and only if $q(\theta_i, \theta_{-i}) \in \mathcal{A}$. We also write $T_i(\theta_i)$ for agent i's interim expected transfer payment conditional on agent i's type being θ_i:

$$T_i(\theta_i) \equiv \int\limits_{\Theta_{-i}} t_i(\theta_i, \theta_{-i}) d\mu(\theta_{-i} \mid \theta_i). \tag{6.2}$$

If a direct mechanism is derived from a Bayesian equilibrium of some indirect mechanism, then truth-telling will be a Bayesian equilibrium of the direct mechanism. In that case we call the direct mechanism Bayesian incentive compatible.

Definition 6.2. A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is "Bayesian incentive compatible" if for all $\theta \in \Theta$, all $i \in I$ and all $\theta_i' \in \Theta_i$ we have:

$$\int\limits_A u_i(a, \theta_i) dQ_i(\theta_i) - T_i(\theta_i) \geq \int\limits_A u_i(a, \theta_i) dQ_i(\theta_i') - T_i(\theta_i').$$

Our fist concern will be with the characterization of Bayesian incentive compatible decision rules. We shall later bring in additional considerations, such as individual rationality and budget balance.

6.3 Independent Types

We begin with a statement of the revenue equivalence theorem. As in the previous chapter we use the formulation of Krishna and Maenner (2001), now adapted to the Bayesian setting. We make the same assumptions as in the context of Proposition 5.1 in Chapter 5. The result then says that the interim expected payment rules are uniquely determined up to a constant by the interim decision rules. The logic behind the result below is the same as the logic behind all other revenue equivalence results presented in these notes.

Proposition 6.1. *Suppose that for every $i \in I$, the set Θ_i is a convex subset of a finite dimensional Euclidean space. Moreover, assume that for every $i \in I$ the function $u_i(a, \theta_i)$ is a convex function of θ_i. Suppose that $(q, t_1, t_2, \ldots, t_N)$ is a Bayesian incentive compatible mechanism with interim decision rules $Q_i(\theta_i)$ and interim expected payments $T_i(\theta_i)$ for every $i \in I$ and every θ_i. Let $(q', t'_1, t'_2, \ldots, t'_N)$ be another Bayesian incentive compatible mechanism with interim decision rules $Q'_i(\theta_i)$ and interim expected payments $T'_i(\theta_i)$ for every $i \in I$ and every θ_i. Suppose that*

$$Q'_i(\theta_i) = Q_i(\theta_i)$$

for some $i \in I$ and every $\theta_i \in \Theta_i$. Then there is a number $\tau_i \in \mathbb{R}$ such that

$$T'_i(\theta_i) = T_i(\theta_i) + \tau_i$$

for every $\theta_i \in \Theta_i$.

We now focus on decision rules q that can be implemented in Bayesian incentive compatible direct mechanisms. A first point to note is that all dominant strategy incentive compatible mechanisms are also Bayesian incentive compatible, and therefore VCG mechanisms are Bayesian incentive compatible mechanisms implementing efficient decisions.

To obtain characterizations of *all* incentive compatible decision rules, efficient or not, we can translate the results of the previous chapter into our current setting. These characterizations now hold at the interim, not at the ex post level. To translate these results, we treat the set of all probability distributions over A as the set of alternatives. The decision rule Q_i assigns to every type θ_i of player i a probability distribution $Q_i(\theta_i)$ over A. Agent i evaluates alternatives by calculating their expected utility. Finally, agent i's payment rule is given by T_i. As an example we adapt Proposition 5.4 to our setting.

Proposition 6.2. *A decision rule q is part of a Bayesian incentive compatible direct mechanism $(q, t_1, t_2, \ldots, t_N)$ if and only if q is interim cyclically monotone, that is, for every $i \in I$, for every sequence of length $k \in \mathbb{N}$ of types of agent i, $(\theta_i^1, \theta_i^2, \ldots, \theta_i^k) \in \Theta_i^k$, with $\theta_i^k = \theta_i^1$, we have:*

$$\sum_{\kappa=1}^{k-1} \left(\int_A u_i(a, \theta_i^{\kappa+1}) dQ_i(\theta_i^\kappa) - \int_A u_i(a, \theta_i^\kappa) dQ_i(\theta_i^\kappa) \right) \leq 0.$$

In Chapter 5, we considered the special case in which the type space is single-dimensional. In this special case, monotonicity of the decision rule q was shown to be necessary and sufficient for the existence of a dominant strategy incentive compatible mechanism that implements q. A translation of these results into our setting requires that for every agent $i \in I$ the type set Θ_i is single dimensional where the set of alternatives is the set of all probability distributions over Q. The only cases where this condition is met seem to be those in which A consists of only two alternatives. Therefore, there do not seem to be more general results here that go significantly beyond what we showed in Chapter 3.

Turning to individual rationality and budget balance, it again seems that the result on individual rationality with single dimensional type spaces, Proposition 5.14, has no interesting generalization. Concerning budget balance, the existence of an ex post budget balanced dominant strategy mechanism is, of course, sufficient for the existence of a budget balanced Bayesian incentive compatible mechanism. Thus, Proposition 5.15 provides a sufficient condition for the existence of an efficient Bayesian incentive compatible mechanism.

Combining individual rationality and budget balance, if all individually rational and ex post budget balanced mechanisms have, in all states of the world, either a zero surplus, or a deficit, then it follows that there is no Bayesian incentive compatible, individually rational, and ex post budget balanced mechanism either, provided that the revenue equivalence result Proposition 6.1 holds. This is the logic that we used to prove Propositions 3.7 and 3.12.

For second best considerations in settings in which efficient rules cannot be implemented, the equivalence between ex ante and ex post budget balance that we used in Chapter 3 is useful, and it generalizes in a straightforward way.

Definition 6.3. A direct mechanism $(q, t_1, t_2, \dots, t_N)$ is "ex ante budget balanced" if

$$\int_{\Theta} \left(\sum_{i=1}^{N} t_i(\theta) \right) d\mu(\theta) = 0.$$

The proof of the equivalence of ex ante and ex post budget balance, Proposition 3.6, that we gave in Chapter 3 directly applies here. For convenience we repeat the result.

Proposition 6.3. *For every direct mechanism with decision rule q that is ex ante budget balanced, there is an equivalent direct mechanism with the same decision rule q that is ex post budget balanced.*

Here, two mechanisms are "equivalent" if the have the same decision rule, and if for all agents $i \in I$ and for all types $\theta_i, \theta_i' \in \Theta_i$, agent i's expected transfers, conditional on agent i's type being θ_i and agent i reporting to be type θ_i', is the same in the two mechanisms.

6.4 Correlated Types

6.4.1 Framework

We now turn to the case that the distribution μ reflects some correlation among types. Throughout this part of the notes, we shall assume that each of the sets Θ_i is finite. The results that we discuss in this part of the notes are more easily proved with finite type spaces, although they are also true with infinite type spaces. Note that this is in contrast with earlier sections, where results were more easily obtained with continuous type spaces rather than discrete type spaces. Although the finiteness assumption is thus, in principle, innocuous, it might mislead us when we consider the question what is true generically, and which cases are exceptional cases. We shall return to this point below. We also assume that every $\theta \in \Theta$ has positive probability: $\mu(\theta) > 0$. This, too, simplifies the exposition.

Observe that the definitions and results of Section 6.2 were not restricted to the case of independent types, and therefore we can use them here. We restrict attention to Bayesian incentive compatible direct mechanisms. These are defined as in Section 6.2.

6.4.2 Failure of Revenue Equivalence

In the setup with independent types we have obtained that, under some technical assumptions, the decision rule q determines the interim expected transfer rules T_i for each player $i \in I$ up to a constant. This result, to which we have referred as the "revenue equivalence" result, greatly simplified the problem of describing the set of all incentive compatible mechanisms, and the problem of identifying in this set those mechanisms that are optimal by some criterion.

When types are not independent, then the revenue equivalence result is no longer true. To understand why this is the case, let us look back for a moment at the independent types framework. The key argument that allowed us to establish the payoff equivalence result showed that the derivative of any agent i's interim expected utility with respect to agent i's type depends only on the choice rule, and not on the transfer rules. This point is crucial because it implies that agent i's interim expected utility function is determined up to a constant by the choice rule. This then easily implies, for example, that the expected payments are determined up to a constant by the choice rule.

How did we show that the derivative of interim expected utility with respect to an agent's type depends only on the choice rule? The argument for this is as follows. By the envelope theorem, as we change agent i's type, we may take as given and fixed agent i's type report. A change in the agent's type then affects the agent's valuation of collective decisions, but it does not affect the agent's valuation of expected transfers. Therefore, only collective decisions, but not transfers, enter into the derivative.

Why is an agent's expected transfer not affected by a change in the agent's type if we take the agent's type report as given and fixed? There are two reason for this is. The first is that all types have the same utility function of money, i.e. $u(t_i) = t_i$. The second reason is that all types have the same conditional probability distribution over other agents' type vectors. This is important because it implies that as we change the agent's type, the agent's expected transfer payment does not change, provided we keep the report fixed.

If types are not independent, this last part of the argument is no longer valid. As we change an agent's type, even if we keep the type report fixed, the agent's conditional probability distribution over other agents' types changes. As the agent's transfer payments may depend on other agents' types, the agent's expected transfer payment may change. Thus, not only the decision rule, but also the transfer rule, enters into the derivative of agent i's interim expected utility with respect to type. As a consequence, the same choice rule may be incentive compatible in combination with transfer rules that differ at the interim level by more than an additive constant.

The fact that the payoff equivalence result is not valid implies that we cannot transform the problem of characterizing the incentive compatible direct mechanisms into the problem of characterizing all incentive compatible choice rules. We need to provide a joint characterization of incentive compatible choice rules and transfer rules. A surprising and general characterization is provided in the next subsection.

6.4.3 Characterizing Bayesian Incentive Compatibility

The result that we present in this section relies on a condition regarding the distribution μ that we shall call the "Crémer-McLean condition" as it originates in the work of Crémer and McLean (1988).

Definition 6.4. The probability distribution μ satisfies the *Crémer-McLean condition* if there are no $i \in I$, $\theta_i \in \Theta_i$ and $\lambda_i : \Theta_i \setminus \{\theta_i\} \to \mathbb{R}_+$ for which:

$$\mu(\theta_{-i} \mid \theta_i) = \sum_{\theta_i' \in \Theta_i \setminus \{\theta_i\}} \lambda(\theta_i')\mu(\theta_{-i} \mid \theta_i') \text{ for all } \theta_{-i} \in \Theta_{-i}.$$

The content of this condition is best understood if one thinks of agent i's belief about the other agents' types conditional on agent i's type being θ_i, $\mu(\cdot \mid \theta_i)$, as a vector with as many entries as Θ_{-i} has elements. Agent i's conditional beliefs are described by a collection of vectors of this form, one for each of agent i's type. The Crémer-McLean condition requires that none of these vectors can be written as a convex combination of all the other vectors where the weights are denoted by $\lambda(\theta_i')$. Thus, none of these vectors is contained in the convex hull of the other vectors.

The Crémer-McLean condition is obviously satisfied if the rank of the collection of vectors that describe agent i's conditional beliefs is equal to the number of agent i's types, and hence the vectors are linearly independent. The Crémer-McLean condition is obviously *violated* if at least two of the vectors that describe agent i's conditional beliefs are identical. Thus, the Crémer-McLean condition rules out that agent i's conditional beliefs are independent of his type, as is the case when types are independent.

Under the Crémer-McLean condition, we can obtain the following surprising result:

Proposition 6.4. *Suppose that the distribution μ satisfies the Crémer-McLean condition. Consider any direct mechanism (q, t). Then there is an equivalent direct mechanism (q, t') that is Bayesian incentive compatible; that is:*

1. *the two mechanisms have the same decision rule q;*
2. *the two mechanisms have the same interim expected payments:*

$$\sum_{\theta_{-i} \in \Theta_{-i}} t_i(\theta_i, \theta_{-i})\mu(\theta_{-i} \mid \theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} t_i'(\theta_i, \theta_{-i})\mu(\theta_{-i} \mid \theta_i)$$

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

In words, this result says that *every* direct mechanism can be made Bayesian incentive compatible *without altering the decision rule or interim expected payments*. This is a very permissive result.

As an application of Proposition 6.4 consider the single unit auction environment. Let the decision rule be that the object is allocated to one of the agents with the highest valuation of the agent, and let the payment rule be that the winner of the object pays his true valuation, and that all other agents pay nothing. If the auctioneer could implement this mechanism, he would clearly obtain the largest possible revenue that he could extract from the agents provided that he respects the agents' individual rationality constraint. Proposition 6.4 says that, although this particular direct mechanism is not necessarily Bayesian incentive compatible, one can adjust it so that it does become incentive compatible without changing either the allocation rule or the interim expected payments. If interim expected payments are not altered, of course also the auctioneer's expected revenue is not altered, and hence the auctioneer can extract the entire surplus from trade. Put differently, the auctioneer can achieve the same expected revenue as he can if he directly observes agents' valuations of the object. Agents earn no information rents.

The idea of the proof of Proposition 6.4 is to add to the transfer schemes of the original mechanism (q, t) a transfer scheme that provides agents with incentives to truthfully reveal their beliefs. It is well-known that such incentive schemes exist for risk-neutral subjects. Experimentalists, for example, use such incentive schemes to elicit subjects' beliefs about uncertain events, such as other agents' actions. Moreover, it is easy to see that in such incentive schemes the costs for false reports of one's beliefs can be made arbitrarily large. As no two types have the same beliefs, an agent who truthfully reveals his beliefs also truthfully reveals his type. By making the incentives for truthful revelation of beliefs very large we can undo all possible incentives to lie about one's type in the original mechanism (q, t).

An additional complication arises from the fact that the incentive scheme for belief revelation needs to be such that truthful reports of beliefs generate expected payments of zero, whereas false reports of beliefs generate negative expected payments. This will ensure that the interim expected payments, under truthful revelation of types, will remain unchanged in the transformed mechanism. While incentive mechanisms for truthful belief revelation always exist, the Crémer-McLean condition is used to ensure that the expected payments from truth-telling can be set equal to zero.

How can truthful revelation of beliefs be induced? A very simple incentive scheme is as follows. Suppose the finite set of possible outcomes is Ω, and we want to elicit an experimental subject's belief about the probabilities of different elements of Ω. We ask the subject to announce their probabilities π for the elements of Ω, i.e. $\pi : \Omega \rightarrow [0, 1]$, and we announce that if $\omega \in \Omega$ occurs, the subject makes a payment to the experimenter of $k - \ln(\pi(\omega))$. Here, k is a constant. Thus, for example, if an $\omega \in \Omega$ occurs to which the subject has assigned an extremely small probability, the subject has to pay to the experimenter a comparatively large amount.

Which probabilities will subjects announce? Suppose a subject is risk-neutral. Let $\hat{\pi}$ be the subject's true subjective probabilities. In choosing the report π, the subject will minimize:

$$\sum_{\omega \in \Omega} (k - \ln(\pi(\omega))\hat{\pi}(\omega) \tag{6.3}$$

subject to

$$\sum_{\omega \in \Omega} \pi(\omega) = 1. \tag{6.4}$$

A Lagrange function is:

$$\mathcal{L} = \sum_{\omega \in \Omega} (k - \ln(\pi(\omega)))\hat{\pi}(\omega) + \lambda \left(\sum_{\omega \in \Omega} \pi(\omega) - 1 \right). \tag{6.5}$$

The first order condition for maximizing this requires that for every $\omega \in \Omega$:

$$\frac{\partial \mathcal{L}}{\partial \pi(\omega)} = -\frac{\hat{\pi}(\omega)}{\pi(\omega)} + \lambda = 0 \Leftrightarrow$$
$$\frac{\hat{\pi}(\omega)}{\pi(\omega)} = \lambda. \tag{6.6}$$

Thus, the reported vector of probabilities, $\pi(\omega)$, has to be proportional to the true vector of probabilities, $\hat{\pi}(\omega)$. There is, of course, only one reported vector π that his property, namely the true vector of probabilities, $\hat{\pi}$, itself.

Suppose we added to the transfer rule in the mechanism (q, t) in Proposition 6.4 payments that follow the above payment scheme, i.e. if agent i reports to be type θ_i, and if the agents other than i report type θ_{-i}, then agent i has to pay $t_i(\theta_i, \theta_{-i}) + c(k - \ln \mu(\theta_{-i} \mid \theta_i))$ where k and $c > 0$ are constants. If we choose c sufficiently large, then the incentive to report the probabilities truthfully will override all other incentives that agent i might have, and the mechanism will be Bayesian incentive compatible.

What we have just described is almost, but not quite the mechanism that we will use in the proof of Proposition 6.4. The mechanism that we have described obviously alters agents' interim payments. However, the Crémer-McLean condition is sufficient to allow us to provide strict incentives for truthful revelation of beliefs using a mechanism where an agent's expected payment, if the agent reports his beliefs truthfully, is exactly zero. This can be deduced from the Crémer-McLean condition using Farkas' Alternative:[19]

Proposition 6.5. *Let A be an $n \times m$ matrix and let $b \in \mathbb{R}^n$. Either the equation*

$$Ax = b \qquad \textit{has a solution } x \geq 0$$

or (exclusive)

$$y^T A \geq 0, \quad y^T b < 0 \qquad \textit{has a solution } y.$$

To apply Farkas' Alternative, we take some type θ_i of agent i as given and fixed. We take b to be the the column vector of agent i's beliefs if agent i is type θ_i, and we take A to be a matrix of dimensions $\#\Theta_{-i} \times (\#\Theta_i - 1)$ where each column of A represents agent i's conditional beliefs if agent i is of one of the types other than θ_i. The Crémer-McLean condition says that the first of the two conditions in Farkas' Alternative is not satisfied. We deduce that the second condition holds. The column vector y, which has $\#\Omega_{-i}$ entries, represents agent i's payments if agent i reports type θ_i, and all other agents report to be type θ_{-i}. The first condition says that agent i's expected payment, conditional on any type other than type θ_{-i}, is non-negative. On the other hand, conditional on the type θ_i, agent i's expected payment is negative, that is, he expects to receive a payment. By subtracting a constant from all payments, we can achieve that conditional on type θ_i, agent i expects a zero payment, and conditional on all other types he expects to make a strictly positive payment.

We can repeat the construction of the previous paragraph for each type θ_i of agent i. Putting together all the transfer vectors y that we obtain in this way, we construct a transfer scheme for agent i with the desired properties, that incentives to report true beliefs are strict, and that expected payments conditional on truth-telling are zero. We can add this transfer scheme, times a positive constant, to agent i's transfers in our original mechanisms (q, t), and, if the constant is sufficiently large, agent i's incentives to report his beliefs, and thus his type, truthfully, override all other incentives that agent i might have in (q, t). We can proceed in a similar way for all agents i. This completes the proof of Proposition 6.4.

[19] Franklin (1980), p.56.

	1	2	3
1	$\frac{4}{20}$	$\frac{2}{20}$	$\frac{1}{20}$
2	$\frac{2}{20}$	$\frac{2}{20}$	$\frac{2}{20}$
3	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{4}{20}$

Figure 6.1: Joint probability distribution of types.

	1	2	3
1	$(\frac{1}{2}, \frac{1}{2})$	(0,1)	(0,1)
2	(1.0)	$(\frac{1}{2}, \frac{1}{2})$	(0,1)
3	(1,0)	(1,0)	$(\frac{1}{2}, \frac{1}{2})$

Figure 6.2: Decision Rule

6.4.4 A Numerical Example

The numerical example in this Subsection underlines the general message of this section that one should view the Crémer McLean result as a paradox rather than a guidance to the construction of mechanisms that could work in practice.

Example 6.1. *We consider a single unit auction example. There are two agents. Each of the two agents has one of three types:* $\Theta_1 = \Theta_2 = \{1, 2, 3\}$*. An agent's type is the agent's valuation for the good. The joint probability distribution of types is shown in Figure 6.1. In this figure, rows correspond to types of player 1, and columns correspond to types of player 2.*

Suppose that the auctioneer wants to allocate the object to the agent with the highest valuation, where ties are broken randomly. Moreover, the auctioneer would like the winner of the object to pay her reservation price, that is, her type. This corresponds to the allocation and payment rules shown in Figures 6.2 and 6.3.

In the decision rule, the first entry in each box indicates the probability with which agent 1 obtains the object, and the second entry indicates the probability with which player 2 obtains the object. In the payment rule, the

	1	2	3
1	$(\frac{1}{2}, \frac{1}{2})$	(0,2)	(0,3)
2	(2,0)	(1, 1)	(0,3)
3	(3,0)	(3,0)	$(\frac{3}{2}, \frac{3}{2})$

Figure 6.3: Payment Rule

	1	2	3
1	-1	1	2
2	1	-2	1
3	2	1	-1

Figure 6.4: Belief revelation

first entry in each box shows the payment by player 1, and the second entry shows the payment by player 2.

The rule that we have described is clearly not incentive compatible. Types 2 and 3 of each player have an incentive to pretend that they are lower types. That will reduce the probability with which they get the object, but at least it will give them a positive surplus if they obtain the object. We now want to demonstrate Proposition 1 for this example, and construct a payment rule that leaves interim payments unchanged, and that makes the decision rule incentive compatible.

As explained in Subsection 6.4.3, we start by constructing transfer rules that give each player strict incentives to reveal their true beliefs about the other player's type, and that give expected utility of zero to each player if they reveal their beliefs truthfully. For player 1 such a transfer rule is indicated in Figure 6.4.

In Figure 6.4, positive numbers are payments by player 1, and negative numbers are payments to player 1. The idea of the payment rule is that agent 1 is rewarded if the type that she reports is the same as that of player 2, but she has to make a payment if the type that she reports is different from the one that player 2 reports. This reflects that agents' types are positively

	1	2	3
1	$(-\frac{5}{2}, -\frac{5}{2})$	(3,5)	(6,9)
2	(5,3)	(-5, -5)	(3,6)
3	(9,6)	(6,3)	(-$\frac{3}{2}, -\frac{3}{2}$)

Figure 6.5: Modified payment rule

correlated in this example. We omit the simple check that this payment rule has the required properties.

A symmetric rule can, of course, be used for player 2. We shall now add a positive multiple of the transfers in Figure 6.4 to the transfers in Figure 6.3. We need to determine by how much we need to multiply the transfers in Figure 6.4. We need to overcome all incentives to deviate in the mechanism of Figures 6.2 and 6.3. One can calculate that 3 is the smallest integer by which we can multiply all payments in Figure 6.4 and obtain a transfer rule that eliminates all incentives to lie in the mechanism of Figures 6.2 and 6.3. If we multiply the transfer payments in Figure 6.4 by 3 and add them to the transfer payments in Figure 6.3 we get the payments shown in Figure 6.5.

One can now verify that the auction that has the allocation rule indicated in Figure 6.2 and the transfer rules of Figure 6.5 extracts the full surplus for the auctioneer, and is incentive compatible and individually rational. Note that individual rationality only holds at the interim, not at the ex post level.

6.4.5 Individual Rationality and Budget Balance

We will want to combine Bayesian incentive compatibility with other requirements, such as individual rationality and ex post budget balance. We begin by defining individual rationality.

Definition 6.5. A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is "individually rational" if for every agent i and every type $\theta_i \in \Theta_i$ we have:

$$\int_A u_i(a, \theta_i) dQ_i(\theta_i) - T_i(\theta_i) \geq 0$$

In the transformation that was described in Subsection 6.4.3, whenever the mechanism with which we started is individually rational, then also the transformed mechanism is individually rational. Thus, any pair of decision and transfer rules that are individually rational can be made incentive compatible and individually rational, provided the Crémer-McLean condition holds.

We next consider ex post budget balance. The construction that demonstrates the equivalence of ex ante and ex post budget balance and proves Proposition 6.3 requires that types are independent. That result therefore does not straightforwardly generalize to the context of correlated types. However, Kosenok and Severinov (2008) have shown that an additional condition for the prior distribution of types that goes beyond the Crémer McLean condition, guarantees that budget balance can always be achieved. The additional condition is called "identifiability."

Definition 6.6. The probability distribution μ satisfies the *identifiability condition* if, for all distributions $\nu \neq \mu$ such that $\nu(\theta) > 0$ for all $\theta \in \Theta$, there is at least one agent i and one type $\theta_i \in \Theta_i$ such that for any collection of nonnegative coefficients $\left(\lambda_{\theta'_i}\right)_{\theta'_i \in \Theta_i}$ we have:

$$\nu(\theta_{-i} \mid \theta_i) \neq \sum_{\theta'_i \in \Theta_i} \lambda_{\theta'_i} \mu(\theta_{-i} \mid \theta'_i)$$

for at least one $\theta_{-i} \in \Theta_{-i}$.

Intuitively, this condition says that for any alternative distribution ν of types, there is at least one agent and one type of that agent such that this agent cannot randomize over reports in a way that makes the conditional distribution of all other types under ν indistinguishable from the conditional distribution of all other types μ. Kosenok and Severinov (2008) prove:

Proposition 6.6. *Suppose that the distribution μ satisfies the Crémer-McLean and the identifiability condition. Consider any direct mechanism $(q, t_1, t_2, \ldots, t_N)$ that is individually rational and ex ante budget balanced. Then there is a Bayesian incentive compatible and ex post budget balanced mechanism $(q, t'_1, t'_2, \ldots, t'_N)$ that is equivalent, that is:*

1. *the two mechanisms have the same decision rule q;*

2. *the two mechanisms have the same interim expected payments:*

$$\sum_{\theta_{-i} \in \Theta_{-i}} t_i(\theta_i, \theta_{-i}) \mu(\theta_{-i} \mid \theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} t'_i(\theta_i, \theta_{-i}) \mu(\theta_{-i} \mid \theta_i)$$

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

Kosenok and Severinov's construction that proves this result is quite different from Crémer and McLean's construction, and we omit a discussion of this construction.

6.4.6 Discussion

A number of authors have investigated the question how the set-up presented by Crémer and McLean needs to be changed to obtain less paradoxical results. Much of this discussion has focused on the particular case of a single unit auction in which the decision and transfer rules to be implemented are the full extraction rules, as illustrated by our numerical example. However, it seems plausible that the results of this literature generalize to other settings.

Suppose that in the Crémer and McLean setting the prior distribution of types is close to a product measure, i.e. types are close to being independent. Then a belief revelation scheme that provides incentives strong enough to outdo all incentives to lie will require large payments by agents, and it will expose agents to significant risk. Thus, if agents are either liquidity constrained or risk averse, such a mechanism might be impossible, or agents might require compensation for the risk that reduces the attractiveness of the mechanism to the mechanism designer. This intuition has been formalized by Robert (1991).

The fact that in a belief revelation scheme the reports of all agents except agent i determine how much agent i has to pay suggests that there may be an incentive to collude. Laffont and Martimort (2000) have shown that in environments in which types are close to independent collusion might prevent the mechanism designer from full surplus extraction.

Heifetz and Neeman, in a recent series of papers, have pursued a different line of argument. They point out that the construction of Crémer and McLean requires that there is a one-to-one relation between an agent's preferences and an agent's beliefs about other types. This precisely is the reason why a belief extraction mechanism can help with implementation. Neeman (2004) and Heifetz and Neeman (2006) have shown that information structures with this property are in some sense rare among all conceivable common prior information structures. While it is often true that *for fixed finite type space* generic probability distributions will satisfy Crémer and McLean's conditions, the same distributions are, as Heifetz and Neeman show, a very small subset of the infinite dimensional space of general information structures.

6.5 Complete Information

6.6 Efficiency and Utilitarianism

6.7 Remarks on the Literature

Proposition 6.4 is adapted from Crémer and McLean (1988). My favorite reference for Farkas' Lemma, which is central to the proof of Proposition 1, is Franklin (1980). Franklin's proof is not much more than half a page long, and all that you need to follow the proof is a basic knowledge of separating hyperplane theorems. Proposition 6.6 is a combination of Theorem 1 and Corollary 1 in Kosenok and Severinov (2008).

6.8 Problems

A seller has a single indivisible object to sell. The seller values the object at zero, and seeks to maximize his expected revenue. There are two potential buyers: $i = 1, 2$. Buyer i's von Neumann Morgenstern utility equals $v_i - t_i$ if he obtains the good and pays t_i, and his utility equals $-t_i$ if he does not obtain the good and pays t_i.

For each buyer i, the valuation v_i is a random variable that only buyer i observes, but that the other buyer and the seller don't observe. The random variable v_1 takes only two possible values: $v_1 \in \{3, 4\}$. The random variable v_2 takes only two possible values: $v_2 \in \{1, 2\}$. Note that the two buyers are *not* symmetric. The joint distribution of v_1 and v_2 is given by the following table:

	$v_2 = 1$	$v_2 = 2$
$v_1 = 3$	0.25+ε	0.25-ε
$v_1 = 4$	0.25-ε	0.25+ε

Figure 6.6: Distribution of Valuations

Here, ε is a constant that satisfies $0 < \varepsilon < 0.25$. Each entry indicates the probability with which v_1 takes the value indicated in the row, and at the same time v_2 takes the value indicated in the column. The distribution indicated in this table is common knowledge among the buyers and the seller.

(a) Verify that the Crémer McLean condition holds for buyer 1.

(b) Construct a payment scheme that provides buyer 1 with incentives for truthful revelation of his beliefs about buyer 2's type, and that implies that, if he truthfully reveals his beliefs, his expected transfer payment conditional on each of his types will be zero.

(c) Use the payment scheme that you found in part (b) to construct a Bayesian incentive compatible selling mechanism that always allocates the object to buyer 1, and that implies that buyer 1's expected payment to the seller equals his expected valuation of the object. The selling mechanism that you construct should also offer each type of each agent an interim expected utility of at least zero in equilibrium.

(d) Investigate the limit for $\varepsilon \to 0$ of the transfer payments in the selling mechanism that you obtained in part (c).

Chapter 7

Non-Transferrable Utility

7.1 Introduction

So far we have assumed that all agents' utility is additively separable in an allocative decision and money, and moreover that all agents are risk-neutral in money. This is a very restrictive assumption. It is easy to imagine situations in which the assumption is not satisfied. The simplest case is that agents are not risk neutral in money. Another way in which the assumption might be violated is that there are interactions between money and the allocative decision. Finally, it might be the case that we are considering situations, such as voting, in which monetary payments are typically not invoked to provide incentives.

One implication of the assumption of additively separable utility and risk neutrality is that monetary payments that are made by one agent and received by another agent are welfare neutral as long as we allocate the same welfare weight to all agents. By contrast, in a more general model, welfare depends not only on the allocative decision but also on the distribution of money among agents. We shall refer to the case that we have considered so far as the case of "transferrable utility" and we shall refer to the case considered in this chapter as the case of "non-transferrable utility."

On the specter of possible assumptions, we shall move in this chapter to the opposite extreme of the assumption that we have made so far. We shall consider situations in which there is some arbitrary set of possible collective decisions, and, at least initially, no particular assumption is made about agents' preferences over these decisions. Each collective decision is interpreted as a decision about all issues relevant to agents, including the allocation of money.

The emphasis of theoretical work in this area has been on dominant strategy incentive compatibility. We shall start with this part of the literature. There are a vast number of theoretical results in this area. We shall offer a discussion of only an eclectic selection from these results. Later, we shall briefly discuss some work on Bayesian incentive compatibility without transferrable utility.

7.2 The Gibbard Satterthwaite Theorem

7.2.1 Set Up

There is a finite set of agents, $i \in I = \{1, 2, \ldots, N\}$. These agents have to choose one alternative from a finite set A of mutually exclusive alternatives. Each agent i has a preference relation R_i over A. We assume that R_i is a linear order; that is, it is complete, transitive, reflexive, and the only indifference is among identical elements. We denote the strict order derived from R_i by P_i. We read "aR_ib" as "a is weakly preferred to b," and we read "aP_ib" as "a is strictly preferred to b." The set of all linear orders over A is denoted by $\mathcal{R}$. We write R for the list of all agents' preference relations: $R = (R_1, R_2, \ldots, R_N)$, and we write R_{-i} for the list of all agents' preference relations leaving out agent i's: $R_{-i} = (R_1, R_2, \ldots, R_{i-1}, R_{i+1}, \ldots, R_N)$.

We consider a mechanism designer who does not know the agents' preference relations, but who determines the rules of the strategic interaction among the agents by which an alternative from A is chosen. The mechanism designer can thus construct an extensive game with outcomes in A. We study in this section the case in which the mechanism designer seeks to construct a game such that every player, with every conceivable preference relation in $\mathcal{R}$, has a dominant strategy in the sense in which we used this phrase in earlier chapters of these notes. The revelation principle then applies, and we can restrict attention to direct mechanisms.

Definition 7.1. A "direct mechanism" is a function $f : \mathcal{R}^N \to A$.

In the literature direct mechanisms in the sense of definition 7.1 are sometimes also called "social choice functions." To maintain consistency with earlier parts of these notes, we shall speak of "direct mechanisms."

Definition 7.2. A direct mechanism f is "dominant strategy incentive compatible" if for every agent $i \in I$ and all preference relations R_i, R_i' in $\mathcal{R}$:

$$f(R_i, R_{-i}) R_i f(R_i', R_{-i})$$

In the literature dominant strategy incentive compatibility is sometimes also referred to as "strategy proofness." In this section we shall characterize direct mechanisms that are dominant strategy incentive compatible. Before we proceed we make two further remarks about the set up described here. The first concerns the role of randomization in this section. The reader might notice that we have not introduced probability distributions over the set A. But, of course, the elements of set A could be outcomes that are not deterministic, but stochastic. If we have this case in mind, then it might seem more natural to have an infinite set A rather than a finite set. But it is not immediately obvious that in practice an infinite number of probability distributions can be implemented. Physical limitations might well force us to choose from a finite set of probability distributions, and the reader can think of A as the set of all these distributions. Thus, our setup does not really rule out randomization.

The second point that needs emphasis is that the domain of our direct mechanism contains *all* linear orders of A. In practice we might have some knowledge about agents' preferences over A, and can therefore restrict attention to mechanisms that perform as we want them to perform only for some, but not for all preference vectors. Natural restrictions may be that all agents rank either all elements of some subset $\hat{A}$ of A higher than all elements of the complement $A \setminus \hat{A}$, or that they have the opposite ranking. This would seem plausible if the elements of A are candidates for a political office, and if the candidates in $\hat{A}$ have the opposite ideology from candidates in $A \setminus \hat{A}$. Another natural restriction may be that agents' preferences are of the von Neumann Morgenstern form. This would seem natural if A consists of lotteries, as suggested in the previous paragraph. We shall consider later in this Chapter the implications of some such domain restrictions, but first we explore the implications of assuming an unrestricted domain.

7.2.2 Statement of the Result and Outline of the Proof

Definition 7.3. A direct mechanism f is called "dictatorial" if there is some individual $i \in I$ such that for all $R \in \mathcal{R}^N$:

$$f(R) R_i a \text{ for all } a \in A.$$

Proposition 7.1. *Suppose that A has at least three elements, and that the range of f is A. A direct mechanism f is dominant strategy incentive compatible if and only if it is dictatorial.*

This is one of the most celebrated results in the theory of mechanism design. It is due to Gibbard (1973) and Satterthwaite (1975), and is called the "Gibbard-Satterthwaite Theorem." The Gibbard-Satterthwaite theorem is a paradoxical result. It shows that the requirements that the requirements of dominant strategy incentive compatibility and unlimited domain together are too strong. In later parts of this Chapter, we shall discuss weaker requirements that lead to more positive results.

In this result the assumption that the range of f equals A rather than being a strict subset of A is immaterial. If the range of f is a strict subset of A, we can re-define the set of alternatives to be the range of f. Alternatives that are not in the range of f will never be chosen, and therefore agents' preferences over these alternatives cannot influence the outcome. We can therefore analyze the situation as if such alternatives didn't exist.

The assumption that A has at least three elements is important. Without this assumption the result is not true. Consider the only remaining non-trivial case, that is, the case that A has two elements. In that case many mechanisms are dominant strategy incentive compatible. For example, giving each agent the opportunity to vote for one of the two alternatives, and then choosing the alternative with the highest number of votes, with some arbitrary tie breaking rule, is dominant strategy incentive compatible, and clearly not dictatorial.

The "sufficiency part" of Proposition 7.1 is obvious. We therefore focus on proving the "necessity part." Our presentation of this proof is based on Reny (2001). We shall proceed in two steps. First, we show that every dominant strategy incentive compatible, direct mechanism is monotone. We also report some simple implications of monotonicity. The core of the proof is then in the next subsection, where we show in a second step that every monotone direct mechanism is dictatorial.

Definition 7.4. A direct mechanism f is "monotone" if, whenever $f(R) = a$, and for every agent i and every alternative b the preference relation R'_i ranks a above b if R_i does, then $f(R') = a$. Formally:

$$f(R) = a \text{ and for all } i \in I: \; [aR'_i b \text{ for all } b \in A \text{ such that } aR_i b] \Rightarrow f(R') = a.$$

Proposition 7.2. *If f is dominant strategy incentive compatible, then it is monotone.*

Proof. Suppose first that the preference profiles R and R' referred to in Definition 7.4 differed only in the i-th component. Suppose also $f(R') =$

$b \neq a$. Because with preference R_i it is a dominant strategy of agent i in mechanism f to report R_i truthfully, we must have: aR_ib. Because the ranking of a does not fall as we move from R_i to R_i', this implies: $aR_i'b$. But then, with true preference R_i', agent i has an incentive to report preference R_i. This contradicts dominant strategy incentive compatibility. Thus we have arrived at a contradiction, and we can conclude $f(R') = a$. If R' and R differ in more than one component, then we apply the above argument successively to each component in which R' and R differ. Again, we arrive at the conclusion $f(R) = a$. □

Now we introduce two simple implications of monotonicity.

Definition 7.5. A direct mechanism f is "set-monotone" if, whenever $f(R) \in B$ for some $B \subseteq A$, and for every agent i, preference relation R_i' differs from preference relation R_i only regarding the ranking of elements of B, then $f(R') \in B$. Formally:

$$f(R) \in B \text{ and } [aR_i'a' \Leftrightarrow aR_ia' \text{ whenever } a \notin B \text{ or } a' \notin B] \Rightarrow f(R') \in B.$$

Proposition 7.3. *If f is monotone then f is set-monotone.*

Proof. Let f be monotone, and let R' differ from R only regarding the ranking of elements of B. Assume $f(R') = a \notin B$. Then monotonicity implies that $f(R) = a$, which contradicts $f(R) \in B$. □

Definition 7.6. A direct mechanism f "respects unanimity" if, whenever an alternative a is at the top of every individual's preference relation, then a is chosen by f:

$$aR_ib \text{ for all } i \in I \text{ and } b \in A \Rightarrow f(R) = a.$$

Proposition 7.4. *If f is monotone and the range of f is the A, then f respects unanimity.*

Proof. Consider any $a \in A$. Because the range of f is A, there is some R such that $f(R) = a$. Now raise a to the top of everyone's ranking. By monotonicity, the social choice remains a. Now re-order alternatives below a in arbitrary ways. Again, by monotonicity, the choice must remain a. This proves that f respects unanimity. □

7.2.3 Every monotone direct mechanism is dictatorial

Proposition 7.5. *Suppose that A has at least three elements, and that the range of f is A. If f is monotone, then it is dictatorial.*

Proof. We shall show that for every alternative $a \in A$, there is a "dictator for a," i.e., there is an agent i such that whenever a is at the top of i's ranking, then a is chosen. If there is such a dictator for every alternative a, then the dictator must be the same individual i for every alternative. Otherwise, if the dictator for a had a at the top of her ranking, and the dictator for $b(\neq a)$ had b at the top of his ranking, the outcome would not be well-defined. We conclude that the dictator must be the same agent i for every alternative, and therefore this agent i is a dictator.

Now fix an alternative $a \in A$. To show that there is a dictator for a, it is sufficient to find one preference profile where a is at the top of some agent i's ranking, but at the bottom of everybody else's ranking, and where the social choice is a. One such profile is shown in Figure 7.1. In this figure, and in subsequent figures, each column corresponds to one player's ranking of alternatives in A, with the highest ranked alternative at the top. Figure 7.1 shows a specific ranking, involving alternatives other than a, but for the moment this is of no relevance to the argument. In Figure 7.1 agent n ranks a top, but all other agents rank a bottom. Finding one such profile in which the social choice is a is sufficient because every other profile in which agent n ranks a at the top can be obtained from the one shown in Figure 7.1 by changing preferences without moving any alternative above a, and therefore by monotonicity the social choice for every other profile in which agent n ranks a at the top is a.

We shall arrive at the conclusion that the social choice in Figure 7.1 has to be a through a sequence of steps starting with the profile shown in Figure 7.2. In the profile in Figure 7.2 the social choice has to be a because f respects unanimity. Now suppose that we move b up in agent 1's ranking, until it is just below a, as shown in Figure 7.3. Then, by monotonicity, the social choice has to remain a.

Now suppose that we move b one step further up, to the top of agent 1's ranking, as shown in Figure 7.4. Then set-monotonicity as defined above (setting $B = \{a, b\}$) implies that the social choice is either a or b.

We will now identify an agent n of whom we shall show that he or she is a dictator for b. If the social choice is b in Figure 7.4, then we set $n = 1$. However, if the social choice remains a in Figure 7.4, then we repeat the

R_1	…	R_{n-1}	R_n	R_{n+1}	…	R_N
.		.	a	.		.
.		.	c	.		.
.		.	b	.		.
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
c	…	c	.	c	…	c
b	…	b	.	b	…	b
a	…	a	.	a	…	a

Figure 7.1: Social choice is a

R_1	…	R_{n-1}	R_n	R_{n+1}	…	R_N
a	…	a	a	a	…	a
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
b	…	b	b	b	…	b

Figure 7.2: Social choice is a

R_1	…	R_{n-1}	R_n	R_{n+1}	…	R_N
a	…	a	a	a	…	a
b	…	a	a	a	…	a
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
.		b	b	b	…	b

Figure 7.3: Social choice is a

R_1	...	R_{n-1}	R_n	R_{n+1}	...	R_N
b	...	a	a	a	...	a
a	...	a	a	a	...	a
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
.		b	b	b	...	b

Figure 7.4: Social choice is a or b

R_1	...	R_{n-1}	R_n	R_{n+1}	...	R_N
b	...	b	a	a	...	a
a	...	a	b	.	...	.
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
.		.	.	b	...	b

Figure 7.5: Social choice is a

same procedure for agent 2, etc. The first agent for whom the social choice switches from a to b is the agent whom we identify as our candidate dictator n. There has to be one such agent because after we have worked our way through all agents we arrive at a profile where all agents put alternative b at the top of their ranking, and f respects unanimity so that the social choice is b. In Figures 7.5 and 7.6 we show the generic situation for agent n before and after the social choice switches from a to b.

Figures 7.7 and 7.8 show the same profiles as Figures 7.5 and 7.6 except

R_1	...	R_{n-1}	R_n	R_{n+1}	...	R_N
b	...	b	b	a	...	a
a	...	a	a	.	...	.
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
.		.	.	b	...	b

Figure 7.6: Social choice is b

R_1	...	R_{n-1}	R_n	R_{n+1}	...	R_N
b	...	b	a	.	...	.
,	...		b	.	...	.
.		.	.	.		.
.		.	.	.		.
.		.	.	a		a
a		a	.	b	...	b

Figure 7.7: Social choice is a

R_1	...	R_{n-1}	R_n	R_{n+1}	...	R_N
b	...	b	b	.	...	.
.	...	.	a	.	...	.
.		.	.	.		.
.		.	.	.		.
.		.	.	a		a
a		a	.	b	...	b

Figure 7.8: Social choice is b

that we have moved alternative a to the bottom of the ranking for agents $i < n$, and we have moved them to the second position from the bottom for agents $i > n$. We now argue that in these profiles the social choices have to be the same as in Figures 7.5 and 7.6. For Figure 7.8 this follows from Figure 7.6 and monotonicity that the choice has to be b. Now consider the profile in Figure 7.7. Comparing Figure 7.7 and Figure 7.8 we can conclude from set monotonicity (with $B \equiv \{a, b\}$) that the social choice in Figure 7.7 has to be either a or b. But if the alternative chosen were b, then it would also have to be b in Figure 7.5, by monotonicity. Therefore, the choice has to be a in Figure 7.7.

Now consider the preferences in Figure 7.9. The position of alternative a has not changed relative to other alternatives in comparison to Figure 7.7, and therefore the social choice has to be a.

Finally, compare Figure 7.9 to Figure 7.1. The choice in Figure 7.1 has to be a or b, by set monotonicity, setting $B = \{a, b\}$. But if the choice were b in Figure 7.1, then we could move alternative c to the top of everyone's preferences, and by monotonicity the choice would still have to be b, which would contradict that f respects unanimity. Therefore, the choice in Figure

R_1	...	R_{n-1}	R_n	R_{n+1}	...	R_N
.		.	a	.		.
.		.	c	.		.
.		.	b	.		.
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
c	...	c	.	c	...	c
b	...	b	.	a	...	a
a	...	a	.	b	...	b

Figure 7.9: Social choice is a

7.1 has to be a. This concludes the proof.

□

7.3 Dominant Strategy Incentive Compatibility On Restricted Domains

Two natural ways of relaxing the stringent requirements of Proposition 7.1 so that more positive results obtain are firstly to consider a more restricted domain of preferences, and secondly to consider a less demanding solution concept than dominant strategies. In this section, we shall discuss the former approach, whereas in the next section we discuss the latter approach. In this section, thus, the set of preferences for any individual i that we are considering is no longer the complete set $\mathcal{R}$, but some subset $\hat{\mathcal{R}}$ of $\mathcal{R}$.

The best known restriction on the domain of preferences that allows dominant strategy incentive compatible mechanisms that are not dictatorial is single-peakedness. Suppose the alternatives in A are labeled with the integers $1, 2, \ldots, K$. A preference relation R_i of agent i is called "single-peaked" if there is a $k(i) \in \{1, 2, \ldots, K\}$ such that (i) agent i prefers alternative $k(i)$ to all other alternatives: $k(i)R_ij$ for all $j \in A$, and (ii) agent i's preferences decline monotonically "to the left" and "to the right" of $k(i)$, that is: if $\ell \geq k$ then $\ell R_i \ell + 1$ and if $\ell \leq k$ then $\ell - 1 R_i \ell$. Denote by $\hat{\mathcal{R}}$ the set of all single-peaked preferences. The definition of this set depends on how the alternatives have been labeled with numbers. We keep this labeling fixed in this section, and, for simplicity, don't reflect the dependency of the set $\hat{\mathcal{R}}$ on the labeling in our notation. The restricted domain of single-peaked preferences is then $\hat{\mathcal{R}}^N$.

The domain of single-peaked preferences may appear natural in a the following environment. Alternatives are candidates for some political position. Direct mechanisms are methods for selecting one candidate out of a set of candidates. Candidates are labeled according to their position on the "left-right" spectrum, and the alternative $k(i)$ reflects voter i's ideal position on this spectrum. It then appears plausible that preferences decline monotonically as candidates are further away from agent i's ideal position.

Proposition 7.6. *If preferences are single-peaked, then there are dominant strategy direct mechanisms that are not dictatorial.*

Proof. Suppose N is odd. For every agent i, denote by $k(i)$ the alternative that is ranked highest according to R_i. The rule f that picks the alternative a_m, where m is the median of the vector $(k(1), k(2), \ldots, k(N))$, is dominant strategy incentive compatible. Consider any agent i, and take the other agents' reported preferences R_{-i} as given and fixed. The median of $(k(1), k(2), \ldots, k(N))$ will be between the $\frac{N-1}{2}$-th largest and the $\frac{N+1}{2}$-th largest most preferred alternative among the most preferred alternatives of the $N-1$ agents other than i. Denote these by m_- and m_+. The median will be these two numbers, independent of what agent i reports. If the alternative that agent i most prefers is between these two numbers, then agent i can ensure by reporting his preferences truthfully that his most preferred alternative is chosen. If agent i's most preferred alternative is lower than m_-, then by reporting her preferences truthfully agent i ensures that alternative m_-, her most preferred alternative from the range of possible medians, is chosen. If agent i's most preferred alternative is higher than m_+, then by reporting her preferences truthfully, agent i can ensure that alternative m_+, her most preferred alternative from the range of possible medians, is chosen. Thus, in either case agent i, by reporting her preferences truthfully, ensures that her most preferred alternative from the range of possible medians is chosen. Thus, the rule is dominant strategy incentive compatible, and it is obviously non-dictatorial.

Now suppose that N is even. Then we can use the same rule as in the case that N is odd, with the following small modification. We arbitrarily pick some alternative $a \in A$ and pretend that there was some $N+1$-th agent, so that the total number of agents is again odd, and that this agent had expressed a preference that lists a at the top. The same argument as above proves that this rule is dominant strategy incentive compatible. $\square$

Other restricted domains on which dominant strategy incentive compatibility does not imply dictatorships are, of course, those that we studied in Chapter 4. In those environments we assumed that agents' utility functions were of a particular form, for example, additive separability and risk neutrality. This domain restriction lead to the dominant strategy incentive compatibility of simple rules such as the fixed price rule for bilateral trade.

A particularly interesting domain restriction might be natural in the case in which the set A consists of all lotteries over some given finite set of outcomes. In this case one might assume that individuals' preferences satisfy the von Neumann Morgenstern axioms, and can therefore be represented by Bernoulli utility functions. One might hope that this domain restriction allows the implementation of social choice functions other than dictatorial social choice functions. However, a theorem due to Hylland (1980) shows that in this setting, without further domain restrictions, the only direct mechanisms that respect unanimity and are dominant strategy incentive compatible are random dictatorships. In random dictatorships each of the N agents is assigned a probability, and is then picked with this probability as the dictator, i.e. the alternative is picked that maximizes this individual's preference relation.

7.4 Bayesian Incentive Compatibility

Although much of the literature on mechanism design with non-transferrable utility has focused on dominant strategy incentive compatibility, there seems to be no fundamental reason why we should not Bayesian incentive compatibility as well. To explore this approach, Börgers and Postl (2005) consider a stylized model of compromising. There are two agents, $N = 2$, and three alternatives: $A = \{a, b, c\}$. It is commonly known that agent 1 ranks the alternatives as a, b, c. Agent 1's Bernoulli utility of alternative a is 1, and her Bernoulli utility of alternative c is 0, but her Bernoulli utility of the intermediate alternative b is θ_1, agent 1's privately observed type. Agent 2 is analogous, except that agent 2 ranks alternatives in the opposite order c, b, a. A direct mechanism maps the vector (θ_1, θ_2) of agents' types into a probability distribution over the set of alternatives A. We assume that θ_1 and θ_2 are stochastically independent variables that are both distributed on the interval $[0, 1]$ with the same cumulative distribution function F with density f where $f(\theta_i) > 0$ for all $\theta_i \in [0, 1]$. Each agent i observes θ_i but not θ_j where $j \neq i$, and the mechanism designer observes neither of the two types.

Börgers and Postl (2005) interpret the alternative B as a "compromise" as both agents rank B as their middle alternative, even though the two agents have opposing preferences. The focus of the paper is on the question of whether there is an incentive compatible mechanism that implements the alternative B whenever it is efficient, that is, whenever $\theta_1 + \theta_2 \geq 1$. This is a natural notion of efficiency because the sum of utilities from the compromise B is $\theta_1 + \theta_2$, whereas alternatives A and C both yield a sum of utilities of 1. To induce agents to correctly reveal their types the mechanism designer can choose the probabilities of the individuals' favorites, A and C, judiciously. These probabilities can be used as an instrument for providing incentives because they are welfare neutral: Both alternatives yield exactly the same amount of welfare.

Börgers and Postl (2005) note that their framework is equivalent to a public goods model similar to the one we considered in Section 3.3. One can think of the "compromise" as a public good for which both agents have to give up probability of their preferred alternative. The cost-function is one-to-one: for each unit of the public good (i.e. probability of the compromise) produced some agent has to give up one unit of money (i.e. probability of their preferred alternative). An important difference with the public goods models that we considered in Section 3.3 is that agents are liquidity constrained: probabilities have to be between zero and one, and therefore agents cannot give up more than sum upper bound of probability of their preferred alternative.

Using arguments similar to our arguments in Section 3.3, Börgers and Postl (2005) show that no Bayesian incentive compatible mechanism implements the first best, i.e. choice of the compromise with probability 1 if and only if $\theta_1 + \theta_2 \geq 1$. The problem of finding a second best is complicated by the fact that agents face liquidity constraints. Therefore, to determine the second best, one cannot use the approach that was explained in Section 3.3. Börgers and Postl (2005) have analytical results only for small parametrized classes of direct mechanisms. They complement these results with a numerical investigation of second best mechanisms.

The compromise problem described here is a sandbox model for the more general model of determining an optimal voting scheme when there are three or more alternatives, that is, the Bayesian equivalent of the problem that Gibbard and Satterthwaite studied. The difficulties that we found analyzing the much simpler model indicate that the more general problem is analytically hard, and that it might benefit from numerical investigation.

7.5 Remarks on the Literature

The large literature on the Gibbard-Satterthwaite theorem is surveyed by Barbera (2001). Barbera describes several different approaches to the proof of the theorem. An important aspect of the Gibbard-Satterthwaite theorem that we have not mentioned in the text is that, as Satterthwaite (1975) pointed out, the theorem is equivalent to Arrow's famous theorem on the impossibility of preference aggregation; see Arrow (1963).

Our treatment of the case of single-peaked preferences is based on Barbera (2001) although the original result is in Moulin (1980). In fact, Moulin (1980) obtains a converse to Proposition 7.6 and shows that all dominant strategy incentive compatible direct mechanisms on a single-peaked domain need to be based on a generalized version of the median decision rule.

Our discussion of the case of Hylland's result is based on Dutta et al. (2007) who also offer an independent proof of the result. That paper also contains a detailed discussion of the relation between Hylland's result and Gibbard's earlier theorems on random dictatorship (Gibbard (1977), Gibbard (1978)).

Other papers that investigate Bayesian incentive compatibility without transferrable utility include Abdulkadiroglu and Loertscher (2007). The literature on the design of matching markets surveyed in Roth (2008) is also in a setting without transferrable utility. This literature is less focused on incentive compatibility than the mechanism design literature. Strategic properties of matching algorithms are, however, discussed in Section 2.1 of Roth (2008).

7.6 Problems

Chapter 8

Interdependent Types

8.1 Introduction

In all previous parts of these notes we have assumed that each agent's private information is all that matters for that agent's preferences over group choices. In this section, we shall instead consider the case in which an agent's preference over group choices depends not only on this agent's own private information, but also on the information of other agents. We shall refer to this as the case of "informational interdependence." By contrast, we refer to the case that we have addressed so far as the case of "private values."

One can think of many real world examples in which informational interdependence seems important. When choosing a new chair of a committee, different committee members will have different pieces of information about the abilities of the candidates. Each member's evaluation of a candidate will depend not only on the member's own information about the candidate, but also on other members' information about the candidate. Similarly, in auctions for licenses to use radio spectrum to operate a telephone service, each bidder will have some private information about the likely profitability of various services. The private information about market conditions that each bidder holds is likely to affect not only that bidder's valuation of licenses but also other bidders' valuations of licenses.

In this section we shall present some elements of a formal analysis of mechanism design with informational interdependence. Most of our discussion will focus on Bayesian incentive compatibility. We shall, however, also comment on ex post implementation, which is related to dominant strategy implementation.

We shall maintain the assumption of transferrable utility that we made in all previous parts of these notes except Chapter 7. This will make it easier to contrast the results of this section with those of previous sections.

For the case of informational independence, as for the case of private values, an important distinction is whether agents' signals are assumed to be independent random variables, or whether dependence is allowed. We shall focus here on the case that these signals are independent. For the case that signals are dependent, the permissive results obtained in Section 6.4 can be generalized. However, as in the context of that Section, these results are paradoxical, and it seems likely that some modification of the setting in which these results are obtained is needed in order to obtain more plausible results.

The dimensions of the agents' signal space will play an important role in this part of the lecture notes. The role that the dimensions play in the case of informational interdependence is, however, quite different from the role that they play in the case of private values. With informational interdependence the dimensions of the signal spaces are, for example, crucial for the question whether efficient decision rules can be part of a Bayesian incentive compatible mechanism. Recall that in the case of private values the answer to this question was positive, and independent of the dimensions of the signal spaces. Efficient decision rules could be implemented using Vickrey-Clarke-Groves (VCG) mechanisms.

The focus of this section will be on the case in which signal spaces are subsets of finite-dimensional Euclidean spaces where the dimension is larger than one. For this case, we explain a series of impossibility results that were obtained by Philippe Jehiel and Benny Moldovanu in a sequence of papers.

8.2 An Example

We start off with an extremely simple example. Suppose that there are 2 agents, $I = \{1, 2\}$, and two alternatives: $A = \{a, b\}$. Agent 1 observes a two-dimensional private signal $(\theta_1, \theta_2) \in [0, 1]^2$. We assume that (θ_1, θ_2) is a random variable with a distribution that has support $[0, 1]^2$. Agent 2 has no private information. Agent i's utility if alternative a is chosen, and agent i pays transfer t_i is: $\theta_i - t_i$. Agent i's utility if alternative b is chosen and agent i pays transfer t_i is: $0.5 - t_i$. Note the essential feature of this example: agent 1's private signal is relevant not only to his own utility but also to agent 2's utility.

By the revelation principle, we restrict attention to direct mechanisms. A decision rule $q : [0,1]^2 \rightarrow [0,1]$ maps agent 1's private signal into the probability with which alternative a is chosen for the given signal. A transfer rule $t : [0,1]^2 \rightarrow \mathbb{R}$ maps agent 1's private signal into a transfer to be paid by agent 1 for the given signal. Obviously, only transfers to be paid by agent 1 matter. A direct mechanism is "incentive compatible" if, for all realizations $(\theta_1, \theta_2) \in [0,1]^2$ of his private signal, agent 1 finds it in his interest to truthfully report the realization of the signal.

Let us ask a simple question: Can we find transfers that make welfare maximizing decisions incentive compatible? Here, we define welfare maximization in the usual way as maximizing the sum of the agents' utilities. More specifically, the "first best" decision rule q^* satisfies: $q^*(\theta_1, \theta_2) = 1$ if $\theta_1 + \theta_2 > 1$, and $q^*(\theta_1, \theta_2) = 0$ if $\theta_1 + \theta_2 < 1$. Our question is: Can we find a transfer rule t^* so that the direct mechanism (q^*, t^*) is incentive compatible?

As a first step, observe that we require that $t^*(\theta_1, \theta_2)$ is constant for all (θ_1, θ_2) such that $q^*(\theta_1, \theta_2) = 1$, because otherwise agent 1 would distort his report and only report that (θ_1, θ_2) for which t^* is minimal. Denote the constant payment by t_a. Similarly, we require that $t^*(\theta_1, \theta_2)$ is constant for all (θ_1, θ_2) such that $q^*(\theta_1, \theta_2) = 0$. Denote the constant payment by t_b.[20]

Now observe that for every $\theta_1 \in (0,1)$ agent 1 can report a θ_2 such that $q(\theta_1, \theta_2) = 0$, and he can also report a θ_2 such that $q(\theta_1, \theta_2) = 1$. Depending on which θ_2 he actually observes, we sometimes want him to choose the first, and sometimes the second option. Because his utility does not depend on θ_2, he must be indifferent between the two choices:

$$\begin{aligned} \theta_1 - t_a &= 0.5 - t_b \Leftrightarrow \\ t_a - t_b &= \theta_1 - 0.5. \end{aligned} \tag{8.1}$$

This has to be true for all $\theta_1 \in (0,1)$. Obviously, no two real numbers t_a and t_b have this property. We conclude that it is impossible to implement the first best.

The result that we observed for this example is not very surprising. Intuitively, we would like the group decision to be conditional on a component of an agent's private signal that has no implications for that agent's payoff. We have no instrument that would allow us to elicit this information from the agent.

[20] This is sometimes called the "taxation principle." Agent 1 is implicitly offered the choice between two alternatives, a and b, with associated taxes t_a and t_b.

Jehiel and Moldovanu (2001) have generalized the above example in two ways. Firstly, they consider a model in which each agent makes private observations that are relevant to other agents' utility but not to the agent's own utility. They show that it is, in general, impossible to implement a first-best decision rule. Their argument for this more general case is slightly different from the above argument. Secondly, they consider a model in which each agent makes observations that affect the agent's own utility as well as other agents' utilities, but potentially with different weights. Again they show the impossibility to implement first best decision rules. The proof is related to the argument in the above example. Our exposition below will focus on Jehiel/Moldovanu's second extension of the example.

8.3 Impossibility of Implementing Welfare Maximizing Decision Rules

We assume that there are N agents. The set of agents is denoted by $I = \{1, 2, \ldots, N\}$. They have to choose an alternative a out of some finite set $A = \{a_1, a_2, \ldots, a_K\}$ of K mutually exclusive alternatives. Each agent i observes a K-dimensional signal: $\theta^i = (\theta^i_1, \theta^i_2, \ldots, \theta^i_K) \in [0, 1]^K$.[21] The signal θ^i has a distribution with density f^i that is positive everywhere. Different agents' signals are independent.

Agent i's von Neumann Morgenstern utility if alternative a_k is chosen, and if agent i has to pay transfer t_i is:

$$\sum_{j=1}^{N} \alpha^j_{ki} \theta^j_k - t_i. \tag{8.2}$$

Thus, agent i's utility from alternative a_k is a linear function of the k-th component of all agents' signals. The factor in front of the k-th component of agent j's signal in agent i's utility function is: α^j_{ki}. We assume that $\alpha^j_{ki} \neq 0$ for all $i, j \in I$ and all $k = 1, 2, \ldots, K$. Note that we do not restrict these factors to have the same sign for all agents. The linear form of the utility function assumed above is not essential to the argument, but makes the exposition easier.

A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ consists firstly of a mapping $q : [0, 1]^{IK} \rightarrow \Delta(A)$ that assigns to each vector of reported private signals $\theta \equiv (\theta^1, \theta^2, \ldots,$

[21] Jehiel and Moldovanu (2001) allow more general convex signal spaces with non-empty interior, but for ease of exposition we restrict attention to the case that signals are contained in $[0, 1]^K$.

θ^N) a probability distribution $q(\theta)$ over A. The set of all such probability distributions is denoted by $\Delta(A)$. For every agent $i \in I$ a direct mechanism also specifies a mapping $t_i : [0,1]^{IK} \to \mathbb{R}$ that assigns to each vector θ of private signals a transfer $t_i(\theta)$ to be paid by agent i.

Given a direct mechanism we can define for every $i \in I$ and every $a \in A$ a function $Q_a^i : [0,1]^K \to [0,1]$ that assigns to every signal θ^i of agent i the interim probability that alternative a is chosen. We can also define for every $i \in I$ a function $T^i : [0,1]^K \to \mathbb{R}$ that assigns to every signal θ^i of agent i the interim expected value of agent i's transfer payment.

A direct mechanism is incentive compatible if and only if for every agent i, and every pair of types $\theta^i, \tilde{\theta}^i \in [0,1]^K$, agent i's expected utility from reporting type θ^i is at least as large as his expected utility from reporting type $\tilde{\theta}^i$ if agent i's true type is θ^i.

A choice rule q^* is "first best" if for every $\theta \in [0,1]^{IK}$ we have that $q^*(\theta)$ assigns positive probability only to alternatives that maximize the sum of agents' utilities. A direct mechanism is "first best" if its choice rule is first best.

A first question to ask is whether we can use a VCG mechanism to implement a first best decision rule. Recall that in previous sections we defined agent i's transfer rule in a VCG mechanism by:

$$t_i(\theta) = -\sum_{j \neq i} u_j(q(\theta), \theta_j) + \tau_i(\theta_{-i}) \text{ for all } \theta \in \Theta, \tag{8.3}$$

where τ_i is an arbitrary function. Under this rule, if agent i is of type θ_i and reports that he is of type θ_i', his utility is:

$$\begin{aligned} & u_i(q(\theta_i', \theta_{-i}), \theta_i) + \sum_{j \neq i} u_j(q(\theta_i', \theta_{-i}), \theta_j) - \tau_i(\theta_{-i}) \\ = & \sum_{j=1}^{N} u_j(q(\theta_i', \theta_{-i}), \theta_j) - \tau_i(\theta_{-i}). \end{aligned} \tag{8.4}$$

Thus, when agent i reports θ_i', his payoffs will be true social welfare, for actual types, if the collective decision is $q(\theta_i', \theta_{-i})$, minus some constant that does not affect agent i's incentives. Because the first best rule maximizes true social welfare, agent i will find it in his interest to report his type truthfully. Now, in our context, with informational interdependence, the definition of the transfer rule needs to be modified, because now any agent

j's utility does not only depend on agent j's type, but on all agents' types. Thus, we might define:

$$t_i(\theta) = -\sum_{j \neq i} u_j(q(\theta), \theta) + \tau_i(\theta_{-i}) \text{ for all } \theta \in \Theta. \tag{8.5}$$

Now, if agent i is of type θ_i and reports that he is of type θ_i', then his utility becomes:

$$u_i(q(\theta_i', \theta_{-i}), \theta) + \sum_{j \neq i} u_j(q(\theta_i', \theta_{-i}), (\theta_i', \theta_{-i})) - \tau_i(\theta_{-i}). \tag{8.6}$$

Note that agent i's utility is no longer aligned with true welfare. This is because the mechanism perceives a different utility for agents $j \neq i$ if agent i changes his report. Agent i is rewarded according to this perceived utility of agents j, rather than according to their true utility.

The previous paragraph showed that a generalization of the VCG mechanism cannot be used to implement a first best decision rule. The following result, which is the main result of Jehiel and Moldovanu (2001), shows that in fact no direct mechanism implements a first best decision rule.

Proposition 8.1. *Assume that any first best rule q^* is such that interim expected probabilities are differentiable, and, moreover, their derivative is never zero:*

$$\frac{\partial Q_a^i(\theta^i)}{\partial \theta_b^i} \neq 0$$

for every $i \in I$, $a, b \in A$ and $\theta^i \in [0,1]^K$. Assume also that there are some agent $i \in I$ and two alternatives $a, b \in A$ such that:

$$\frac{\alpha_{ai}^i}{\alpha_{bi}^i} \neq \frac{\sum_{j=1}^N \alpha_{bj}^i}{\sum_{j=1}^N \alpha_{aj}^i}.$$

Then no first best direct mechanism is Bayesian incentive compatible.

The first condition in this Proposition assumes that the interim expected probability of alternative a, conditioning on agent i's signal, changes as agent i's signal value for some arbitrary other alternative b changes. This is a regularity condition that is needed in the proof of Proposition 8.1 as presented below. It can be much weakened. For example, instead of requiring it to be true *for all* types of agent i, it is evident from the proof below that we might also require it to be true *for some* type of agent i.

The second condition says that there is at least some agent i, and a pair of alternatives a, b, such that the relative weight that agent i attaches to his signal for alternative a and his signal for alternative b is different from the relative weight that the social welfare function attaches to these two signals. If we pick the weights in agents' utility functions randomly from some continuous distribution, then this condition will be satisfied with probability 1. It is for this reason that Proposition 8.1 presents a generic impossibility result.

Note the contrast between Jehiel and Moldovanu's impossibility result and the impossibility results that we presented earlier in private value settings. Those earlier impossibility results only obtained when welfare maximization is combined with other requirements, such as individual rationality and budget balance. By contrast, the Jehiel Moldovanu impossibility results refer only to welfare maximization, with no other requirements.

Proof. Suppose there were a first best direct mechanism that is incentive compatible. We shall present a proof that assumes that for every agent i the interim expected utility function U_i is twice continuously differentiable. The proof in Jehiel and Moldovanu (2001) does not assume this.

The Envelope Theorem implies for every agent i and every $a \in A$:

$$\frac{\partial U^i}{\partial \theta^i_a} = \alpha^i_{ai} Q^i_a(\theta^i). \tag{8.7}$$

Now suppose we differentiate this expression again, this time with respect to θ^i_b for some $b \in A$. Then we obtain:

$$\frac{\partial^2 U^i}{\partial \theta^i_a \partial \theta^i_b} = \alpha^i_{ai} \frac{\partial Q^i_a(\theta^i)}{\partial \theta^i_b}. \tag{8.8}$$

If we change the order of differentiation, we obtain similarly:

$$\frac{\partial^2 U^i}{\partial \theta^i_b \partial \theta^i_a} = \alpha^i_{bi} \frac{\partial Q^i_b(\theta^i)}{\partial \theta^i_a}. \tag{8.9}$$

The fact that interim expected utility is twice continuously differentiable implies that the order of differentiation does not matter (Schwarz's Theorem).

Thus the two derivatives that we computed have to be the same.

$$\alpha^i_{ai} \frac{\partial Q^i_a(\theta^i)}{\partial \theta^i_b} = \alpha^i_{bi} \frac{\partial Q^i_b(\theta^i)}{\partial \theta^i_a} \Leftrightarrow$$

$$\frac{\alpha^i_{ai}}{\alpha^i_{bi}} = \frac{\frac{\partial Q^i_b(\theta^i)}{\partial \theta^i_a}}{\frac{\partial Q^i_a(\theta^i)}{\partial \theta^i_b}}. \tag{8.10}$$

Here, we could divide by the partial derivatives of the interim probabilities because, by the first assumption in Proposition 8.1, these derivatives are not zero.

Now suppose agents' utility functions were different. Suppose they were equal to social welfare. Then the given rule, without transfers, would also be Bayesian incentive compatible. An argument like the one that we just displayed would lead to the conclusion that the equation just derived is true with the weights from agents' original utility functions being replaced by the weights from the social welfare function:

$$\frac{\sum_{j=1}^N \alpha^i_{aj}}{\sum_{j=1}^N \alpha^i_{bj}} = \frac{\frac{\partial Q^i_b(\theta^i)}{\partial \theta^i_a}}{\frac{\partial Q^i_a(\theta^i)}{\partial \theta^i_b}}. \tag{8.11}$$

We can now deduce from our results so far that:

$$\frac{\alpha^i_{ai}}{\alpha^i_{bi}} = \frac{\sum_{j=1}^N \alpha^i_{aj}}{\sum_{j=1}^N \alpha^i_{bj}}. \tag{8.12}$$

This has to be true for every $i \in I$, and any two alternatives $a, b \in A$. But this contradicts the second condition in Proposition 8.1. □

8.4 Characterizing All Incentive Compatible Mechanisms

The impossibility result presented in the previous section is remarkable because it holds even if no individual rationality or budget balance. However, as in the case of the impossibility results that we obtained in the private value settings with individual rationality and budget balance, it is natural to ask next which mechanisms would be second best if we wanted to maximize ex ante expected social welfare. We might also investigate revenue maximizing mechanisms, or mechanisms that are chosen according to some other objective function.

For all these questions it is important to have a characterization of all incentive compatible direct mechanisms. With independent signals, which is the case that we are considering here, the characterizations that we developed for the case of private values generalize. In particular, if each agent's utility from the collective decision is a linear function of the agent's own type, then interim expected utilities are convex, and are determined by the collective decision rule up to a constant. Moreover, collective decision rules q can be part of a Bayesian incentive compatible mechanism if and only if at the interim level they generate for each agent the sub-gradient of a convex function. These characterizations are hard to use in practice.

Jehiel et al. (2006) study ex-post incentive compatibility instead of Bayesian incentive compatibility. Ex-post incentive compatibility means that for every realization of all other agents' types each agent finds it in his interest to report his type truthfully rather than distort it. If truth telling is an ex-post equilibrium of a direct mechanism, it is a Bayesian equilibrium for every belief that agents might hold about the other agents' types, including beliefs that are not product measures. With private values, truth telling is an ex post equilibrium if and only if truth telling is a dominant strategy. Thus, ex-post incentive compatibility is a generalization of dominant strategy incentive compatibility to the case of interdependent valuations. Jehiel et. al.'s main finding is that for generic utility functions only constant choice rules that choose the same alternative for every type realization are incentive compatible if each agent's signal space is at least two-dimensional.

A question one might ask about this result is why dictatorial choice rules are not incentive compatible. Suppose some agent i is allowed to pick the collective choice based on i's signal only, and suppose there are no transfers. Agent i will then pick the alternative that maximizes his expected utility, given his information. Jehiel et. al.'s result implies that such a rule is not ex-post incentive compatible. The reason is as follows: all agents other than i would be willing to reveal their private information, as it doesn't affect the collective decision. Once this information is revealed, agent i has an ex post incentive to deviate, and to change his decision.

Note that ex post incentive compatibility thus requires that we can uniquely predict an agent's choice independent of what this agent's beliefs about the other agent's types are. If we relaxed this requirement, and allowed that choices depend on beliefs, then more rules become feasible.

The logic of Jehiel et. al.'s proof of their result is as follows. Consider any two alternatives, a and b, and consider for two agents i and j the set of signals where a is chosen, and the set of signals where b is chosen. Suppose that these two sets have a boundary in common. Along this boundary, both agents, agent 1 and agent 2 have to be indifferent between a and b. Moreover, along this boundary, agent 1's payment is determined by agent 2's report, and agent 2's payment is determined by agent 1's report. Thus, fixing the other agent's signal, payments along the boundary are constant. But for generic utility functions it is impossible to find surfaces such that for fixed payments both agents are indifferent between two alternatives along these surfaces.

8.5 Remarks on the Literature

The example in Section 2 is a special case of an example that appears in Jehiel and Moldovanu (2001). This paper is also the main reference for Proposition 8.1. However, the short proof of Proposition 8.1 that I have given is taken from Jehiel and Moldovanu (2005). I have also used Jehiel et al. (2006).

8.6 Problems

Chapter 9

Robust Mechanism Design

To be written.

Chapter 10

Multiple Equilibria and Implementation

To be written.

Chapter 11

Dynamic Mechanism Design

To be written.

Chapter 12

Conclusion

To be written.

Bibliography

Atila Abdulkadiroglu and Simon Loertscher. Dynamic house allocations. 2007. 153

William J. Adams and Janet L. Yellen. Commodity bundling and the burden of monopoly. *Quarterly Journal of Economics*, 90:475–498, 1976. 30

Kenneth Arrow. *Social Choie and Individual Values*. Wiley, New York, second edition, 1963. 153

Salvador Barbera. An introduction to strategy-proof social choice functions. *Social Choice and Welfare*, 18:619–653, 2001. 153

Sushil Bikhchandani, Shurojit Chatterji, Ron Lavi, Ahuva Mu'alem, Noam Nisan, and Arunava Sen. Weak monotonicity characterizes deterministic dominant strategy implementation. *Econometrica*, 74:1109–1132, 2006. 113, 121

Patrick Bolton and Mathias Dewatripont. *Contract Theory*. MIT Press, Cambridge and London, 2005. 2, 30

Tilman Börgers and Peter Norman. A note on budget balance under interim participation constraints: The case of independent types. *Economic Theory*, 2008. 78

Tilman Börgers and Peter Postl. Efficient compromising. 2005. 151, 152

Kim Sau Chung and Jeffrey C. Ely. Foundations of dominant strategy mechanisms. *Review of Economic Studies*, 74:447–476, 2007. 81

Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, 8: 19–33, 1971. 121

Jacques Crémer and Richard McLean. Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 56:1247–1257, 1988. 129, 138

Claude d'Aspremont and Louis-Andre Gerard-Varet. Incentives and incomplete information. *Journal of Public Economics*, 11:25–45, 1979. 77

Joseph L. Doob. *Measure Theory*. Springer, 1994. 16

Bhaskar Dutta, Hans Peters, and Arunava Sen. Strategy-proof cardinal decision schemes. *Social Choice and Welfare*, 28:163–179, 2007. 153

Joel Franklin. *Methods of Mathematical Economics*. Springer, Heidelberg and New York, 1980. 132, 138

Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, Cambridge and London, 1993. 2, 4, 33

Alan Gibbard. Manipulation of voting schemes. *Econometrica*, 41:587–601, 1973. 143

Alan Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45:665–681, 1977. 153

Alan Gibbard. Straightforwardness of game forms with lotteries as outcomes. *Econometrica*, 46:595–614, 1978. 153

Jerry Green and Jean-Jacques Laffont. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica*, 45:727–738, 1977. 121

Theodore Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973. 121

Werner Güth and Martin Hellwig. The private supply of a public good. *Journal of Economics, Supplement 5*, pages 121–159, 1986. 78

Kathleen Hagerty and William Rogerson. Robust trading mechanisms. *Journal of Economic Theory*, 42:94–107, 1987. 97

Aviad Heifetz and Zvika Neeman. On the generic (im)possibility of full surplus extraction in mechanism design. *Econometrica*, 74:213–233, 2006. 137

Benjamin Hermalin. *Lecture Notes for Economics*. 2005. URL http://faculty.haas.berkeley.edu/hermalin/. 30

Bengt Holmström. Groves' scheme on restricted domains. *Econometrica*, 47:1137–1144, 1979. 121

Bengt Holmström and Roger Myerson. Efficient and durable decision rules with incomplete information. *Econometrica*, 51(6):1799–1819, November 1983.

Aanund Hylland. Strategy proofness of voting procedures with lotteries as outcomes and infinite sets of strateges. 1980. 151

Philippe Jehiel and Benny Moldovanu. Efficient design with interdependent valuations. *Econometrica*, 69:1237–1259, 2001. 157, 159, 160, 163

Philippe Jehiel and Benny Moldovanu. Allocative and informational externalities in auctions and related mechanisms. In *Proceedings of hte 9th World Congress*. Econometric Society, London, 2005. 163

Philippe Jehiel, Moritz Meyer ter Vehn, Benny Moldovanu, and Bill Zame. The limits of ex-post implementation. *Econometrica*, 74:585–611, 2006. 162, 163

Grigory Kosenok and Sergei Severinov. Individually rational, budget-balanced mechanisms and allocation of surplus. *Journal of Economic Theory*, 140:126–161, 2008. 136, 138

Vijay Krishna. *Auction Theory*. Academic Press, San Diego etc., 2002. 38, 42, 77

Vijay Krishna and Eliot Maenner. Convex potentials with an application to mechanism design. *Econometrica*, 69:1113–1119, 2001. 100, 121, 124

Jean-Jacques Laffont and David Martimort. Mechanism design with collusion and correlation. *Econometrica*, 68:309–342, 2000. 137

David G. Luenberger. *Optimization By Vector Space Methods*. John Wiley and Sons, Inc., New York etc., 1969. 56, 73

Alejandro M. Manelli and Daniel R. Vincent. Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *Journal of Economic Theory*, 137:153–185, 2007. 29, 30

Paul Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, Cambridge etc., 2004. 77, 121

Dilip Mookherjee and Stefan Reichelstein. Dominant strategy implementation of bayesian incentive compatible allocation rules. *Journal of Economic Theory*, 56:378–399, 1992. 97

Herve Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35:437–455, 1980. 153

Roger Myerson. Optimal auction design. *Mathematics of Operations Research*, 6:58–73, 1981. 77

Roger B. Myerson and Mark A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29:256–281, 1983. 65, 78

Zvika Neeman. The relevance of private information in mechanism design. *Journal of Economic Theory*, 117:55–77, 2004. 137

Efe A. Ok. *Real Analysis with Economic Applications*. Princeton University Press, Princeton and Oxford, 2007. 17

Philip Reny. Arrow's theorem and the gibbard-satterthwaite theorem: A unified approach. *Economics Letters*, 70:99–105, 2001. 143

Jacques Robert. Continuity in auction design. *Journal of Economic Theory*, 55:169–179, 1991. 137

Kevin Roberts. The characterization of implementable choice rules. In Jean-Jacques Laffont, editor, *Aggregation and Revelation of Preferences*, pages 321–348. North Holland, Amsterdam, 1979. 115, 121

Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16: 191–200, 1987. 104, 121

Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970. 108, 121

Alvin Roth. Deferred acceptance algorithms: History, theory, practie, and open questions. *International Journal of Game Theory*, 36:537–569., 2008. 153

Halsey Royden. *Real Analysis*. Prentice Hall, 3rd edition, 1988. 4, 12, 13, 16

Mark Satterthwaite. Strategy-profness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975. 143, 153

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961. 121

Index