AN INTRODUCTION TO THE THEORY OF **MECHANISM DESIGN**

TILMAN BÖRGERS

with a chapter by
DANIEL KRÄHMER and
ROLAND STRAUSZ

An Introduction to the Theory of Mechanism Design

An Introduction to the Theory of Mechanism Design

Tilman Börgers

With a chapter by

Daniel Krähmer and Roland Strausz

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

Library of Congress Cataloging-in-Publication Data
Börgers, Tilman.
An introduction to the theory of mechanism design / Tilman Börgers; with a chapter by
Daniel Krähmer and Roland Strausz.
p. cm.
Includes bibliographical references and index.
ISBN 978–0–19–973402–3 (alk. paper)
1. Game theory. I. Title.
HB144.B67 2015
519.3—dc23
2014046427

1 3 5 7 9 8 6 4 2

Printed in the United States of America
on acid-free paper

The authors will donate all payments that
they receive from the publisher for this book to
Amnesty International.

CONTENTS

PREFACE

The first objective of this text is to give rigorous but accessible explanations of classic results in the theory of mechanism design. The second objective is to take the reader in selected areas to the frontiers of research. The description of recent areas of research is, by necessity, a little more subjective than the description of classic results. The reader may turn to recent literature in the field to learn about perspectives that are different from the authors'.

This book is meant for advanced undergraduate and graduate students of economics who have a good understanding of game theory. Fudenberg and Tirole (1993) contains more than the reader needs for this book. I shall also assume a basic knowledge of real analysis that can, for example, be acquired from Rudin (1976).

This book started out as lecture notes for a class that I taught at the University of Michigan in the Winter Semester 2006. At the time, as far as I was aware, no books on mechanism design existed. As I was working excruciatingly slowly on my text, several excellent books appeared that cover topics similar to mine. Whenever I discovered such a book, to maintain my own motivation, I needed to persuade myself that there are important differences between other authors' books and my own. These excellent books, along with the justifications for the existence of my own book, are listed below in chronological order. I include them here, even though they are in some sense competitors of my own text, to serve the primary purpose of a book such as mine, which is to help others to obtain a comprehensive picture of the field. The following books, hopefully together with my own book, will be extremely useful as the reader embarks on this endeavor.

- Leonid Hurwicz and Stanley Reiter. *Designing Economic Mechanisms*. Cambridge: Cambridge University Press, 2006.

  The theory of mechanism design was created by Leonid Hurwicz, who won the 2007 Nobel Prize for it, together with Eric Maskin and Roger Myerson. The focus of this text is on informational efficiency and privacy preservation in mechanisms. Incentive aspects play a much smaller role than they do in this book.

- Steven R. Williams. *Communication in Mechanism Design: A Differential Approach*. Cambridge: Cambridge University Press, 2008.

  This book covers material similar to that of Hurwicz and Reiter. The emphasis that both books place on the size of the message space in a mechanism differentiates them from more modern treatments of mechanism design. However, as we shall discuss in this book, in particular in Chapter 10, the complexity or simplicity of mechanisms, one aspect of which is the size of the message space, seems to be of continuing importance and may be central for future research.

- Dmitrios Diamantaras, with Emina I. Cardamone, Karen A. Campbell, Scott Deacle, and Lisa A. Delgado. *A Toolbox for Economic Design*. New York: Palgrave MacMillan, 2009.

  This book is closest to mine among those listed here, but it covers more than I do, such as the theory of Nash implementation, the theory of matching markets, and empirical evidence on mechanisms. Sometimes I wish I had written this book. My own book is more narrowly focused, perhaps goes somewhat into greater depth, and places a greater emphasis on the relation between game theoretic foundations and mechanism design.

- Rakesh Vohra. *Mechanism Design, A Linear Programming Approach*. Cambridge: Cambridge University Press, 2011.

  This is a superb book, demonstrating how large parts of the theory of mechanism design can be developed as an application of results from linear programming. Vohra puts less emphasis than I do on the game theoretic aspects of mechanism design.

My comments on the books listed above already indicate some limitations of this text. Two more require emphasis. First, I have not covered the theory of implementation. I explain in the remainder of this paragraph what specifically I mean by this, but I note that this paragraph may be comprehensible to readers only after they have read this book. I understand the literature on implementation to be that part of the literature on mechanisms that requires the mechanism designer to consider *all* equilibria of the mechanism that she proposes, as opposed the literature on mechanism design, which allows the mechanism designer to select one among the equilibria of the mechanism that she proposes. This terminology is not universally used, but it will be useful in this book. Note that whenever we invoke the revelation principle, we are in the realm of the theory of mechanism design rather than the theory of implementation. I mentioned already that the book by Diamantaras et al. cited above includes chapters on implementation. Extremely useful is also:

- Matthew O. Jackson. A crash course in implementation theory. *Social Choice and Welfare*, **18** 655–708, 2001.

I have left out the subject of implementation, not because it would not be important, but because it has been explained so well by others and because it would require many techniques and arguments very different from the ones invoked in this book. Having said this, I continue to feel guilty for leaving out this subject.[1]

I have also not covered the contributions to the theory of mechanism design made by computer scientists. A survey of these contributions can be found in Chapters 9–16 of:

- Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge: Cambridge University Press, 2007.

I have omitted this work from my book because I am not sufficiently familiar with it. For example, I am currently reading at my usual snail's pace through the survey chapters mentioned above. But while I am doing this, the number of papers on mechanism design written by computer scientists seems to grow at breathtaking speed. Finding out how to stay abreast of these developments and how to integrate my knowledge of economists' research with whatever I can learn about computer scientists' research is a task so large that I cannot realistically tackle it in this book.

A very large subject that readers might find fascinating is the subject of mechanism design in practice. This subject requires a separate book, or multiple books. Examples of surveys which the reader might find useful are:

- Alvin E. Roth. What have we learned from market design? *Economic Journal*, **118** 285–310, 2008.
- Peter Cramton, Spectrum auction design. *Review of Industrial Organization*, **42** 161–190, 2013.

In contrast to these works, the emphasis of this book is on the methodology of the theory of mechanism design. I may now already have lost readers with a more practical bent. But, to keep those readers who have stuck with me up to now, it is time to begin.[2]

Ann Arbor, August 2014 Tilman Börgers

ACKNOWLEDGMENTS

This text explains the parts of the theory of mechanism design that I believe I have understood. Whatever I have understood, I have mostly learned from others' books and articles, from students who have politely listened to me and told me where I was going wrong, and from my co-authors during our joint research.

One area in which my lack of knowledge was particularly comprehensive is the theory of dynamic mechanism design. Daniel Krähmer and Roland Strausz generously agreed to take over the task of writing a chapter on this subject. Their work is included here as Chapter 11.

Christoph Kuzmics allowed me to see some problems that he gave to his students when teaching a class based on this book. I have included some problems here that were inspired by Christoph's problems.

I would also like to thank the anonymous referees of Oxford University Press as well as Stefan Behringer, George Chen, Shaowei Ke, Christoph Kuzmics, Xian Li, Vitor Farinha Luz, Stephen Morris, Colin von Negenborn, Martin Pollrich, Arunava Sen, Xianwen Shi, Jan-Henrik Steg, Roland Strausz, and Tobias Widmer for very helpful corrections, comments, and suggestions.

Trevor Burnham, Yan-Min Choo, and Nikoleta Scekic have proofread various parts of this book. They have caught many embarrassing errors. I am immensely grateful to them.

I have benefited a lot from Vijay Krishna's (2002) and Paul Milgrom's (2004) books on auction theory. The exposition in Chapters 2 and 3 owes a lot to these books. The idea to present the screening problem in Chapter 2 before turning to mechanism design proper has its origin in a conversation with Mark Armstrong.

I am grateful to Daniel and Roland for agreeing to donate all authors' income from this book to Amnesty International. Even if you feel after reading the book that you have wasted your time, rest assured that you have not completely wasted your money. Some of it benefits a good cause.

Ann Arbor, August 2014 Tilman Börgers

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 the refrigerator is worth to different people. How can you find out whether the sum of the amounts that everyone would at most be 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 noncooperative 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. In other words, the participants in these procedures will understand that they are playing a noncooperative game. 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 it 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. First, 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. One could call this first aspect of the theory of mechanism design the "normative" side of mechanism design. Second, 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, as well as 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. One could call this second aspect of the theory of mechanism design the "positive" side of mechanism design.

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, however, 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, which 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 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 Introduction, 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, like the theory of mechanism design, concerned with the optimal choice of rules for a game. 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 this book, 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 this book 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 this book 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.

I have chosen a somewhat unusual beginning for the book 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.

Chapter 3 then reviews the classic "Bayesian" theory of mechanism design in the context of some prominent examples. Chapter 5 develops a generalization of the single agent theory of Chapter 2, covering a very large class of models. Chapter 6 then develops in an analogous way the "Bayesian" theory of mechanism design for a very general framework.

The game theoretic models that are used in Chapters 3 and 6 are 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 Chapter 4, in the context of examples, and in Chapter 7, in general.

In Chapters 8 and 9 we relax other assumptions of the classic model. In Chapter 10 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.

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 therefore 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

Consider the following situation: 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 he purchases the good and pays a monetary transfer t to the seller is $\theta - t$. The buyer's utility[1] if he does not purchase the good is zero. Here, $\theta \geq 0$ 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 (a) paying θ and obtaining the good and (b) not obtaining the good.

Two aspects of the assumptions about the buyer's utility deserve emphasis. First, 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 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 (a) the buyer's utility from consuming the good is independent of the amount of money that he pays for it and (b) 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 this book. Much of the classical theory of mechanism design is based on these assumptions. We shall consider a more general case in Chapter 8.

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 8 will relax this assumption, too. Utility functions that satisfy additive separability of utility and linearity of utility in money are referred to as "quasi-linear" utility functions.

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 p 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 buyer. The seller could offer the buyer 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 that it is to her advantage, or at least not disadvantageous, to commit to a strategy and to announce this strategy to the buyer before play begins. An example would be that the seller announces that she will bargain over the price with the buyer, and that she announces also in advance that she 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 that game is 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. The buyer faces a single person decision problem. In subsequent chapters of this book, when we consider mechanism design proper, there will be multiple agents in roles similar to the buyer's role in this chapter, and their strategic interaction will be a game, even after the seller has committed to a strategy.

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, as well as by committing to selling the good at that price whenever the buyer is willing to pay the price?

A rather silly answer to our question could be this: The seller should choose the game according to which the buyer has only a single choice, and this single choice does not give him the object, but obliges her to pay x dollars, where 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 as well as the constraint that the buyer, within the opportunities that the game offers, will choose an expected utility maximizing strategy.

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, without loss of generality, to a small subset of mechanisms, called "direct mechanisms."

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

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

and

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

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 his type is θ, and the buyer has to pay the seller $t(\theta)$ if he reports that his 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 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.

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*

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

that is, 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 his 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 his 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 he would have obtained had he played the optimal strategy of type θ'—that is, $\sigma(\theta')$—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 his type.

Given a direct mechanism, we define the buyer's expected utility $u(\theta)$ conditional on his 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 his type.

Definition 2.2 *A direct mechanism is "incentive-compatible" if truth telling is optimal for every type $\theta \in [\underline{\theta}, \bar{\theta}]$, that is, if*

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

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 his type, is willing to participate, that is, if*

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

Note that in this definition we require voluntary participation *after the buyer has learned his type*. A weaker requirement would be to require voluntary participation *before the buyer has learned his type*. Economic theorists say that we have required voluntary participation at the "interim" level rather than at the "ex ante" level. This terminology will make more sense when, in the next chapter, we consider the case of multiple potential buyers rather than just one buyer.

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

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

and

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

Subtracting these two inequalities, we obtain

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

□

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." Thus, this result is just a statement of the envelope theorem in our context. For completeness, we provide a self-contained proof.

Proof

Incentive compatibility means that for all θ we have

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

Given any value of θ', $\theta q(\theta') - t(\theta')$ is an increasing and affine (and hence convex) function of θ. The maximum of increasing functions is increasing, and the maximum of convex functions is convex. Therefore, u is increasing and convex. Convex functions are not differentiable in at most countably many points [Theorem 18 in Chapter 6 of Royden and Fitzpatrick (2010)]. Consider any θ for which u is differentiable. Let $\delta > 0$. Then, by incentive compatibility we have

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

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

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.

$\square$

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 $\theta \in [\underline{\theta}, \bar{\theta}]$ we have*

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

Proof

The fact that u is convex implies, by Corollary 17 in Chapter 6 of Royden and Fitzpatrick (2010), that it is absolutely continuous. By Theorem 10 in Chapter 6 of Royden and Fitzpatrick (2010), 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 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 obtained 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 makes to the seller. This is shown in the next lemma.

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

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

Proof

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

Lemma 2.4 shows that the expected payments of the different buyer types are pinned down by 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 $\theta \in [\underline{\theta}, \bar{\theta}]$ we have

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

Proof

To prove sufficiency, we have to show that no type θ prefers to pretend to be a type θ' if (i) and (ii) in the proposition hold. What we have to show is

$$u(\theta) \geq \theta q(\theta') - t(\theta') \Leftrightarrow \tag{2.10}$$

$$u(\theta) \geq \theta q(\theta') - \theta' q(\theta') + \theta' q(\theta') - t(\theta') \Leftrightarrow \tag{2.11}$$

$$u(\theta) \geq \theta q(\theta') - \theta' q(\theta') + u(\theta') \Leftrightarrow \tag{2.12}$$

$$u(\theta) - u(\theta') \geq (\theta - \theta') q(\theta') \Leftrightarrow \tag{2.13}$$

$$\int_{\theta'}^{\theta} q(x) dx \geq \int_{\theta'}^{\theta} q(\theta')\, dx. \tag{2.14}$$

To see that the left-hand sides of the last two inequalities are the same, we note that the formulas in Lemmas 2.3 and 2.4 are actually equivalent, as the calculation which we describe in the proof of Lemma 2.4 shows. Therefore, condition (ii) in Proposition 2.2 implies that we can use in our proof the formula in Lemma 2.3. But this formula says exactly that the left-hand sides of the last two inequalities are the same.

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 nonnegative for all θ if and only if it is nonnegative 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 $t(\underline{\theta}) \leq \underline{\theta} q(\underline{\theta})$. 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

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

that is, 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:

$$g = \alpha f \Leftrightarrow \left[g(x) = \alpha f(x) \quad \forall x \in [\underline{\theta}, \bar{\theta}] \right] \qquad \text{for all } \alpha \in \mathbb{R}, f, g \in \mathcal{F}, \tag{2.16}$$

$$h = f + g \Leftrightarrow \left[h(x) = f(x) + g(x) \quad \forall x \in [\underline{\theta}, \bar{\theta}] \right] \qquad \text{for all } f, g, h \in \mathcal{F}. \tag{2.17}$$

This makes $\mathcal{F}$ a vector space. We also give $\mathcal{F}$ the L^1-norm—that is, the norm of any function $f \in \mathcal{F}$ is $\int_{\underline{\theta}}^{\bar{\theta}} \mid f \mid d\mu$ where μ is the Lebesgue measure—and we endow $\mathcal{F}$ with the metric induced by this norm.

We denote by $\mathcal{M} \subset \mathcal{F}$ the set of all *increasing* functions in $\mathcal{F}$ such that $f(x) \in [0,1]$ $\forall x \in [\underline{\theta}, \bar{\theta}]$. 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 selection theorem (Rudin, 1976, Exercise 13, p. 167) and the bounded convergence theorem of Lebesgue integration (Rudin, 1976, Corollary, p. 322). We omit the details.

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 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 $f : X \rightarrow \mathbb{R}$ be a continuous linear function. Then the set E of extreme points of X is nonempty, and there exists an $e \in E$ such that*

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

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*[4] $\theta \in [\underline{\theta}, \bar{\theta}]$.

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.[5] 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; that is, there is some θ^* such that $q(\theta^*) \in (0, 1)$.[6] 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 $\mathcal{M}$. □

The seller can thus restrict her attention to nonstochastic mechanisms. But a nonstochastic 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 *The following direct mechanism maximizes the seller's expected revenues among all incentive-compatible, individually rational direct mechanisms. Suppose $p^* \in \text{argmax}_{p \in [\underline{\theta}, \bar{\theta}]} p(1 - F(p))$. Then:*

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

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 lengths 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, you might 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 nonobvious.

2.3 NONLINEAR PRICING

Next, 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 also because its analysis introduces additional arguments, 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 profit. 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 buyer is risk-neutral in money.

The parameter θ reflects how much the buyer values the good. More precisely, the larger the value of θ, the larger the buyer's absolute willingness to pay $\theta \nu(q)$ and the buyer's marginal willingness to pay $\theta \nu'(q)$ for any given quantity q. The parameter θ can take any value between $\underline{\theta}$ and $\bar{\theta}$.

Now we turn to information. The buyer knows θ and $\nu(q)$. The seller does not know θ but does know $\nu(q)$. Obviously, knowing the precise form of $\nu(q)$ is knowing a lot. By assuming that θ is the only parameter of the buyer's utility function that the seller does not know, we have reduced the potentially high-dimensional uncertainty of the seller to a single-dimensional uncertainty. That makes our maximization problem below substantially easier than it would otherwise be. The seller's beliefs about θ are described 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 $\bar{\theta}v'(0) > c$. 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 $\lim_{q\to\infty} \bar{\theta}v'(q) < c$. 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*

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

and

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

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 if the buyer's report is θ. 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 nonnegative number, not a probability distribution over nonnegative numbers. We make this assumption for simplicity.[7] We do not state the revelation principle formally for our context. It is analogous to the revelation principle in the previous section.

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 $\theta \in [\underline{\theta}, \bar{\theta}]$:

$$t(\theta) = t(\underline{\theta}) + (\theta \nu(q(\theta)) - \underline{\theta}\nu(q(\underline{\theta}))) - \int_{\underline{\theta}}^{\theta} \nu(q(x))\, dx. \tag{2.18}$$

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

$$t(\underline{\theta}) \leq \underline{\theta}\nu(q(\underline{\theta})). \tag{2.19}$$

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,

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

Substituting this into equation (2.18) yields

$$t(\theta) = \theta\nu(q(\theta)) - \int_{\underline{\theta}}^{\theta} \nu(q(x))\, dx. \tag{2.21}$$

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 (2.21), 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 (2.21) to study in more detail the seller's expected profit. If the seller chooses $q(\cdot)$, then her expected profit is

$$\begin{aligned} &\int_{\underline{\theta}}^{\bar{\theta}} \left[\theta\nu(q(\theta)) - \int_{\underline{\theta}}^{\theta} \nu(q(x))\, dx - cq(\theta)\right] f(\theta)\, d\theta \\ &\quad = \int_{\underline{\theta}}^{\bar{\theta}} \theta\nu(q(\theta)) f(\theta)\, d\theta - \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} \nu(q(x))\, dx f(\theta)\, d\theta - \int_{\underline{\theta}}^{\bar{\theta}} cq(\theta) f(\theta)\, d\theta. \end{aligned} \tag{2.22}$$

We seek to simplify the expression in (2.22). 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 (2.22).

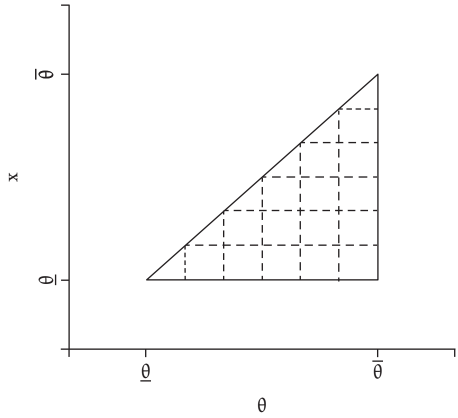

Figure 2.1 Changing the Order of Integration.

$$
\begin{aligned}
&\int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} v(q(x))dx f(\theta)d\theta \\
&\quad = \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\theta} v(q(x))f(\theta)dx d\theta \\
&\quad = \int_{\underline{\theta}}^{\bar{\theta}} \int_{x}^{\bar{\theta}} v(q(x))f(\theta)d\theta dx \\
&\quad = \int_{\underline{\theta}}^{\bar{\theta}} v(q(x)) \int_{x}^{\bar{\theta}} f(\theta)d\theta dx \\
&\quad = \int_{\underline{\theta}}^{\bar{\theta}} v(q(x))(1 - F(x))dx \\
&\quad = \int_{\underline{\theta}}^{\bar{\theta}} v(q(\theta))(1 - F(\theta))d\theta. \qquad (2.23)
\end{aligned}
$$

When moving from the second to the third line in (2.23), 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 (2.23) into the seller's objective function in (2.22) and rearrange:

$$
\begin{aligned}
&\int_{\underline{\theta}}^{\bar{\theta}} \left[\theta v(q(\theta)) - cq(\theta)\right] f(\theta)\, d\theta - \int_{\underline{\theta}}^{\bar{\theta}} v(q(\theta))(1 - F(\theta))\, d\theta \\
&\quad = \int_{\underline{\theta}}^{\bar{\theta}} \left[\theta v(q(\theta)) - cq(\theta)\right] f(\theta)\, d\theta - \int_{\underline{\theta}}^{\bar{\theta}} v(q(\theta)) \frac{1 - F(\theta)}{f(\theta)} f(\theta)\, d\theta \\
&\quad = \int_{\underline{\theta}}^{\bar{\theta}} \left[v(q(\theta)) \left(\theta - \frac{1 - F(\theta)}{f(\theta)} \right) - cq(\theta) \right] f(\theta)\, d\theta. \qquad (2.24)
\end{aligned}
$$

The seller thus chooses q to maximize the expected value of the expression that is in large square brackets in (2.24). Expected values are taken over θ. The seller must choose an increasing function q.

The expression that is in the square brackets in equation (2.24) is also called the "virtual surplus" generated by the seller's transaction with the buyer of type θ. The actual surplus that trade with the buyer of type θ generates is $\nu(q(\theta))\theta - cq(\theta)$, that is, the buyer's utility minus the production cost. If the seller can observe θ, this actual surplus is what the seller and the buyer can share. A seller who has commitment power, as we have assumed here, would maximize the actual surplus and then recover this surplus by including it in the price charged to the buyer, on top of the production cost. But in our setting, where θ is not observable, the surplus generated by the transaction with buyer type θ that the seller can recover from the buyer is lower than it would be if θ were observable. This can be seen from the fact that in the square brackets we subtract a positive term from θ. The subtracted term represents the reduction in surplus recovered by the seller that is due to the constraint that she must offer incentives to reveal θ truthfully. In the literature this subtracted term is called the buyer's "information rent."

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

$$\nu'(q)\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) - c = 0 \Leftrightarrow \tag{2.25}$$

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

We now investigate the existence of a solution to (2.26). If

$$\theta - \frac{1-F(\theta)}{f(\theta)} \leq 0, \tag{2.27}$$

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

$$q(\theta) = 0. \tag{2.28}$$

Now consider

$$\theta - \frac{1-F(\theta)}{f(\theta)} > 0. \tag{2.29}$$

Recall that we have assumed that v' is differentiable, and therefore continuous and decreasing, and that $\bar{\theta}v'(q)$ tends to less than c as q tends to infinity. Obviously, the left-hand side of (2.26) shares all these properties. If

$$v'(0)\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) \leq c, \tag{2.30}$$

then it is again obvious that the optimal choice is

$$q(\theta) = 0. \tag{2.31}$$

If

$$v'(0)\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) > c, \tag{2.32}$$

then our assumptions imply that there is a unique solution to (2.26) and that this solution 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 (2.24). 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 (2.26) 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 the following:*

(i) If $v'(0)\left(\theta - \frac{1-F(\theta)}{f(\theta)}\right) \leq c$, *then*

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

(ii) Otherwise,

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

The corresponding expected profit maximizing t is given by

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

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

$$v'(q(\bar{\theta}))\bar{\theta} = c. \tag{2.33}$$

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 maximizes the surplus generated by the transaction between the buyer and the seller. We refer to this quantity as the "first best" quantity. The fact that the first best quantity is offered to the top buyer type $\bar{\theta}$ is sometimes expressed by economic theorists as follows: "There is no distortion at the top."

For lower types $\theta < \bar{\theta}$ for whom the quantity supplied is determined by equation (2.26), a quantity that is lower than the surplus maximizing quantity is chosen. This is because equation (2.26) differs from the first best condition (2.33) (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. We might call this the "virtual marginal benefits," in line with the earlier term "virtual surplus." This means that all types $\theta < \bar{\theta}$ for whom the quantity supplied is determined by equation (2.26) are offered a quantity that is smaller than the "first-best" quantity. Finally, types which are offered a quantity of zero in the expected profit maximizing direct mechanism are obviously offered weakly less than they are offered in the first best. We thus find a phenomenon familiar from introductory economics classes: The profit maximizing monopolist distorts the quantity supplied (weakly) downwards.

We conclude with a numerical example.

Example 2.1 *$c = 1$, $v(q) = \sqrt{q}$, θ is uniformly distributed on $[0,1]$; that is, $F(\theta) = \theta$ 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:*

$$\theta - \frac{1 - F(\theta)}{f(\theta)} = \theta - \frac{1-\theta}{1} = 2\theta - 1, \tag{2.34}$$

which is obviously the case.

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

$$\begin{aligned} v'(0)\left(\theta - \frac{1 - F(\theta)}{f(\theta)}\right) &\leq c \Leftrightarrow \\ \theta - \frac{1 - F(\theta)}{f(\theta)} &\leq 0 \Leftrightarrow \\ 2\theta - 1 &\leq 0 \Leftrightarrow \\ \theta &\leq 0.5. \end{aligned} \tag{2.35}$$

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

$$\begin{aligned} v'(q(\theta))\left(\theta - \frac{1 - F(\theta)}{f(\theta)}\right) &= c \Leftrightarrow \\ \frac{1}{2\sqrt{q}}(2\theta - 1) &= 1 \Leftrightarrow \\ \sqrt{q} &= \theta - \frac{1}{2} \Leftrightarrow \\ q &= \left(\theta - \frac{1}{2}\right)^2. \end{aligned} \tag{2.36}$$

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

$$\begin{aligned} t(\theta) &= \theta v(q(\theta)) - \int_{\underline{\theta}}^{\theta} v(q(x))\,dx \\ &= \theta\left(\theta - \frac{1}{2}\right) - \int_{0.5}^{\theta} x - \frac{1}{2}\,dx \\ &= \theta\left(\theta - \frac{1}{2}\right) - \left[\frac{1}{2}x^2 - \frac{1}{2}x\right]_{0.5}^{\theta} \\ &= \theta\left(\theta - \frac{1}{2}\right) - \left(\frac{1}{2}\theta^2 - \frac{1}{2}\theta - \frac{1}{8} + \frac{1}{4}\right) \\ &= \frac{1}{2}\theta^2 - \frac{1}{8}. \end{aligned} \tag{2.37}$$

We want to translate the solution into an optimal nonlinear 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:

$$\begin{aligned} q(\theta) &= q \Leftrightarrow \\ \left(\theta - \frac{1}{2}\right)^2 &= q \Leftrightarrow \\ \theta - \frac{1}{2} &= \sqrt{q} \Leftrightarrow \\ \theta &= \sqrt{q} + \frac{1}{2} \end{aligned} \tag{2.38}$$

The payment by type θ is

$$\begin{aligned} t(\theta) &= \frac{1}{2}\theta^2 - \frac{1}{8} \\ &= \frac{1}{2}\left(\sqrt{q} + \frac{1}{2}\right)^2 - \frac{1}{8} \\ &= \frac{1}{2}q + \frac{1}{2}\sqrt{q} \end{aligned} \tag{2.39}$$

Essentially, the monopolist thus offers to the buyers the deal that they can buy any quantity $q \in [0, \frac{1}{4}]$ and then have to pay $t(q) = \frac{1}{2}q + \frac{1}{2}\sqrt{q}$. This optimal nonlinear pricing scheme is shown in Figure 2.2. We note that there is a quantity discount. The per unit price,

$$\frac{t}{q} = \frac{\frac{1}{2}q + \frac{1}{2}\sqrt{q}}{q} = \frac{1}{2} + \frac{1}{2\sqrt{q}}, \tag{2.40}$$

decreases in q.

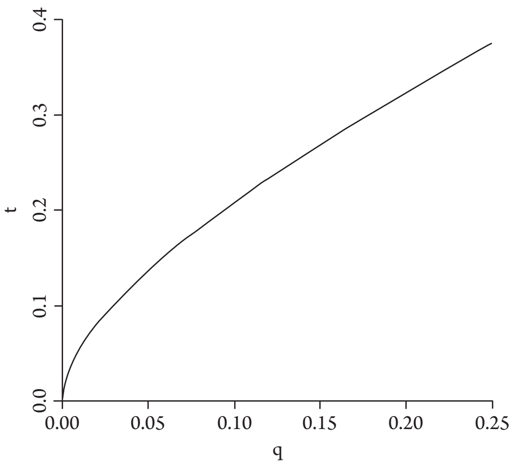

Figure 2.2 The Optimal Nonlinear Pricing Scheme.

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 this book.

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

$$I_A v_A + I_B v_B - t. \tag{2.41}$$

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 without loss of generality to direct mechanisms. We shall instead simplify in a different way and only consider a very small class of indirect mechanisms. We should emphasize, though, that we offer no reason that we could restrict attention to this small class without loss of generality. We restrict attention to this small class only to make our analysis easily tractable.

We assume that the seller quotes 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

$$
\begin{aligned}
p_A = p_B &= \frac{2}{3} \\
p_{AB} &= \frac{1}{3}\left(4 - \sqrt{2}\right) \approx 0.862.
\end{aligned} \tag{2.42}
$$

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 buyers, but she 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 she quotes the three optimal prices. We present this example mainly to illustrate that even in the

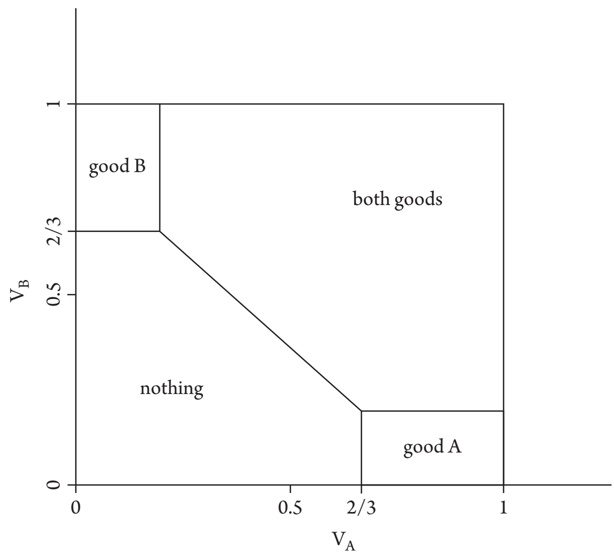

Figure 2.3 Buyer Behavior Given Optimal Prices.

simple case of screening, multidimensional 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 buyer'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. Here, we cannot go into this. As this example indicates, the general theory of screening with multiple goods and multidimensional private information is rather complicated.

2.5 REMARKS ON THE LITERATURE

Our exposition in Section 2.2 is a modified version of Manelli and Vincent (2007). The theory of nonlinear pricing in Section 2.3 is due to Mussa and Rosen (1978). It is discussed further in Bolton and Dewatripont (2005). Throughout this book we shall, whenever it is helpful, make a regularity assumption such as Assumption 2.1. If this assumption does not hold, one needs to use a technique that is commonly referred to in the literature as the "ironing technique." It is explained in Myerson (1981). 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) In the setting of Section 2.2 consider the following two-stage mechanism: In stage 1 the seller posts a price 0.5. Then the buyer can choose to buy or not to buy. If the buyer buys, the game is over. If he does not buy, then a third party draws a price randomly from the interval $[\underline{\theta}, \bar{\theta}]$, using the uniform distribution. The buyer can then either buy or not buy at the random price. Find the buyer's optimal strategy for this mechanism. Then find an equivalent direct mechanism in which truth telling is an optimal strategy for the buyer.

(b) 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.

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

(d) Consider the setting of Section 2.3. Suppose that the seller's beliefs about θ are described by the exponential distribution with density $f(\theta) = \lambda e^{-\lambda\theta}$ and cumulative distribution function $F(\theta) = 1 - e^{-\lambda\theta}$ for all $\theta \geq 0$. Assume that $v(q) = q^{\alpha}$

with $\alpha \in (0, 1)$. Find the schedule (q, t) that maximizes the seller's expected revenue. You may want to assume for some calculations that $\alpha = 0.5$. (Note that in this example the support of F is not bounded from above. Does that cause a problem for the analysis of Section 2.3?)

(e) Prove that the conditions in (2.18) are sufficient for incentive compatibility.

(f) Prove that the prices in (2.42) maximize the seller's profits.

3

BAYESIAN MECHANISM DESIGN: EXAMPLES

3.1 INTRODUCTION

This chapter describes three classic mechanism design problems. We shall use the notion of 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 because we first want to illustrate the general analysis before we present that analysis in later chapters. The examples are also interesting in their own right. We juxtapose optimal Bayesian mechanisms and 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 Setup

The model in this section is the same as in Section 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 $I = \{1, 2, \ldots, N\}$. Buyer i's utility if he purchases the good and pays a transfer t_i to the seller is $\theta_i - t_i$. Buyer i's utility if he does not purchase the good and pays a transfer of t_i to the seller is $0 - t_i$. The seller's utility if she 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 the random variables $\theta_1, \theta_2, \ldots, \theta_N$ are mutually 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 word "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 Section 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 seller 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 personal 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 her good. For example, she could pick a price, ask each buyer to indicate whether he 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 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 no, 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." We can write $q(\theta)$ as $(q_1(\theta), q_2(\theta), \ldots, q_N(\theta))$. 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. We shall also refer to it as the "payment rule" for buyer i. Note that we have assumed that this transfer payment is deterministic. This is without loss of generality as all agents are risk-neutral.

We now state the "Revelation Principle," which as in Section 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 truthtelling, 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 $\Theta_{-i} \equiv [\underline{\theta}, \bar{\theta}]^{N-1}$. We denote by F_{-i} the cumulative distribution of θ_{-i}, and we denote by f_{-i} the density of F_{-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

$$Q_i(\theta_i) = \int_{\Theta_{-i}} q_i(\theta_i, \theta_{-i}) f_{-i}(\theta_{-i}) \, d\theta_{-i}. \tag{3.1}$$

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

$$T_i(\theta_i) = \int_{\Theta_{-i}} t_i(\theta_i, \theta_{-i}) f_{-i}(\theta_{-i})\, d\theta_{-i}. \tag{3.2}$$

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.[1] 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*

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

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

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

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 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 frequently in this book. For example, we shall refer to $T_i(\theta_i)$ as the interim expected transfer of agent *i* if he is of type θ_i, and we shall refer to $U_i(\theta_i)$ as the interim expected utility of agent *i* if he is of type θ_i. The individual rationality condition in Definition 3.3 is also called the condition of "interim individual rationality."

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 Section 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 $\theta_i \in [\underline{\theta}, \bar{\theta}]$ we have*

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

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

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

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, as well as 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 the 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 Section 2.2, we can collect the observations made so far and can 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: first the allocation rule q and second 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, she can pick the interim expected payments by the lowest types in any arbitrary way and also 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 Proposition 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 rational direct mechanism maximizes the seller's expected revenue, then for every $i \in I$ we have*

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

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 $\theta_i \in [\underline{\theta}, \bar{\theta}]$ the following:

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

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

We shall now focus on the optimal choice of q. We shall proceed as in Section 2.3 and not as in Section 2.2. The reason is that there are many extreme points of the seller's choice set. These were the focus of Section 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 Section 2.3

to equation (2.24), 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

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

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

As in Section 2.3, we first ask which function q the seller would choose if she 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 $\psi_j(\theta_j)$ 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 $\theta \in \Theta$:*

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

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

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. For example, we can simply set $t_i(\theta) = T_i(\theta_i)$ for all $i \in I$ and $\theta \in \Theta$.

The expression $\psi_i(\theta_i)$ is sometimes referred to as seller i's "virtual valuation," which is similar to the terminology that we introduced in the previous chapter. 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 valuation, provided that this valuation is at least zero.

If the buyers are symmetric—that is, 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 formats, with appropriately chosen minimum bids, are optimal. To show this, one has to derive equilibrium bidding functions for these auctions and also 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 Welfare Maximization

Suppose that the seller were not maximizing expected profits but expected welfare. Let us assume that the seller uses the following utilitarian welfare function, where each agent has equal weight:

$$\sum_{i \in I} q_i(\theta)\theta_i. \tag{3.8}$$

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. She 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 if the object is not transferred to one of the potential buyers, welfare is assumed to be zero, and hence the welfare-maximizing seller always wants to transfer the object. Because transferring to the buyer for whom θ_i is largest maximizes welfare for every type vector, it also maximizes expected welfare. Can the expected welfare-maximizing seller allocate the object to the buyer with the highest value even if she 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 the following:

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

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

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

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: Monopoly sellers who don't know buyers' values 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 $N = 2$, $\underline{\theta} = 0$, $\bar{\theta} = 1$ and that θ_i is uniformly distributed so that $F(\theta_i) = \theta_i$ for $i = 1, 2$. We begin by calculating for $i = 1, 2$:*

$$\begin{aligned}\psi_i(\theta_i) &= \theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \\ &= \theta_i - \frac{1 - \theta_i}{1} \\ &= 2\theta_i - 1.\end{aligned} \tag{3.9}$$

Note that the regularity assumption 3.1 is satisfied.

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

$$\begin{aligned}\psi_i(\theta_i) &< 0 \Leftrightarrow \\ 2\theta_i - 1 &< 0 \Leftrightarrow \\ \theta_i &< \frac{1}{2}\end{aligned} \tag{3.10}$$

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

$$\begin{aligned}\psi_1(\theta_1) &> \psi_2(\theta_2) \Leftrightarrow \\ 2\theta_1 - 1 &> 2\theta_2 - 1 \Leftrightarrow \\ \theta_1 &> \theta_2.\end{aligned} \tag{3.11}$$

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

$$\begin{aligned}\psi_1(\theta_1) &= \theta_1 - \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} \\ &= \theta_1 - \frac{1 - (\theta_1)^2}{2\theta_1} \\ &= \frac{3}{2}\theta_1 - \frac{1}{2\theta_1}\end{aligned} \tag{3.12}$$

and

$$\begin{aligned}\psi_2(\theta_2) &= \theta_2 - \frac{1 - F_2(\theta_2)}{f_2(\theta_i)} \\ &= \theta_2 - \frac{1 - 2\theta_2 + (\theta_2)^2}{2 - 2\theta_2} \\ &= \theta_2 - \frac{1 - \theta_2}{2} \\ &= \frac{3}{2}\theta_2 - \frac{1}{2}.\end{aligned} \tag{3.13}$$

Again the regularity assumption is satisfied.

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

$$\begin{aligned}\psi_1(\theta_1) &< 0 \Leftrightarrow \\ \frac{3}{2}\theta_1 - \frac{1}{2\theta_1} &< 0 \Leftrightarrow \\ \theta_1 &< \sqrt{\frac{1}{3}}\end{aligned} \tag{3.14}$$

and

$$\begin{aligned}\psi_2(\theta_2) &< 0 \Leftrightarrow \\ \frac{3}{2}\theta_2 - \frac{1}{2} &< 0 \Leftrightarrow \\ \theta_2 &< \frac{1}{3}.\end{aligned} \tag{3.15}$$

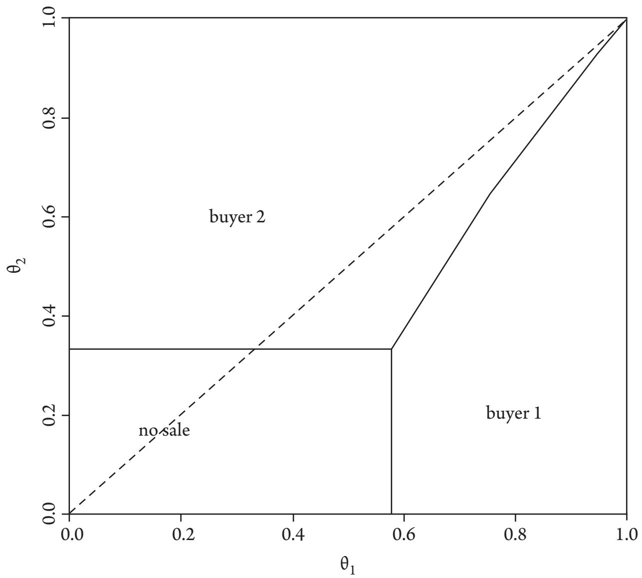

Figure 3.1 Expected Revenue-Maximizing Allocation in Example 3.2.

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

$$\psi_1(\theta_1) > \psi_2(\theta_2) \Leftrightarrow$$

$$\frac{3}{2}\theta_1 - \frac{1}{2\theta_1} > \frac{3}{2}\theta_2 - \frac{1}{2} \Leftrightarrow$$

$$\theta_2 < \theta_1 - \frac{1}{3\theta_1} + \frac{1}{3}. \tag{3.16}$$

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 buyer 1. If the good is sold, buyer 1 wins the object only in a subset of all cases where his value is higher than buyer 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.

3.3 PUBLIC GOODS

3.3.1 Setup

Our next example is a public goods problem. The theory of Bayesian 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 $N \geq 2$. These agents have to choose whether to produce some indivisible, nonexcludable 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 agents' 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.

That the public good is nonexcludable is reflected by the fact that the same variable g enters into each individual's utility function. An alternative model, which is also of interest, is a model in which individuals can be selectively excluded from consuming the public good. But we do not investigate that alternative model here.

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

$$\left(\sum_{i \in I} \theta_i\right) g - \sum_{i \in I} t_i. \tag{3.17}$$

The mechanism designer's objective is to maximize the expected value of (3.17).

3.3.2 Incentive-Compatible and Individually Rational Direct Mechanisms

As in previous parts of this book, 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 nonstochastic. 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*

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

and

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

The function q assigns to each type vector θ the collective decision about the public good in case that 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 in expectations, before agents' types are realized.

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

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

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 they have the same decision rule and 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.*

Note 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 that is ex ante budget balanced, there is an equivalent direct mechanism 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 exactly ex ante budget balanced 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 this deficit conditional on her own type. Formally, if the vector of types is θ, then agent 1's payment in the modified mechanism is

$$t_1(\theta) + \left(cq(\theta) - \sum_{i \in I} t_i(\theta) \right) - \left(cQ_1(\theta_1) - \sum_{i \in I} T_i(\theta_1) \right). \tag{3.18}$$

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'; 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

$$t_2(\theta) + cQ_1(\theta_1) - \sum_{i\in I} T_i(\theta_1). \tag{3.19}$$

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. □

Intuitively, the ex post budget balanced mechanism in the above proof is constructed so that agent 1 covers the ex post deficit, and agent 2 compensates agent 1 by paying to agent 1 a "fair" insurance premium. Note that a crucial assumption that guarantees that this construction leaves agent 1 and 2's incentives for truthful reporting of their types intact is that both agents are risk-neutral.

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 difference between revenue and cost 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:

$$\sum_{i \in I} -U_i(\underline{\theta}) + \int_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta) \, d\theta. \tag{3.20}$$

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 (3.17), she would never raise transfers larger than what is required to cover cost. From (3.17) it is then clear that the optimal decision rule is[2]

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

As welfare function (3.17) 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

$$\sum_{i \in I} t_i^*(\theta) = \begin{cases} c & \text{if } q^*(\theta) = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.22}$$

are optimal. We call these direct mechanisms "first best." They maximize welfare for any type vector θ, and therefore they also maximize ex ante expected welfare.

The following impossibility result shows that in all nontrivial 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 possible 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 not to produce the public good. These are trivial cases. For all nontrivial 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 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.

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 nontrivial 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

$$\theta_i q^*(\theta_i', \theta_{-i}) - q^*(\theta_i', \theta_{-i}) \left(c - \sum_{j \neq i} \theta_j \right)$$
$$= q^*(\theta_i', \theta_{-i}) \left(\sum_{j=1}^{N} \theta_j - c \right). \tag{3.23}$$

Thus agent i's utility is the 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 3.8. 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 the decision rule is first best and if, moreover, every agent's payment consists of two terms: first, a part that depends on the agent's report and that has the effect of aligning the agent's incentives with social welfare; second, 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."[3] VCG mechanisms play a central part in the general theory of dominant strategy mechanisms. They are given detailed attention in Chapter 7. 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 ex ante expected budget surplus—that is, revenue minus cost—than the pivot mechanism.*

Proof

Expression (3.20) shows that 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 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 the trivial cases.

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

Proof

We show that the ex post budget surplus of the pivot mechanism is always nonpositive and with positive probability negative. This implies that the ex ante expected budget 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, $N\underline{\theta} < c$. 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

$$
\begin{aligned}
&\sum_{i \in P}\left(c - \sum_{j \neq i}\theta_j\right) + \sum_{i \in NP}\underline{\theta} \\
&\quad = Pc - P\sum_{j \in NP}\theta_j - (P-1)\sum_{j \in P}\theta_j + \sum_{i \in NP}\underline{\theta} \\
&\quad = Pc - (P-1)\sum_{j \in NP}\theta_j - (P-1)\sum_{j \in P}\theta_j - \sum_{i \in NP}(\theta_i - \underline{\theta}) \\
&\quad = Pc - (P-1)\sum_{j \in I}\theta_j - \sum_{i \in NP}(\theta_i - \underline{\theta}). \qquad (3.24)
\end{aligned}
$$

By construction in states in which $q^*(\theta) = 1$, we have $c \leq \sum_{j \in I}\theta_j$. Therefore the last line above is not larger than

$$
\begin{aligned}
&Pc - (P-1)c - \sum_{i \in NP}(\theta_i - \underline{\theta}) \\
&\quad = c - \sum_{i \in NP}(\theta_i - \underline{\theta}). \qquad (3.25)
\end{aligned}
$$

and thus the transfers are not more than c.

States in which the public good is produced and some agents are pivotal occur with positive probability under our assumptions. 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

$$\int_{\Theta} q(\theta)\left(\sum_{i\in I}\theta_i - c\right) f(\theta)\, d\theta. \tag{3.26}$$

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:

for every $i \in I$ the function Q_i is monotonically increasing

(*incentive constraint*), (3.27)

for every $i \in I$ *we have* $U_i(\underline{\theta}) \geq 0$

(*individual rationality constraint*), (3.28)

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

(*budget constraint*), (3.29)

where the budget constraint uses expression (3.20) for the expected revenue.

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

$$\int_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta)\, d\theta \geq 0 \tag{3.30}$$

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 (3.27). Then we shall discuss conditions under which the solution to the problem without constraint (3.27) actually happens to satisfy condition (3.27) 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, pp. 217 and 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

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

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

$$\int_{\Theta} q(\theta) \left[\sum_{i \in I} \left(\theta_i - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta)\, d\theta > 0. \tag{3.32}$$

We don't go through the details of checking the applicability of these results. However, 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 (3.31) as

$$\int_{\Theta} q(\theta)(1 + \lambda) \left[\sum_{i \in I} \left(\theta_i - \frac{\lambda}{1 + \lambda} \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \right) - c \right] f(\theta)\, d\theta. \tag{3.33}$$

It is evident, from pointwise 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:

$$q^*(\theta) = \begin{cases} 1 & \text{if } \sum_{i \in I} \theta_i > c + \sum_{i \in I} \left(\frac{\lambda}{1+\lambda} \frac{1-F_i(\theta_i)}{f_i(\theta_i)} \right), \\ 0 & \text{otherwise.} \end{cases} \tag{3.34}$$

Note that we must have $\lambda > 0$ if Proposition 3.7 applies, because with $\lambda = 0$ the rule (3.34) becomes the first best rule.

Now we introduce an assumption under which the rule (3.34) satisfies the monotonicity constraint (3.27) 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; that is, 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

$$\theta_i - \frac{\lambda}{1+\lambda} \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \tag{3.35}$$

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 (3.35). This implies that the second best rule in (3.34) satisfies the monotonicity condition (3.27). Therefore, we can conclude the following:

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

(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 implement 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. One obtains

$$\int_{\Theta} q(\theta)\left[\sum_{i\in I}\left(\theta_i - \frac{1-F_i(\theta_i)}{f_i(\theta_i)}\right) - c\right] f(\theta)\,d\theta \tag{3.36}$$

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 $\theta \in \Theta$:*

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

(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

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

$$\theta_1 + \theta_2 > \frac{1+\lambda}{1+2\lambda}c + \frac{\lambda}{1+2\lambda}2. \tag{3.37}$$

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 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 (3.30) 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 mechanisms with threshold s. We distinguish two cases.

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

$$C(s) = \left(1 - \frac{1}{2}s^2\right)c. \tag{3.38}$$

Next we calculate the expected payment by the agent 1:

$$\int_{\Theta} q(\theta)\left(\theta_1 - \frac{1-F_1(\theta_1)}{f_1(\theta_1)}\right) f(\theta)\, d\theta$$
$$= \int_0^s \int_{s-\theta_1}^1 (2\theta_1 - 1)\, d\theta_2 d\theta_1 + \int_s^1 \int_0^1 (2\theta_1 - 1)\, d\theta_2 d\theta_1$$
$$= -\frac{1}{3}s^3 + \frac{1}{2}s^2. \tag{3.39}$$

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

$$R(s) = -\frac{2}{3}s^3 + s^2. \tag{3.40}$$

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

$$C(s) = \frac{1}{2}(2-s)^2 c. \tag{3.41}$$

The expected payment by agent 1 is given by

$$\int_{\Theta} q(\theta)\left(\theta_1 - \frac{1-F_1(\theta_1)}{f_1(\theta_1)}\right) f(\theta)\, d\theta$$
$$= \int_{s-1}^1 \int_{s-\theta_1}^1 (2\theta_1 - 1)\, d\theta_2 d\theta_1$$
$$= \frac{1}{6} - \frac{1}{2}(s-1)^2 + \frac{1}{3}(s-1)^3. \tag{3.42}$$

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

$$R(s) = \frac{1}{3} - (s-1)^2 + \frac{2}{3}(s-1)^3. \tag{3.43}$$

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

$$0 < s < 1 \Rightarrow D'(s) = -2s^2 + (2+c)s = s(2(1-s)+c) > 0; \tag{3.44}$$
$$1 < s < 2 \Rightarrow D'(s) = -2(s-1) + 2(s-1)^2 + (2-s)c$$
$$= (2-s)(c - 2(s-1)). \tag{3.45}$$

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

$$\begin{aligned} c - 2(s-1) > 0 \Leftrightarrow \\ s < 1 + \frac{c}{2}. \end{aligned} \tag{3.46}$$

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

$$D(0) = -c < 0 \tag{3.47}$$

and

$$D(2) = 0. \tag{3.48}$$

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

$$\begin{aligned} D\left(1 + \frac{c}{2}\right) &= \frac{1}{3} - \left(\frac{c}{2}\right)^2 + \frac{2}{3}\left(\frac{c}{2}\right)^3 - \frac{1}{2}\left(1 - \frac{c}{2}\right)^2 c \\ &= \frac{1}{3} - \frac{1}{2}c + \left(\frac{c}{2}\right)^2 - \frac{1}{3}\left(\frac{c}{2}\right)^3. \end{aligned} \tag{3.49}$$

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*

$$\begin{aligned} &-\frac{1}{2} + \frac{c}{2} - \frac{1}{2}\left(\frac{c}{2}\right)^2 \\ &= \frac{c}{2}\left(1 - \frac{c}{2}\right) - \frac{1}{2} \\ &\leq \frac{1}{4} - \frac{1}{2} \\ &= -\frac{1}{4}. \end{aligned} \tag{3.50}$$

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

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* $D(1)$:

$$\begin{aligned} D(1) = \frac{1}{3} - \frac{1}{2}c > 0 \Leftrightarrow \\ c < \frac{2}{3}. \end{aligned} \tag{3.51}$$

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:*

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

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

$$\frac{1}{3} - (s-1)^2 + \frac{2}{3}(s-1)^3 - \frac{1}{2}(2-s)^2 c = 0. \tag{3.53}$$

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

$$s = \frac{1}{2} + \frac{3}{4}c. \tag{3.54}$$

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*

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

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

To illustrate our findings, we have plotted in Figure 3.2 the optimal threshold s for the utilitarian case (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*

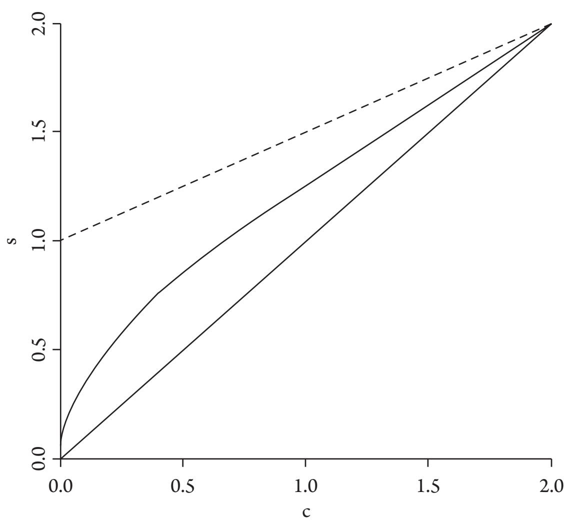

Figure 3.2 Thresholds in First Best, Second Best, and Profit-Maximizing Mechanisms.

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 institutions, such as stock exchanges or commodity markets. Methodologically, it is interesting to see in this example 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 she sells the good and receives a transfer payment t is equal to t. If she does not sell the good and receives a transfer t, then her 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*

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

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 no trade takes place $(q(\theta) = 0)$. We shall refer to q 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[4] that $t_S = t_B$, that is, 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 a more general framework in which these variables are not identical and then to introduce an ex post budget balance condition which requires 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 $Q_i : [\underline{\theta}_i, \bar{\theta}_i] \to [0, 1]$ and $T_i : [\underline{\theta}_i, \bar{\theta}_i] \to \mathbb{R}$ 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. Incentive compatibility 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. Intuitively, this is because high types of the seller are the ones least willing to trade. 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

$$T_S(\theta_S) = T_S(\bar{\theta}_S) + (1 - Q_S(\bar{\theta}_S))\bar{\theta}_S - (1 - Q_S(\theta_S))\theta_S - \int_{\theta_S}^{\bar{\theta}_S} (1 - Q_S(x))\, dx \text{ for all } \theta_S. \tag{3.55}$$

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 her utility would be if she 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:

$$\theta_S(1 - Q_S(\theta_S')) + T_S(\theta_S'). \tag{3.56}$$

This exceeds θ_S by

$$T_S(\theta_S') - \theta_S Q_S(\theta_S'). \tag{3.57}$$

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

$$U_S(\bar{\theta}_S) - \int_\Theta (1 - q(\theta)) \left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)\, d\theta \tag{3.58}$$

and the buyer's expected payment is

$$-U_B(\underline{\theta}_B) + \int_\Theta q(\theta) \left(\theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)}\right) f(\theta)\, d\theta. \tag{3.59}$$

3.4.3 Welfare Maximization

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

$$\begin{aligned} &q(\theta)\theta_B - t_B + (1 - q(\theta))\theta_S + t_S \\ &\quad = \theta_S + q(\theta)\,(\theta_B - \theta_S) + t_S - t_B. \end{aligned} \tag{3.60}$$

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

$$q^*(\theta) = \begin{cases} 1 & \text{if } \theta_B \geq \theta_S; \\ 0 & \text{otherwise,} \end{cases} \tag{3.61}$$

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. These mechanisms maximize expected welfare because they maximize welfare ex post for any type vector θ.

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 expost 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, regardless of agents' realized types. The condition $\underline{\theta}_S \geq \bar{\theta}_B$ implies that efficiency never requires trade, regardless of agents' realized types. Thus, these are trivial cases. In all nontrivial cases, there is no incentive-compatible and individually rational first best mechanism.

Proof

The "if" part of Proposition 3.12 is trivial. If $\underline{\theta}_B \geq \bar{\theta}_S$, a mechanism where trade always takes place and the buyer always pays the seller some price $p \in [\bar{\theta}_S, \underline{\theta}_B]$ is first best and has the required properties. 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.

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 neither of the conditions of Proposition 3.12 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 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

$$q^*(\theta_S', \theta_B)(\theta_B - \theta_S). \tag{3.62}$$

Note that the first best decision rule q^* maximizes the social surplus $q^*(\theta)(\theta_B - \theta_S)$.[5] 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 $\bar{\theta}_S$ and $\underline{\theta}_B$. From the payment rule, it is clear that the type $\bar{\theta}_S$ receives utility $\bar{\theta}_S$ whatever the buyer's type, and the type $\underline{\theta}_B$ receives utility zero whatever the seller's type. Individual rationality follows. □

Lemma 3.10 *The ex ante expected difference between the buyer's and the seller's transfers 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 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 ex ante expected difference between the buyer's and the seller's transfers is negative under the pivot mechanism if the condition of Proposition 3.12 is violated, that is, 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 nonempty 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 the first term in this expression, θ_S, is not affected by the mechanism designer's decisions. Therefore, we shall assume that the mechanism designer seeks to maximize:

$$\int_\Theta q(\theta)\,(\theta_B - \theta_S) f(\theta)\,d\theta. \tag{3.63}$$

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 as follows:

Q_S is increasing and Q_B is decreasing

(*incentive compatibility*); (3.64)

$U_S(\bar{\theta}_S) \geq \bar{\theta}_S$ and $U_B(\underline{\theta}_B) \geq 0$

(*individual rationality*); (3.65)

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

(*budget balance*). (3.66)

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

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

Next, we set the choice variables $U_S(\bar{\theta}_S)$ and $U_B(\underline{\theta}_B)$ to their lowest values, thus relaxing the budget constraint, and then we rewrite the budget constraint as

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

$$\bar{\theta}_S - \int_{\Theta}\left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)\, d\theta. \tag{3.68}$$

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 (3.68) 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, pp. 217 and 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:

$$\int_{\Theta} q(\theta)\left[(1+\lambda)\theta_B - \lambda\frac{1-F_B(\theta_B)}{f_B(\theta_B)} - (1+\lambda)\theta_S - \lambda\frac{F_S(\theta_S)}{f_S(\theta_S)}\right] f(\theta)\, d\theta - \lambda K \tag{3.69}$$

and moreover the budget constraint (3.68) 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:

$$q(\theta) = \begin{cases} 1 & \text{if } \theta_B - \frac{\lambda}{1+\lambda}\frac{1-F_B(\theta_B)}{f_B(\theta_B)} \geq \theta_S + \frac{\lambda}{1+\lambda}\frac{F_S(\theta_S)}{f_S(\theta_S)}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.70}$$

Can we have $\lambda = 0$? Then the optimal decision rule (3.70) 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 (3.64) of the mechanism designer's maximization problem.

Assumption 3.3 *The seller's distribution function F_S is "regular," that is,*

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

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

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

is monotonically increasing.

If F_S and F_B are regular,

$$\theta_S + \frac{\lambda}{1+\lambda}\frac{F_S(\theta_S)}{f_S(\theta_S)} \quad \text{and} \quad \theta_B - \frac{\lambda}{1+\lambda}\frac{1-F_B(\theta_B)}{f_B(\theta_B)} \tag{3.71}$$

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 the following:

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 it maximizes expected welfare among all such mechanisms if and only if:*

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

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

(ii) Exact budget balance holds:

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

(iii) The payment rules create incentive compatibility, that is, for all $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$ and all $\theta_S \in [\underline{\theta}_S, \bar{\theta}_S]$:

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

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

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 increment, 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 (3.58) and (3.59), we can thus write expected revenue as

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

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

$$\int_{\Theta} q(\theta)\left(\theta_B - \frac{1 - F_B(\theta_B)}{f_B(\theta_B)} - \theta_S - \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)\, d\theta. \tag{3.73}$$

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

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 as follows:*

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

(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

$$\theta_B - \frac{\lambda}{1+\lambda}\frac{1-F_B(\theta_B)}{f_B(\theta_B)} > \theta_S + \frac{\lambda}{1+\lambda}\frac{F_S(\theta_S)}{f_S(\theta_S)} \Leftrightarrow \tag{3.74}$$

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

$$\theta_B - \theta_S > \frac{\lambda}{1+2\lambda}. \tag{3.76}$$

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 expected payment is given by

$$\begin{aligned}
&\int_\Theta q(\theta)\left(\theta_B - \frac{1-F_B(\theta_B)}{f_B(\theta_B)}\right) f(\theta)\, d(\theta) \\
&= \int_s^1 \int_0^{\theta_B - s} (2\theta_B - 1)\, d\theta_S d\theta_B \\
&= \frac{1}{3}s^3 - \frac{1}{2}s^2 + \frac{1}{6}.
\end{aligned} \tag{3.77}$$

The seller's expected transfer is given by

$$\begin{aligned}
&\bar{\theta}_S - \int_\Theta (1-q(\theta))\left(\theta_S + \frac{F_S(\theta_S)}{f_S(\theta_S)}\right) f(\theta)\, d\theta \\
&= 1 - 2\int_\Theta (1-q(\theta))\, \theta_S d\theta \\
&= 1 - 2\left(\int_0^{1-s}\int_0^{\theta_S+s} \theta_s\, d\theta_B d\theta_S + \int_{1-s}^1\int_0^1 \theta_s\, d\theta_B d\theta_S\right) \\
&= \frac{1}{3}(1-s)^3.
\end{aligned} \tag{3.78}$$

Budget balance is achieved when

$$\frac{1}{3}s^3 - \frac{1}{2}s^2 + \frac{1}{6} = \frac{1}{3}(1-s)^3. \tag{3.79}$$

This has two solutions:

$$s = \frac{1}{4} \quad \textit{and} \quad s = 1. \tag{3.80}$$

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*

$$\theta_B - \theta_S > \frac{1}{4}. \tag{3.81}$$

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*

$$\theta_B - \theta_S > \frac{1}{2}. \tag{3.82}$$

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 REMARKS ON THE LITERATURE

The classic paper on expected revenue-maximizing single unit auctions is Myerson (1981). In preparing this chapter 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 setup 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 (2009). For the bilateral trade problem the classic paper is Myerson and Satterthwaite (1983). My exposition of the bilateral trade problem has also benefited from the books by Krishna and Milgrom cited above.

3.6 PROBLEMS

(a) Consider the symmetric version of the setup of Section 3.2; that is, the distributions F_i are the same for all i. Suppose the seller uses an all pay auction, that is, the highest bidder wins the object, but all bidders have to pay their bids. Use the revenue equivalence theorem to determine equilibrium bids, assuming that there is an equilibrium that allocates the good to the bidder with the highest type. Is there such an equilibrium?

(b) 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.

(c) 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 $i \in I$. For each case, determine payment rules that guarantee incentive compatibility, individual rationality, and ex post budget balance of the second best mechanism.

(d) 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.

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 explained 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 are the same for all types of the agent, 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 agents' beliefs. Imperfect information about others' beliefs seems at least as pervasive as imperfect information about others' preferences, however.

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 may not come true.

In this chapter we revisit the examples of the previous chapter, 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 chapter the mechanism designer was willing to trust a single hypothesis about agents' beliefs, in this chapter 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 agents do. In short: Each type is required to have a dominant strategy.[1] 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 chapter are a strict subset of those considered in the previous chapter. 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 and instead opting out. 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 called the individual rationality constraint of that chapter an "interim individual rationality constraint." In that chapter, agents evaluated the outside option once they knew their type using a belief about the other agents' types that was known to the mechanism designer. The mechanism designer had to offer a mechanism that ensured that when conducting this evaluation, agents didn't have an incentive to opt out. Ex post individual rationality restricts the set of mechanisms that the mechanism designer can recommend further. This is because, obviously, whenever participation dominates nonparticipation 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 chapter, but focus on the characterization of classes of mechanisms that satisfy incentive compatibility, individual rationality, and budget balance requirements.

The argument that a mechanism designer who is uncertain about agents' beliefs about each other will choose a dominant strategy mechanism will be further scrutinized in Chapter 10. We shall find in that chapter that upon more careful investigation the argument is not as persuasive as it seems. Nonetheless, dominant strategy mechanisms have played a very significant role in the literature. This is why we cover them here in some detail.

We conclude this introduction with one formal clarification. We shall say in this book 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 "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 book, we shall be sloppy with terminology 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 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 Setup

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 $\{0, 1, 2, \ldots, N\} \times \mathbb{R}^N$—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 requirements 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 nonnegative 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 requirements for the strategy seller proposed by the seller that we 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 this chapter 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 mechanism design, as for Bayesian mechanism design, we can restrict our attention without loss of generality to direct mechanisms. Moreover, we can require truth telling to be 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 all $\theta_{-i} \in \Theta_{-i}$ we have*

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

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 $\theta_{-i} \in \Theta_{-i}$ we have*

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

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 compatibility 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) $q_i(\theta_i, \theta_{-i})$ *is increasing in* θ_i.
(ii) *For every* $\theta_i \in [\underline{\theta}, \bar{\theta}]$ *we have*

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

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 types 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 pin down 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."[2] 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 some of 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 $\psi_i : [\underline{\theta}, \bar{\theta}] \rightarrow \mathbb{R}$ for $i \in I$:

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

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 functions 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 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 one. 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 $\psi_i : [\underline{\theta}, \bar{\theta}] \rightarrow \mathbb{R}$ 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 where

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

for all $\theta \in \Theta$.

It is worth considering the transfer rule in detail. One can interpret the transfer payment formula in Definition 4.3 as the expected transfers resulting from the following rule. If bidder i does not win the auction, either because her bid is too low or because she won, but there was a tie and she was not selected, then bidder i does not have to pay anything. If bidder i does win the auction, then bidder i's payment equals the lowest type that she might have had that would have allowed her to win the auction with positive probability. The assumed continuity of ψ_i guarantees that this minimum exists. If bidder i wins with probability 1, then the minimum will either be the value of θ_i at which i would have tied, or, if no such value exists, $\underline{\theta}$. If bidder i ties, then the minimum equals her θ_i.

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.

Consider next the case that $q_i(\theta_i, \theta_{-i}) = 1$. Note first that reporting θ_i truthfully yields strictly positive utility, because buyer i obtains the object with probability 1 and pays less than θ_i. 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 $0 < q_i(\theta_i', \theta_{-i}) < 1$ (i.e., there is a tie), her utility decreases because she obtains the object with probability less than 1 and, if she obtains it, pays the same as she would if she reported truthfully. In other words, she obtains the positive surplus that results from truthful reporting with probability less than 1. Lastly, if buyer i reports a type θ_i' for which $q_i(\theta_i', \theta_{-i}) = 0$, then she loses the 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 already also shows that a buyer's utility is always nonnegative 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 (a) loses the auction and has utility zero or (b) 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. First, 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, that gives the lowest types expected utility zero, and that ensures dominant strategy incentive compatibility. This auction 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 is always at least one that makes 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 second price 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 Setup

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, which is in contrast with the auction example. 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, nonexcludable 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 $[\underline{\theta}, \bar{\theta}]$ 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 is 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 auction example of the previous section.

Regarding budget balance, the ex ante version of this requirement that we introduced in Section 3.3 has no meaning in our current context, because there is no prior probability measure over Θ. We shall instead only consider an ex post version. In the ex post version of budget balance in Section 3.3, which is Definition 3.5, we wrote the ex post budget constraint as an inequality rather than an equality. By contrast, in the current chapter, we shall write it as an equality. In the previous chapter, the distinction between an equality budget constraint and an inequality budget constraint was not of much 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 if for every $i \in I$ and for every $\theta_{-i} \in \Theta_{-i}$, there are a type[3] $\hat{\theta}_i \in \mathbb{R}$ and two payments τ_i and $\hat{\tau}_i \in \mathbb{R}$ such that*

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

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

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

$$q(\theta_i, \theta_{-i}) = 1 \text{ and } t_i(\theta_i, \theta_{-i}) = \hat{\tau}_i;$$

$$\hat{\tau}_i - \tau_i = \hat{\theta}_i.$$

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 $q(\theta_i, \theta_{-i}) = 1$ 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 $q(\theta_i, \theta_{-i}) = 0$ for all $\theta_i \in [\underline{\theta}, \bar{\theta}]$).

Proof

First we show sufficiency. Consider an agent i with type θ_i, and some fixed $\theta_{-i} \in \Theta_{-i}$. If the outcome is independent of agent i's report, then obviously reporting the truth is a dominant strategy. Otherwise, agent i has the choice between reporting a type for which the public good is produced and the agent has to pay $\hat{\tau}_i$, or a type for which the public good is not produced and the agent has to pay τ_i. If in truth $\theta_i \geq \hat{\theta}_i$, then reporting a type above $\hat{\theta}_i$, such as the true type, is optimal, because the agent obtains more than $\hat{\theta}_i$ in comparison to reporting a low type for which the public good is not produced, and the payment that he has to make increases by only $\hat{\tau}_i - \tau_i = \hat{\theta}_i$. A symmetric argument applies if $\theta_i \leq \hat{\theta}_i$. Finally, if $\theta_i = \hat{\theta}_i$, agent i is indifferent between the outcome where the public good is produced and the outcome where it is not produced, and therefore reporting θ_i is optimal. 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 $q(\theta_i, \theta_{-i}) = 0$. 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 $q(\theta_i, \theta_{-i}) = 1$.

Consider first the case that $q(\theta_i, \theta_{-i}) = 0$ 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 $\hat{\tau}_i = \tau_i + \hat{\theta}_i$, all conditions of Proposition 4.5 hold. The case that $q(\theta_i, \theta_{-i}) = 1$ for all $\theta_i \in [\underline{\theta}, \bar{\theta}]$ can be dealt with symmetrically.

Now suppose that $q(\theta_i, \theta_{-i}) = 0$ for some θ_i and $q(\theta_i, \theta_{-i}) = 1$ 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 if 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 the ψ_i's be some arbitrary strictly increasing functions. To simplify the exposition, we assume that the functions ψ_i are continuous. The decision rules in the mechanisms defined below are of the same form as the decision rule in the expected welfare- or profit-maximizing mechanisms in Section 3.3. 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 incentive-compatible and interim 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 $i \in I$ we have

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

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 $i \in I$ we have $u_i(\underline{\theta}, \theta_{-i}) = 0$ 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 Exact Budget Balance

Now we ask which direct mechanisms are dominant strategy incentive-compatible and ex post individually rational and satisfy ex post budget balance. In this subsection we only provide a partial answer to this question. The answer that we give is for the case that there are only two 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 $\{\theta | q(\theta) = 1\}$ 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*

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

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 that any mechanism satisfying these conditions must be of the form displayed in Proposition 4.8.

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

Suppose that $\tilde{\theta}_1 > \hat{\theta}_1$. By Proposition 4.5, $t_1(\tilde{\theta}_1, \hat{\theta}_2) = \tilde{\theta}_1$ and $t_2(\tilde{\theta}_1, \hat{\theta}_2) = \hat{\theta}_2$. Budget balance requires $\tilde{\theta}_1 + \hat{\theta}_2 = c$. By Proposition 4.5, we have that $q(\tilde{\theta}_1, \tilde{\theta}_2) = 1$, $t_1(\tilde{\theta}_1, \tilde{\theta}_2) = \hat{\theta}_1$ and $t_2(\tilde{\theta}_1, \tilde{\theta}_2) = \hat{\theta}_2$. Thus, the sum of contributions if types are $(\tilde{\theta}_1, \tilde{\theta}_2)$ is $\hat{\theta}_1 + \hat{\theta}_2$. If $\tilde{\theta}_1 > \hat{\theta}_1$, then the sum of these contributions would be strictly less than $\tilde{\theta}_1 + \hat{\theta}_2$, of which we have just shown that it equals c. Therefore, budget balance would be violated. (The same reasoning applies if $\tilde{\theta}_2 > \hat{\theta}_2$.) We conclude that $\tilde{\theta}_1 = \hat{\theta}_i$ for $i \in \{1, 2\}$. Hence $\theta_i \geq \hat{\theta}_1$ 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 $q(\theta_1, \theta_2) = 0$. 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 $\tau_i = \hat{\theta}_i$, the mechanism is of the form described in the Proposition.

Consider next the case is that $\hat{\theta}_i = 0$ for both agents i. In this case 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—that is, 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. We have a contradiction with budget balance at both type vectors. Therefore, $\tau_1(\theta_2)$ is constant. We write τ_1 for this constant. Budget balance then requires $\tau_2(\theta_1)$ to be constant too. Again we find that the mechanism is of the form described in the Proposition.

The remaining case—where $\hat{\theta}_i = 0$ for one, but not both, agents i—is dealt with analogously. □

Proposition 4.5 does not generalize to the case of three or more agents, as the following example shows.[4] Suppose $N = 3$. The public good is produced if and only if $\theta_1 + \theta_2 \geq b$, where b is some number that we determine in a moment. If the public good is produced, then agent 1 pays $b - \theta_2$ and agent 2 pays $b - \theta_1$. The smallest total payment that we thus obtain from agents 1 and 2 is $2(b - \bar{\theta})$. Now let us choose b such that $2(b - \bar{\theta}) \geq c$, so that agent 1 and 2's payments alone cover the cost of producing the public good. It is straightforward to see that this mechanism has dominant strategies for agents 1 and 2 and that it is ex post individually rational for agents 1 and 2. The mechanism will never have an ex post budget deficit, but it potentially leads to an ex post budget surplus and is thus not exactly budget balanced. This is where agent 3 comes in. If the public good is produced and the sum of agent 1 and 2's payments exceed c, then agent 3 gets the surplus. Agent 3 has no influence on whether or not the public good is produced, nor does his type affect agent 1 and 2's payments. Obviously, the mechanism is now also dominant strategy incentive-compatible and ex post individually rational for agent 3. It is now also ex post budget balanced, and it is clearly not of the form described in Proposition 4.5.[5]

4.4 BILATERAL TRADE

4.4.1 Setup

We recapitulate briefly the setup 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. We also assume deterministic trading rules and thus consider the same class of mechanisms that we introduced in Definition 3.9. We require these to satisfy dominant strategy incentive compatibility, ex post individual rationality, and ex post exact 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 brevity, 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*

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

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

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

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 $\theta_B \in [\underline{\theta}_B, \bar{\theta}_B]$ we have*

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

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

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

4.4.3 Canonical Mechanisms

We introduce again a class of mechanisms that encompass 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*

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

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

and

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

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 pays 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 Exact Budget Balance

Now we characterize direct mechanisms that are dominant strategy incentive-compatible and ex post individually rational and satisfy ex post exact budget balance, i.e. $t_B(\theta) = t_S(\theta)$ for all θ. 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 exactly 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

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

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 REMARKS 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. These results have recently been significantly extended in Gershkov et al. (2013a). I have also used Section 5.2 of Vijay Krishna's *Auction Theory* for my Section 4.2. Some dominant strategy mechanisms for public goods economies with more than two agents and exact budget balance are characterized in Serizawa (1999). Dominant strategy mechanisms for bilateral trade are discussed in Hagerty and Rogerson (1987).

4.6 PROBLEMS

(a) Suppose there are only two bidders in an auction. Is there a canonical auction such that bidder 1 wins with probability 1 if only if his bid is larger than twice bidder 2's bid? If there is such a canonical auction, what is the payment rule?

(b) 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.

(c) Are the mechanisms in Proposition 4.5 canonical mechanisms for the public good problem?

(d) Give an example of a canonical mechanism for the public good problem with more than two agents that satisfies ex post budget balance as a weak inequality; that is, for every type vector, payments are at least as large as the cost of producing the public good, but not always as an exact equality. Construct your example such that for every agent there is at least one type vector of the other agents such that the agent's own type determines whether the public good is produced or not.

(e) Consider the following mechanism for the bilateral trade problem: Buyer and seller both name their value of the item. The agent who names the higher value obtains the good, and he or she pays to the other agent the lower of the two values. If there is a draw, then a randomly chosen agent obtains the good, and he or she pays his or her own value. Which of the three conditions does this mechanism satisfy: dominant strategy incentive compatibility, ex post individual rationality, ex post exact budget balance?

5

INCENTIVE COMPATIBILITY

5.1 INTRODUCTION

In this and the next two chapters we shall analyze more general versions of the examples treated in Chapters 2–4. Some of what we shall study is quite abstract. I hope it is not discouragingly abstract. By considering more general models than before, we can understand the most important special features of the examples that we have considered so far. At the same time, we learn about results that can be applied to other examples that we may want to study at some point in the future. The structure of the next three chapters mirrors that of Chapters 2–4, with Chapter 5 generalizing Chapter 2, and so on. Also, as before, the current chapter treats screening problems with just a single agent. Many of the results developed here will apply directly to the models treated in Chapters 6 and 7.

5.2 SETUP

There are a mechanism designer and an agent. They have to choose an alternative a out of some set A of mutually exclusive alternatives. The agent's utility if alternative a

is chosen, and the agent pays transfer t is

$$u(a, \theta) - t.$$

Here, θ is the agent's type. Note that the setup includes the case in which A is the set of lotteries over some set of nonrandom alternatives, and u is an expected utility function. The set of possible types of the agent is Θ, which we now take to be some abstract, nonempty set. Our interest is in characterizing incentive-compatible mechanisms. The revelation principle holds, and we can restrict attention to direct mechanisms.

Definition 5.1 *A direct mechanism (q, t) consists of a mapping*

$$q : \Theta \to A,$$

which maps every type into an alternative, and a mapping

$$t : \Theta \to \mathbb{R},$$

which indicates for each type the transfer that the agent makes (positive t) or receives (negative t).

We call q the "decision rule." Note that we restrict ourselves to mechanisms where the payment is deterministic. Because we assume that the agent is risk-neutral, and since we are only concerned with characterizations of incentive compatibility and not, say, with ex post budget balance, this is without loss of generality.

Definition 5.2 *A direct mechanism is "incentive-compatible" if for all $\theta, \theta' \in \Theta$ we have*

$$u(q(\theta), \theta) - t(\theta) \geq u(q(\theta'), \theta) - t(\theta').$$

5.3 WEAK MONOTONICITY

In this and the next section our objective is to characterize the set of *all* decision rules q that may form part of an incentive-compatible direct mechanism. We shall call such decision rules "implementable."[1] Later, we shall return to a study of the associated payment rules.

Definition 5.3 *A decision rule q is "implementable" if there is a transfer rule $t : \Theta \to \mathbb{R}$ such that (q, t) is incentive-compatible.*

We start with a result that can be proved by an appropriate adaptation of the argument that led in Chapter 2, Lemma 2.1, to the conclusion that decision rules that can be part of an incentive-compatible direct mechanism are increasing in θ. We have to adapt the notion of monotonicity here because the set Θ of types does not necessarily consist of numbers, so that it may not be clear how to order the elements of Θ.

Definition 5.4 *A decision rule q is "weakly monotone" if for all $\theta^1, \theta^2 \in \Theta$, if $q(\theta^1) = a^1$ and $q(\theta^2) = a^2$, then*

$$u(a^1, \theta^1) - u(a^2, \theta^1) \geq u(a^1, \theta^2) - u(a^2, \theta^2).$$

This definition considers a situation where the alternative chosen by q for a type θ^1 is alternative a^1; but if we change the agent's type from θ^1 to θ^2, the decision becomes some alternative a^2. The definition says that the utility difference between a^1 and a^2 must decrease as we switch the agent's type from θ^1 to θ^2. We can interpret the utility difference as the agent's "willingness to pay for alternative a^1 replacing a^2." The definition says that for weak monotonicity the agent's willingness to pay for a^1 replacing a^2 must be at least as large when the agent is of type θ^1 as it is when the agent is of type θ^2. In Section 5.6 we shall relate this notion of monotonicity to the monotonicity of decision rules as defined in earlier chapters. For the moment, it is enough to note that the argument in the proof of the following proposition is very similar to the arguments that we used earlier to prove the necessity of monotonicity.

Proposition 5.1 *Suppose q is implementable, then q is weakly monotone.*

Proof
Incentive compatibility implies that for all $\theta^1, \theta^2 \in \Theta$, if $q(\theta^1) = a^1$ and $q(\theta^2) = a^2$, then

$$\begin{aligned} &u(a^1, \theta^1) - t(\theta^1) \geq u(a^2, \theta^1) - t(\theta^2) \Leftrightarrow \\ &u(a^1, \theta^1) - u(a^2, \theta^1) \geq t(\theta^1) - t(\theta^2) \end{aligned} \tag{5.1}$$

and

$$\begin{aligned} &u(a^1, \theta^2) - t(\theta^1) \leq u(a^2, \theta^2) - t(\theta^2) \Leftrightarrow \\ &u(a^1, \theta^2) - u(a^2, \theta^2) \leq t(\theta^1) - t(\theta^2). \end{aligned} \tag{5.2}$$

These two inequalities imply

$$u(a^1, \theta^1) - u(a^2, \theta^1) \geq u(a^1, \theta^2) - u(a^2, \theta^2), \tag{5.3}$$

which means that q is weakly monotone. □

	θ^1	θ^2	θ^3
a	0	-1	1
b	1	0	-1
c	-1	1	0

Figure 5.1 Utility Function for the Counterexample.

Is weak monotonicity sufficient for implementability? We shall now provide an example that proves that this is not the case. Suppose that the set of alternatives has three elements: $A = \{a, b, c\}$. There are three types: $\Theta = \{\theta^1, \theta^2, \theta^3\}$. The utility function $u(a, \theta)$ is described in Figure 5.1.

Now suppose that we want to implement the decision rule that gives each type their middle alternative: $q(\theta^1) = a$, $q(\theta^2) = b$, $q(\theta^3) = c$. Let's check that this decision rule is weakly monotone. We begin by checking for types θ^1 and θ^2 the inequality $u(a, \theta^1) - u(b, \theta^1) \geq u(a, \theta^2) - u(b, \theta^2)$. Plugging in numbers, we can verify $0 - 1 \geq -1 - 0$. Replacing θ^2 by θ^3, and b by c, we get $0 - (-1) \geq 1 - 0$. All other inequalities that we need to check are similar.

However, there are no transfers that make this decision rule incentive-compatible. Without loss of generality, we can set the transfer associated with a equal to zero. We denote the transfers associated with b and c by x and y, respectively. Type θ^1 will report her type truthfully if

$$0 \geq 1 - x \quad \text{and} \quad 0 \geq -1 - y \tag{5.4}$$

For type θ^2 the corresponding conditions are

$$-x \geq -1 \quad \text{and} \quad -x \geq 1 - y \tag{5.5}$$

For type θ^3 the inequalities are

$$-y \geq 1 \quad \text{and} \quad -y \geq -1 - x \tag{5.6}$$

Inspection shows that these inequalities can be rewritten as three equalities: $x = 1$, $y = -1$, and $x - y = -1$. Obviously, no numbers x, y satisfy these three equalities. Therefore, there are no transfers that make q incentive-compatible.

It is not surprising that one can find an example in which weak monotonicity is not enough to guarantee that a decision rule can be made incentive-compatible by appropriate transfers. Weak monotonicity ensures that for every pair of types there are transfers such that none of the two types has an incentive to pretend to be the other type. The property that we are looking for is, however, that there are transfers such that for every pair of types, none of the two types has an incentive to pretend to be the other type. Note that the property described in the second sentence differs from weak monotonicity by the order of quantifiers. In the second sentence, the quantifiers are ordered so that the property becomes more demanding. Surprisingly, there is a modified monotonicity property that is exactly equivalent to the more demanding property. We present this modified monotonicity in the next section.

5.4 CYCLICAL MONOTONICITY

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

$$\left(u(a^1,\theta^2) - u(a^1,\theta^1)\right) + \left(u(a^2,\theta^1) - u(a^2,\theta^2)\right) \leq 0. \tag{5.7}$$

This inequality considers on the left-hand side the sum of the changes to the agent's utility that occur in the following two thought experiments. First, we start with type θ^1 with implemented alternative a^1, and we switch the agent's type to θ^2, leaving the chosen alternative hypothetically unchanged as a^1. Next, we start with θ^2 with implemented alternative a^2, and we switch the agent's type to θ^1, leaving again the chosen alternative hypothetically unchanged. Both switches may cause an increase or a decrease in the agent'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 $\theta^1, \theta^2, \ldots, \theta^k$, and assume that the last element of the sequence is the same as the first: $\theta^1 = \theta^k$. We can consider the same sequence of switches as we considered in the previous paragraph and then sum up the changes to the agent's utility caused by these switches. "Cyclical monotonicity" requires the sum to be not positive.

Definition 5.5 *A decision rule q is "cyclically monotone" if for every sequence of length $k \in \mathbb{N}$ of types $(\theta^1, \theta^2, \ldots, \theta^k) \in \Theta^k$ with $\theta^k = \theta^1$, we have*

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

where for all $\kappa = 1, 2, \ldots, k$ we define: $a^\kappa \equiv q(\theta^\kappa)$.

Thus, weak monotonicity is the same as cyclical 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, a somewhat long (but, in terms of the mathematics used, completely elementary) proof shows that cyclical monotonicity is equivalent to implementability.

Proposition 5.2 *A decision rule is implementable if and only if it is cyclically monotone.*

Before we prove Proposition 5.2, let us verify that the counterexample from the previous section is not cyclically monotone. In that example, consider the following sequence of length 4: $(\theta^1, \theta^3, \theta^2, \theta^1)$. For this sequence, we obtain

$$\sum_{\kappa=1}^{3}(u(a^\kappa, \theta^{\kappa+1}) - u(a^\kappa, \theta^\kappa)) = 1 + 1 + 1 = 3. \tag{5.8}$$

Cyclical monotonicity is violated because this sum is positive rather than less than or equal to zero.

Proof

We first show that if q is implementable, then it is cyclically monotone. Incentive compatibility implies for every $\kappa = 1, 2, \ldots, k-1$ that type $\theta^{\kappa+1}$ has no incentive to pretend to be type κ:

$$\begin{aligned} u(a^\kappa, \theta^{\kappa+1}) - t(\theta^\kappa) \leq u(a^{\kappa+1}, \theta^{\kappa+1}) - t(\theta^{\kappa+1}) \Leftrightarrow \\ u(a^\kappa, \theta^{\kappa+1}) - u(a^{\kappa+1}, \theta^{\kappa+1}) \leq t(\theta^\kappa) - t(\theta^{\kappa+1}). \end{aligned} \tag{5.9}$$

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

$$\begin{aligned} \sum_{\kappa=1}^{k-1} \left(u(a^\kappa, \theta^{\kappa+1}) - u(a^{\kappa+1}, \theta^{\kappa+1})\right) \leq 0 \Leftrightarrow \\ \sum_{\kappa=1}^{k-1} u(a^\kappa, \theta^{\kappa+1}) - \sum_{\kappa=1}^{k-1} u(a^{\kappa+1}, \theta^{\kappa+1}) \leq 0. \end{aligned} \tag{5.10}$$

Because $\theta^k = \theta^1$, we can replace the second term in this inequality by the second term in the following inequality:

$$\begin{aligned} \sum_{\kappa=1}^{k-1} u(a^\kappa, \theta^{\kappa+1}) - \sum_{\kappa=1}^{k-1} u(a^\kappa, \theta^\kappa) \leq 0 \Leftrightarrow \\ \sum_{\kappa=1}^{k-1} \left(u(a^\kappa, \theta^{\kappa+1}) - u(a^\kappa, \theta^\kappa)\right) \leq 0, \end{aligned} \tag{5.11}$$

which is what we wanted to show.

Next, we shall show that cyclical monotonicity implies that the rule q is implementable. To construct the transfer payments, we shall need some further definitions. We fix an arbitrary type $\widetilde{\theta} \in \Theta$. Then, for every $\theta \in \Theta$ we define $\mathcal{S}(\theta)$ to be the set of all finite sequences $(\theta^1, \theta^2, \ldots, \theta^k)$ of elements of θ that satisfy $\theta^1 = \widetilde{\theta}$ and $\theta^k = \theta$. Here, k can be any element of $\mathbb{N}$ with $k \geq 2$. Then we define the function $V : \Theta \to \mathbb{R}$ by

$$V(\theta) \equiv \sup_{(\theta^1,\theta^2,\ldots,\theta^k)\in\mathcal{S}(\theta)} \sum_{\kappa=1}^{k-1} \left(u(a^\kappa, \theta^{\kappa+1}) - u(a^\kappa, \theta^\kappa)\right) \tag{5.12}$$

for all $\theta \in \Theta$, where a^κ is defined to be $q(\theta^\kappa)$ 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 whose supremum is the right-hand side of equation (5.12) is bounded from above. This is what we will show.

First we note that cyclical monotonicity means that for $\theta = \tilde{\theta}$ all sums on the right-hand side of (5.12) are nonpositive. Therefore, for $\theta = \widetilde{\theta}$, the supremum is well-defined. Indeed, because the trivial sequence $(\widetilde{\theta}, \widetilde{\theta})$ is contained in $\mathcal{S}(\widetilde{\theta})$, and because for this sequence the sum over which we take the supremum in the definition of V is zero, we can conclude that $V(\widetilde{\theta}) = 0$.

Now consider some $\theta \neq \tilde{\theta}$, fix a sequence $(\hat{\theta}^1, \hat{\theta}^2, \ldots, \hat{\theta}^k) \in \mathcal{S}(\theta)$, and define $a^\kappa = q(\hat{\theta}^\kappa)$ for $\kappa = 1, 2, \ldots, k-1$. Then

$$\begin{aligned} V(\widetilde{\theta}) &= \sup_{(\theta^1,\theta^2,\ldots,\theta^k)\in\mathcal{S}(\widetilde{\theta})} \sum_{\kappa=1}^{k-1} \left(u(a^\kappa, \theta^{\kappa+1}) - u(a^\kappa, \theta^\kappa)\right) \\ &\geq \sum_{\kappa=1}^{k-1} \left(u(a^\kappa, \hat{\theta}^{\kappa+1}) - u(a^\kappa, \hat{\theta}^\kappa)\right) + \left(u(a^k, \widetilde{\theta}) - u(a^k, \theta)\right). \end{aligned} \tag{5.13}$$

Here, the inequality follows from the fact that $\mathcal{S}(\widetilde{\theta})$ includes the set of all finite sequences that have θ as their penultimate element and that have $\widetilde{\theta}$ as their last element. The inequality that we have obtained is equivalent to

$$\sum_{\kappa=1}^{k-1} \left(u(a^\kappa, \hat{\theta}^{\kappa+1}) - u(a^\kappa, \hat{\theta}^\kappa)\right) \leq V(\widetilde{\theta}) + \left(u(a^k, \theta) - u(a^k, \widetilde{\theta})\right). \tag{5.14}$$

Using $V(\widetilde{\theta}) = 0$, which we showed earlier, we can infer

$$\sum_{\kappa=1}^{k-1} \left(u(a^\kappa, \theta^{\kappa+1}) - u(a^\kappa, \theta^\kappa)\right) \leq u(a^k, \theta) - u(a^k, \widetilde{\theta}), \tag{5.15}$$

which shows that the sums that appear on the right-hand side of (5.12) are bounded from above and therefore that V is well-defined.

We now use the function V to construct the payment scheme that make q incentive-compatible. Indeed, we shall construct the payment scheme so that the utility $u(q(\theta),\theta)$ is exactly equal to $V(\theta)$ for all $\theta \in \Theta$. That is, we set

$$\begin{aligned} u(q(\theta),\theta) - t(\theta) = V(\theta) \Leftrightarrow \\ t(\theta) = u(q(\theta),\theta)) - V(\theta). \end{aligned} \tag{5.16}$$

To prove that this implies that the mechanism is incentive-compatible, we need to show for all $\theta, \theta' \in \Theta$:

$$\begin{aligned} u(q(\theta),\theta) - t(\theta) \geq u(q(\theta'),\theta) - t(\theta') \Leftrightarrow \\ u(q(\theta),\theta) - (u(q(\theta),\theta)) - V(\theta)) \geq \\ u(q(\theta'),\theta) - \left(u(q(\theta'),\theta')) - V(\theta')\right) \Leftrightarrow \\ V(\theta) \geq V(\theta') + u(q(\theta'),\theta) - u(q(\theta'),\theta')). \end{aligned} \tag{5.17}$$

This is true by the definition of V because the set of all finite sequences that start with $\tilde{\theta}$ and end with θ includes the set of all finite sequences that start with $\tilde{\theta}$, have θ' as their penultimate element, and end with θ. □

5.5 CYCLICAL MONOTONICITY 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 $\Theta \subseteq \mathbb{R}^M$ consists of vectors $\theta = (\theta_1, \theta_2, \ldots, \theta_M)$ of Bernoulli utilities of the M outcomes. We assume that the set Θ is convex.

Consider a given direct mechanism, and define $U(\theta) \equiv q(\theta) \cdot \theta - t(\theta)$. Here, "$\cdot$" stands for the scalar vector product. Hence, U is the expected utility of the agent if all players report their types truthfully. Incentive compatibility is equivalent to the condition

$$U(\theta) = \max_{\theta' \in \Theta} \left(q(\theta') \cdot \theta - t(\theta')\right). \tag{5.18}$$

Here is one way of parsing the right-hand side of this equation: Consider the function that describes for each θ the expected utility that an agent of type θ obtains when pretending to be type θ'. For every θ' we can define such a function. On the right-hand side of the above equation we have for each θ the maximum of all the functions just described, where the maximum is taken over all θ'. Note that the function on the right-hand side is linear in θ for given θ'. Thus, we have on the right-hand side the maximum of a family of linear functions. Therefore, we can conclude that U is the maximum of functions that are linear in θ and therefore that it is convex in θ.

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, $q(\theta)$ is a subgradient of $U(\theta)$ (as a function of θ) at θ for every θ. This is because

$$\begin{aligned}
&U(\theta') \geq U(\theta) + q(\theta) \cdot (\theta' - \theta) \Leftrightarrow \\
&U(\theta') \geq q(\theta) \cdot \theta - t(\theta) + q(\theta) \cdot (\theta' - \theta) \Leftrightarrow \\
&U(\theta') \geq q(\theta) \cdot \theta' - t(\theta), \qquad (5.19)
\end{aligned}$$

which holds by incentive compatibility for every $\theta' \in \Theta$. 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 5.3 *Suppose that A is the set of all probability distributions over some finite set of outcomes. Suppose that Θ is a convex set of vectors of Bernoulli utility functions. Then a decision rule q is implementable if and only if there is a convex function $U : \theta \to \mathbb{R}$ such that $q(\theta)$ is a subgradient of U at θ.*

Proof

The argument preceding Proposition 5.3 has shown the necessity of the condition in Proposition 5.3. To see that it is also sufficient, define for every $\theta \in \Theta$ the agent's transfer $t(\theta)$ so that the function U of which q is the subgradient is exactly the agent's expected utility. The equivalencies that precede Proposition 5.3 establish incentive compatibility. □

Theorem 24.8 in Rockafellar (1970) says 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.

5.6 ONE-DIMENSIONAL TYPE SPACES

An important characteristic of the models of Chapters 2–4 is that agents' type spaces are subsets of $\mathbb{R}$ and are therefore one-dimensional. Implementability turns out to be much simpler to characterize for one-dimensional type spaces than for general type spaces. Weak monotonicity and cyclical monotonicity coincide for such type spaces, and therefore even weak monotonicity is equivalent to implementability. Moreover, for one-dimensional type spaces, weak monotonicity has the simple interpretation that allocations are increasing in type, which is the necessary and sufficient condition for implementability that we found in Chapters 2–4.

One-dimensionality is at first sight a purely mathematical concept, and it is not clear why the mathematical dimension of the type space should have any economic meaning. Indeed, it does not. In this section we shall be careful to give a definition of one-dimensionality of the type space that refers to alternatives and preferences rather than to the mathematical properties of the type space. This allows us to focus on the economic meaning of one-dimensionality. The type space itself will be an abstract set rather than a set of real numbers or vectors of real numbers.

We begin by describing what "higher alternatives" are. Let R be a complete and transitive order[2] of A. The strict order derived from R is denoted by P: $aPb \Leftrightarrow [aRb \text{ and not } bRa]$. The indifference relation derived from R is denoted by I: $aIb \Leftrightarrow [aRb \text{ and } bRa]$. It is important not to mistake R for the agent's preference relation. That preference relation is over pairs of alternatives and payments, not just over alternatives.

Given an ordering of alternatives, we can now order types.

Definition 5.6 *Let R be a complete and transitive order of A. For any pair of types $\theta, \theta' \in \Theta$, we shall say that $\theta \succ_R \theta'$ ("θ is a higher type than θ' relative to R") if*

$$u(a,\theta) - u(a',\theta) > u(a,\theta') - u(a',\theta') \qquad \textit{for all } a, a' \in A \textit{ with } aPa'$$

and

$$u(a,\theta) - u(a',\theta) = u(a,\theta') - u(a',\theta') = 0 \qquad \textit{for all } a, a' \in A \textit{ with } aIa'.$$

Intuitively, $\theta \succ_R \theta'$ means that θ attaches larger marginal value to higher alternatives than θ' for any two ordered alternatives that we compare. In other words, θ unambiguously has a stronger preference for higher alternatives than θ'. What do we mean here by "higher alternatives"? This is defined by the ordering R on A. Thus, the order $\succ_R$ is conditional on R. It is important that not every pair of types, θ and θ', is ordered by $\succ_R$. In other words, $\succ_R$ is often incomplete.

We are now going to use the orders R and $\succ_R$ to define a notion of monotonicity of decision rules that has the simple interpretation of higher allocations being assigned to higher types. This concept is analogous to the naive notion of monotonicity of decision rules that we used in Chapters 2–4 except that we allow the ordering of types to be incomplete.

Definition 5.7 *Let R be a complete and transitive order of A. A decision rule q is called "monotone with respect to R" if*

$$\theta \succ_R \theta' \Rightarrow q(\theta) R q(\theta').$$

In words, this definition says that a decision rule is "monotone with respect to R" if an increase in the 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 5.4 *Let R be a complete and transitive order of A. If a decision rule q is weakly monotone, then it is monotone with respect to R.*

A remarkable feature of Proposition 5.4 is that it is true for *any* order R of A, as long as we consider the ordering of types $\succ_R$ that corresponds to R.

Proof

Suppose q is weakly monotone. Suppose $\theta \succ_R \theta'$. Define $a \equiv q(\theta)$ and $a' \equiv q(\theta')$. By the definition of weak monotonicity, we must have:
$u(a,\theta) - u(a',\theta) \geq u(a,\theta') - u(a',\theta')$. We now prove indirectly that $\theta \succ_R \theta'$ implies aRa'. Suppose the contrapositive: $a'Pa$. Then $\theta \succ_R \theta'$ implies

$$u(a',\theta) - u(a,\theta) > u(a',\theta') - u(a,\theta') \Leftrightarrow$$
$$u(a,\theta) - u(a',\theta) < u(a,\theta') - u(a',\theta'), \tag{5.20}$$

which contradicts weak monotonicity. □

The reason that we can't prove in general the converse of this result is that monotonicity with respect to R imposes a restriction on collective decisions only for types that are comparable in the order $\succ_R$, and in general this order is incomplete. By contrast, weak monotonicity places restrictions on choices even if types are not ordered in $\succ_R$. But we can strengthen Proposition 5.4 so that it becomes an equivalence if we restrict attention to type spaces that are completely ordered by $\succ_R$. It is such type spaces that we shall call "one-dimensional."

Definition 5.8 *Let R be a complete and transitive order of A. The type space Θ is "one-dimensional with respect to R" if for any $\theta, \theta' \in \Theta$ with $\theta \neq \theta'$ we have either $\theta \succ_R \theta'$ or $\theta' \succ_R \theta$ or both.*

Note that if a type space is in this sense one-dimensional, we can assign to any type a real number, namely the difference in utility for some alternatives, say the difference $u(a',\theta) - u(a,\theta)$ where $a'Pa$, 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. Thus, there is a sense in which we can very naturally embed one-dimensional type spaces into the one-dimensional Euclidean space $\mathbb{R}$.

We can now show that for completely ordered (i.e., one-dimensional) type spaces, monotonicity is not only necessary, as shown by Proposition 5.4, but also sufficient for weak monotonicity.

Proposition 5.5 *Let R be complete and transitive order of A such that Θ is one-dimensional with respect to R. Then a decision rule q is weakly monotone if and only if it is monotone with respect to R.*

Proof

In light of Proposition 5.4, we only have to show that monotonicity implies weak monotonicity. So suppose q is monotone, and consider $\theta \in \Theta$ and $\theta' \in \Theta$ such that $q(\theta) = a$ but $q(\theta') = a'$. We may restrict our attention to the case $a \neq a'$. Because Θ is one-dimensional, we must have $\theta \succ_R \theta'$ or $\theta' \succ_R \theta$. Without loss of generality, assume the former. Then, because q is monotone, we have aRa'. Moreover, because $\theta \succ_R \theta'$ we obtain

$$u(a,\theta) - u(a',\theta) > u(a,\theta') - u(a',\theta') \tag{5.21}$$

if aPa', or

$$u(a,\theta) - u(a',\theta) = u(a,\theta') - u(a',\theta') \tag{5.22}$$

if aIa'. This shows that q is weakly monotone. □

Having established that weak monotonicity and monotonicity are equivalent on one-dimensional type spaces, we next show that monotonicity is sufficient for a decision rule to be implementable by appropriate transfer schemes. This, together with Proposition 5.1, then implies that on one-dimensional domains, monotonicity is necessary and sufficient for implementability and is therefore also equivalent to cyclical monotonicity.

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 space is bounded in the sense of the following definition:

Definition 5.9 *The type space Θ is "bounded" if there is a constant $c > 0$ such that for all $a, a' \in A$ and all $\theta \in \Theta$ we have*

$$-c < u(a',\theta) - u(a,\theta) < c. \tag{5.23}$$

Intuitively, the type space is "bounded" if there is a uniform upper bound for the agent's willingness to pay for a change in the decision.

Proposition 5.6 *Suppose that A is finite. Let R be a complete and transitive order of A, and suppose that the type space Θ is bounded and one-dimensional with respect to R. Let q be a decision rule that is monotone with respect to R. Then q is implementable.*

Proof

We denote the range of q over Θ—that is, the set $\{q(\theta) \mid \theta \in \Theta\}$—by $\{a^1, a^2, \ldots, a^n\}$, where $a^n R a^{n-1} R \ldots R a^1$. For simplicity we assume that all these preference relations are strict: $a^n P a^{n-1} P \ldots P a^1$. 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 $t(\theta) = 0$ for all θ. This will obviously make it optimal for the agent to report her type truthfully because the same outcome will occur regardless of the agent's report.

If $n \geq 2$, define for every $k = 1, 2, \ldots, n$ the set

$$\Theta^k \equiv \{\theta \in \Theta \mid q(\theta) = a^k\}. \tag{5.24}$$

Monotonicity and one-dimensionality imply that the sets Θ^k are ordered in the following sense:

$$k' > k, \theta^k \in \Theta^k, \theta^{k'} \in \Theta^{k'} \Rightarrow \theta^{k'} \succ_R \theta^k. \tag{5.25}$$

For every $k = 2, \ldots, n$, define

$$\tau^k \equiv \inf\{u(a^k, \theta) - u(a^{k-1}, \theta) \mid \theta \in \Theta^k\}. \tag{5.26}$$

The infimum here is well-defined because Θ is bounded. Note that the ordering of types implies

$$\begin{aligned} &k' < k, \theta' \in \Theta^{k'} \Rightarrow u(a^k, \theta') - u(a^{k-1}, \theta') \leq \tau^k \\ &k' > k, \theta' \in \Theta^{k'} \Rightarrow u(a^k, \theta') - u(a^{k-1}, \theta') \geq \tau^k. \end{aligned} \tag{5.27}$$

We define the agent's transfer payment as follows

$$t(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta^1; \\ \sum_{\kappa=2}^{k} \tau^{\kappa} & \text{if } \theta \in \Theta^k \text{ where } k \geq 2. \end{cases} \tag{5.28}$$

We verify that this transfer scheme makes truthful reporting of θ optimal. If the agent's true type is $\theta \in \Theta^k$ and if she reports any type in Θ^k, her utility will be independent of her report. If she reports a type in $\Theta^{k'}$ where $k' > k$, then the change in her utility in comparison to truthful reporting will be

$$\begin{aligned} & u(a^{k'},\theta) - u(a^k,\theta) - \sum_{\kappa=k+1}^{k'} \tau^\kappa \\ &= \sum_{\kappa=k+1}^{k'} \left(u(a^\kappa,\theta) - u(a^{\kappa-1},\theta)\right) - \sum_{\kappa=k+1}^{k'} \tau^\kappa \\ &\leq \sum_{\kappa=k+1}^{k'} \left(u(a^\kappa,\theta) - u(a^{\kappa-1},\theta)\right) - \sum_{\kappa=k+1}^{k'} \left(u(a^\kappa,\theta) - u(a^{\kappa-1},\theta)\right) = 0. \end{aligned} \tag{5.29}$$

The first equality is a simple rewriting. The inequality follows from the first inequality in (5.27). A symmetric argument proves that there is no incentive for the agent to report a type in $\Theta^{k'}$ where $k' < k$. □

Proposition 5.6 is analogous to the result that we obtained in earlier chapters—that is, that monotonicity of q is necessary and sufficient for implementability of q. In all those models, the domain of preferences satisfied a one-dimensionality condition (although we allowed infinite outcome sets). One-dimensional models are very frequently used in applied work.

5.7 RICH TYPE SPACES

The one-dimensional domains described in Section 5.6 are not the only domains for 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 of types of the agent is "rich" if there is some reflexive and transitive, but possibly incomplete, binary relation R on A such that all utility functions that represent R are possible utility functions of the agent. Formally:

Definition 5.10 *The type space Θ is "rich" if there is a possibly incomplete preference relation R on A such that for every $u : A \to \mathbb{R}$ that represents R—that is, that satisfies $aRb \Rightarrow u(a) \geq u(b)$—there is a $\theta \in \Theta$ such that*

$$u(a,\theta) = u(a) \qquad \textit{for all } a \in A.$$

Note that this condition becomes more restrictive as R becomes less complete—that is, as comparisons are dropped from R. The reason is that the less complete R is, the more utility functions may represent R.

Bikhchandani et al.'s theorem 1 is as follows:

Proposition 5.7 *Suppose that A is finite and that the type space Θ is rich. Let q be a weakly monotone decision rule. Then q is implementable.*

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.

5.8 REVENUE EQUIVALENCE

So far, we have focused on implementable decision rules q. Now we add to this a result about the transfer schemes t that correspond to implementable decision rules—that is, the transfer schemes for which (q, t) is incentive-compatible. Our objective is to establish a revenue equivalence result analogous to Lemma 2.4 in Chapter 2, which says that for given q the function t is uniquely determined up to a constant. We already know that for finite type spaces the revenue equivalence theorem does not need to be true.[3] Therefore, it is clear that in the abstract setting of this chapter we need to make an additional assumption regarding the type space Θ if we want to obtain a revenue equivalence result. We shall assume that Θ is a convex subset of some finite-dimensional Euclidean space.

An assumption for the type spaces alone is, however, not enough. We shall also assume that the utility function is a convex function of the agent's type. This assumption will ensure that an agent's utility under an incentive-compatible mechanism depends sufficiently smoothly on their type so that we can employ a version of the envelope theorem. Convexity of the utility function 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 an agent is the vector of this agent'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 5.8 *Suppose that Θ is a convex subset of a finite-dimensional Euclidean space and that $u(a, \theta)$ is a convex function of θ. Suppose that (q, t) is an incentive-compatible direct mechanism. Then a direct mechanism with the same decision rule, (q, t'), is incentive-compatible if and only if there is a $\tau \in \mathbb{R}$ such that*

$$t'(\theta) = t(\theta) + \tau \qquad \text{for all } \theta \in \Theta.$$

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. 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 property of the functions involved needs to be established. We omit the proof.

5.9 INDIVIDUAL RATIONALITY

In the examples that we have seen in Chapters 2–4, individual rationality required that agents who report their types truthfully 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, say, the alternative in A that corresponded to individual rationality was "not obtaining the good." In the bilateral trade example, it was "no trade." To generalize, we shall assume that there is some alternative $a \in A$ such that individual rationality means that the agent cannot be made worse off than he would be if alternative a were chosen and the agent had to pay nothing.

Definition 5.11 *Let $a \in A$. A direct mechanism is "individually rational with outside option a" if for all $\theta \in \Theta$ we have*

$$u(q(\theta), \theta) - t(\theta) \geq u(a, \theta).$$

We obtain a useful characterization of individual rationality if we focus on the case of one-dimensional type spaces. In this case, it is sufficient for the individual rationality constraint to be satisfied that it is holds for the lowest type. We have seen special cases of this result in earlier chapters.

Proposition 5.9 *Let R be a complete and transitive order of A and suppose that the type set Θ is one-dimensional with respect to R. Assume that there is a type $\underline{\theta}$ that is the lowest type in the order $\succ_R$ in Θ—that is, $\theta \succ_R \underline{\theta}$ for all $\theta \in \Theta$ such that $\theta \neq \underline{\theta}$. Finally, assume that $\underline{a}$ is the worst alternative in A in the order R—that is, $bR\underline{a}$ for every $b \in A$ with $b \neq \underline{a}$. Then an incentive-compatible direct mechanism is individually rational with outside option $\underline{a}$ if and only if*

$$u(q(\underline{\theta}), \underline{\theta}) - t(\underline{\theta}) \geq u(\underline{a}, \underline{\theta}).$$

Proof

The condition in Proposition 5.9 is the individual rationality constraint for type $\underline{\theta}$. It is therefore necessary. To show that it is also sufficient, we show that it implies individual rationality for any type $\theta \neq \underline{\theta}$. We first note that

$$
\begin{aligned}
&u(q(\underline{\theta}),\underline{\theta}) - t(\underline{\theta},\theta) \geq u(\underline{a},\underline{\theta}) \Leftrightarrow \\
&u(q(\underline{\theta}),\underline{\theta}) - u(\underline{a},\underline{\theta}) \geq t(\underline{\theta}).
\end{aligned} \tag{5.30}
$$

Now if $\theta \neq \underline{\theta}$, then by the one-dimensionality condition we have $\theta \succ_R \underline{\theta}$. Because, moreover, the alternative $q(\underline{\theta})$ satisfies either $q(\underline{\theta})P\underline{a}$ or $q(\underline{\theta}) = \underline{a}$ we can infer

$$
\begin{aligned}
&u(q(\underline{\theta}),\theta) - u(\underline{a},\theta) \geq t(\underline{\theta}) \Leftrightarrow \\
&\quad u(q(\underline{\theta}),\theta) - t(\underline{\theta}) \geq u(\underline{a},\theta).
\end{aligned} \tag{5.31}
$$

By incentive compatibility we have

$$
u(q(\theta),\theta) - t(\theta) \geq u(q(\underline{\theta}),\theta) - t(\underline{\theta}). \tag{5.32}
$$

Combining the last two inequalities, we obtain

$$
u(q(\theta),\theta) - t(\theta) \geq u(\underline{a},\theta), \tag{5.33}
$$

which is individual rationality with outside option $\underline{a}$ for type θ. □

5.10 REMARKS ON THE LITERATURE

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 (1987). The result on the sufficiency of weak monotonicity is due to Bikhchandani et al. (2006). The revenue equivalence result that we presented is taken from Krishna and Maenner (2001).

5.11 PROBLEMS

(a) Suppose that we modify the counterexample at the end of Section 5.3 as follows (where $x \in \mathbb{R}$):

	θ^1	θ^2	θ^3
a	0	-1	x
b	1	0	-1
c	-1	1	0

For which values of $x \in \mathbb{R}$ is the decision rule that we defined in Section 5.3 cyclically monotone?

(b) For the values of x that you found in problem (a), construct the function V to which the proof of Proposition 5.2 refers.

(c) Prove that the "inf" in the definition of the function V in Proposition 5.2 can be replaced by a "min" if the set of alternatives is finite.

(d) Suppose the set of alternatives is $A = \{1, 2, 3\} \times \{1, 2, 3\}$. Denote the typical element of A by (a, b). Define $\Theta = [\underline{\theta}_a, \bar{\theta}_a] \times [\underline{\theta}_b, \bar{\theta}_b]$, where $0 < \underline{\theta}_a < \bar{\theta}_a$ and $0 < \underline{\theta}_b < \bar{\theta}_b$. Preferences for a type $(\theta_a, \theta_b) \in \Theta$ are given by $u((a, b), (\theta_a, \theta_b)) = \theta_a a + \theta_b b$. For which values of $\underline{\theta}_a$, $\bar{\theta}_a$, $\underline{\theta}_b$, and $\bar{\theta}_b$ is this type space one-dimensional?

(e) Show that the type space in Section 2.3 is one-dimensional.

6

BAYESIAN MECHANISM DESIGN

6.1 INTRODUCTION

To generalize the examples of Bayesian mechanism design that we considered in Chapter 3, we need to take two steps. First, we need to consider more general sets of alternatives and more general specifications of types and preferences. Second, we need to consider more general distributions of types than in Chapter 3, allowing, unlike we did in that chapter, that different agents' types are not stochastically independent. This chapter will correspondingly have two parts. In the first part we deal with more general sets of alternatives, types, and preferences. In this part our results will be straightforward applications of results from Chapter 5.

The second part will deal with not independent—or, in somewhat sloppy language, "correlated"—types. In practice, it seems plausible that types are often not independent. For example, if one agent values an object sold in an auction particularly highly, then this agent might think that it is likely that other agents also value the object highly. This motivates the study of a model of Bayesian mechanism design in which types are not independent.

We shall explain why the results of the previous chapter cannot be used in this case. The nature of the analysis in the second part will then be quite different from what we have seen before. The results in the second part of this chapter will indicate that with dependent types almost all combinations of decision rules and expected 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, we shall begin a discussion of what this missing ingredient might be. This discussion will be continued in Chapter 10.

6.2 SETUP

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 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 $\Theta \equiv \Theta_1 \times \Theta_2 \times \ldots \times \Theta_N$. 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, 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 mechanism designer. We write μ_i for the marginal probability distribution of 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 is also the mechanism designer's belief. We shall explore the role of the common prior assumption in some detail in Chapter 10.

An important special case is that types are stochastically independent. This special case will be treated in the next section. Here, we define what we mean by independent types.

Definition 6.1 *Types are "independent" if for every $i \in I$ we have*

$$\mu(\cdot \mid \theta_i) = \mu_i \qquad \text{for all } \theta_i \in \Theta_i.$$

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 6.2 *A "direct mechanism" $(q, t_1, t_2, \ldots, t_N)$ consists of a mapping*

$$q : \Theta \to A,$$

which 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."
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 6.3 *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*

$$\int_{\Theta_{-i}} u_i(q(\theta_i, \theta_{-i}), \theta_i) - t_i(\theta_i, \theta_{-i})\, d\mu(\theta_{-i}|\theta_i) \geq$$

$$\int_{\Theta_{-i}} u_i(q(\theta_i', \theta_{-i}), \theta_i) - t_i(\theta_i', \theta_{-i})\, d\mu(\theta_{-i}|\theta_i).$$

Our first concern will be with the characterization of Bayesian incentive-compatible decision rules. We shall later bring in additional considerations of individual rationality and ex post budget balance. As in Definition 5.11, we define individual rationality with respect to an outside option. The outside option here is a probability distribution over A. We denote the set of all such probability distributions by $\Delta(A)$. We allow different agents' outside options to differ.

Definition 6.4 *A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is "individually rational with outside options $\alpha_i \in \Delta(A)$ for $i \in I$" if for every agent i and every type $\theta_i \in \Theta_i$ we have*

$$\int_{\Theta_{-i}} u_i(q(\theta_i, \theta_{-i}), \theta_i) - t_i(\theta_i, \theta_{-i})\, d\mu(\theta_{-i}|\theta_i) \geq \int_A u_i(a, \theta_i)\, d\alpha_i.$$

Definition 6.5 *A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ is "ex post budget balanced" if for every type vector $\theta \in \Theta$ we have*

$$\sum_{i \in I} t_i(\theta) = 0.$$

6.3 INDEPENDENT TYPES

To obtain characterizations of *all* incentive-compatible decision rules, we can translate the results of the previous chapter into our current setting. These characterizations apply at the interim level. To translate these results, we treat the set of all probability distributions over A as the set of alternatives. 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

$$Q_i(\mathcal{A} \mid \theta_i) \equiv \int_{\Theta_{-i}} I_{\mathcal{A}} \, d\mu(\theta_{-i} \mid \theta_i), \tag{6.1}$$

where $I_{\mathcal{A}} : \Theta_{-i} \to \{0, 1\}$ 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:

$$T_i(\theta_i) \equiv \int_{\Theta_{-i}} t_i(\theta_i, \theta_{-i}) \, d\mu(\theta_{-i} \mid \theta_i). \tag{6.2}$$

Our results in Chapter 5 can now all be adapted to the current setting, if we take Q_i to be the decision rule and T_i to be the transfer rule. As an example we adapt Proposition 5.2 to our setting.

Proposition 6.1 *Suppose types are independent. Then 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 $i \in I$ and 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*

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

No proof is needed. But, it is very important to understand the role of the assumption that types are independent. It is only this assumption that allows us to adapt the results of Chapter 5 to the Bayesian setting in a straightforward manner. The assumption is used to ensure that the probability distribution over alternatives, $Q_i(\theta_i)$, and the expected payment $T_i(\theta_i)$, are the same for type θ_i and for some other type θ_i' pretending to be type θ_i. In Chapter 5 we take this, rightly, for granted, because the probability distributions over alternatives and payments are chosen by the mechanism designer. But in this chapter, they also depend on the beliefs that a type holds about other agents' types, because they are calculated at the interim level. We can use the results from the previous chapter in a straightforward manner in the Bayesian setup only if the beliefs of a player i's types about the other players' types θ_{-i} are the same whatever type player i is. This is precisely what the assumption that types are independent ensures.

In Chapter 5, we considered the special case in which the type space is one-dimensional. In this special case, monotonicity of the decision rule q was shown to be necessary and sufficient for the existence of an incentive-compatible direct 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 one-dimensional where the set of alternatives is the set of all probability distributions over A. The only cases where it is easy to show that this condition is met seem to be those in which the set A consists of only two alternatives, or in which, from agent i's perspective, regardless of i's type, all alternatives in A can be divided into two indifference sets. What do I mean by this? Consider, for example, the single unit auction example. The set of alternatives consists of either giving the good to some agent i or not selling it at all. But for any given agent i, all that matters is whether i gets the good or not. All types of agent i, in the model that we showed, are indifferent between all alternatives that do not give the good i to them. The other two examples in Chapter 3 were examples in which there were literally only two alternatives.

We continue 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 5.8 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 this text.

Proposition 6.2 *Suppose that types are independent. Suppose also 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 every $i \in I$ *and every* $\theta_i \in \Theta_i$. *Then for every* $i \in I$ *there is a number* $\tau_i \in \mathbb{R}$ *such that*

$$T_i'(\theta_i) = T_i(\theta_i) + \tau_i$$

for every $i \in I$ *and for every* $\theta_i \in \Theta_i$.

Turning to individual rationality, the result on individual rationality with one-dimensional type spaces, Proposition 5.9, directly applies, if the conditions of that result are met for an agent *i*. Concerning budget balance, we shall provide below, in Proposition 7.10, a sufficient condition for the existence of an efficient, ex post budget balanced mechanism that is dominant strategy incentive-compatible. This condition is then, of course, also sufficient condition for the existence of an efficient, ex post budget balanced mechanism that is Bayesian incentive-compatible.

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 6.2 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 6.6 *A direct mechanism* $(q, t_1, t_2, \ldots, t_N)$ *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 *Suppose types are independent. Then for every direct mechanism that is ex ante budget balanced, there is an equivalent direct mechanism that is ex post budget balanced.*

Here, we use the same definition of "equivalent" that we used for Proposition 3.6: Two mechanisms are "equivalent" if they 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', are 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 different agents' types. Throughout this section, we shall assume that each of the sets Θ_i is finite. The results are more easily proved with finite type spaces, although they are also true with infinite type spaces. Note that this is in contrast with some earlier results which were more easily shown for continuous type spaces than for discrete type spaces. Although the finiteness assumption is, 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 assume that every $\theta \in \Theta$ has positive probability: $\mu(\theta) > 0$. This, too, simplifies the exposition.

Observe that the definitions of Section 6.2 were not restricted to the case of independent types, and therefore we can use them here. The revelation principle holds, and 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 also 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 decision 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 decision rule. This then easily implies, for example, that the expected payments are determined up to a constant by the decision rule.

How did we show that the derivative of interim expected utility with respect to an agent's type depends only on the decision rule? The argument for this is as follows. By the envelope theorem, as we change agent i's type, we may take agent i's type report as given and fixed. 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 reasons for this. The first is that all types have the same utility function of money, that is, $u(t_i) = t_i$. 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 that 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 decision 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 implementable decision rules. We need to provide a joint characterization of incentive-compatible decision 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 6.7 *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*

$$\mu(\theta_{-i} \mid \theta_i) = \sum_{\theta_i' \in \Theta_i \setminus \{\theta_i\}} \lambda(\theta_i')\mu(\theta_{-i} \mid \theta_i') \qquad \textit{for all } \theta_{-i} \in \Theta_{-i}.$$

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[1]:

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, that is, $\pi : \Omega \to [0, 1]$, 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

$$\sum_{\omega \in \Omega} \pi(\omega) = 1. \tag{6.4}$$

The 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 $\omega \in \Omega$ we have

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \pi(\omega)} = -\frac{\hat{\pi}(\omega)}{\pi(\omega)} + \lambda = 0 \Leftrightarrow \\ \frac{\hat{\pi}(\omega)}{\pi(\omega)} = \lambda. \end{aligned} \tag{6.6}$$

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 has this 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; that is, 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's Alternative[2]:

Proposition 6.5 *Let A be an $n \times m$ matrix and let $b \in \mathbb{R}^n$. Either*

$Ax = b$ *has a solution* $x \geq 0$

or (exclusive)

$y^T A \geq 0, \quad y^T b < 0$ *has a solution* y.

To apply Farkas's Alternative, we take some type θ_i of agent i as given and fixed. We take b to be 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's alternative is not satisfied. We deduce that the second condition holds. The column vector y, which has $\#\Theta_{-i}$ 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 nonnegative. 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, namely, 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.

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 first entry in each box shows the payment by player 1, and the second entry shows the payment by player 2.

	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.

	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.

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 6.4 for this example, and we 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 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.

	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.

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 now combine Bayesian incentive compatibility with individual rationality and ex post budget balance. We begin by considering individual rationality. In the transformation that was described in Section 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 that 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 be 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, if combined with the Crémer McLean condition, guarantees that budget balance can always be achieved. The additional condition is called "identifiability."

Definition 6.8 *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 the following:

Proposition 6.6 *Suppose that the distribution μ satisfies the Crémer–McLean and the identifiability conditions. Consider any direct mechanism $(q, t_1, t_2, \ldots, t_N)$ that is ex ante budget balanced. Then there is an equivalent Bayesian incentive-compatible and ex post budget balanced mechanism $(q, t'_1, t'_2, \ldots, t'_N)$.*

We use the phrase "equivalent direct mechanism" here with the same meaning as, for example, in Proposition 6.3. 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 of how the setup 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[3]; that is, 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 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.[4]

6.5 REMARKS ON THE LITERATURE

Proposition 6.4 is adapted from Crémer and McLean (1988). My favorite reference for Farkas's 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).

6.6 PROBLEMS

(a) A seller has a single indivisible object to sell. The seller values the object at zero, and he seeks to maximize his expected revenue. There are two potential buyers: $i = 1, 2$. Buyer i's von Neumann–Morgenstern utility equals $\theta_i - t_i$ 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 θ_i is a random variable that only buyer i observes, but that the other buyer and the seller don't observe. The random variable θ_1 takes only two possible values: $\theta_1 \in \{3, 4\}$. The random variable θ_2 takes only two possible values: $\theta_2 \in \{1, 2\}$. Note that the two buyers are *not* symmetric. The joint distribution of θ_1 and θ_2 is given by the Figure 6.6. Here, ε is a constant that satisfies $0 < \varepsilon < 0.25$. Each entry indicates the probability with which θ_1 takes the value indicated in the row, and at the same time θ_2 takes the value indicated in the column. The distribution indicated in Figure 6.6 is common knowledge among the buyers and the seller.

(i) Verify that the Crémer–McLean condition holds for buyer 1.

(ii) 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.

	$\theta_2 = 1$	$\theta_2 = 2$
$\theta_1 = 3$	$0.25 + \varepsilon$	$0.25 - \varepsilon$
$\theta_1 = 4$	$0.25 - \varepsilon$	$0.25 + \varepsilon$

Figure 6.6 Distribution of Valuations.

(iii) Use the payment scheme that you found in part (ii) 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.

(iv) Investigate the limit for $\varepsilon \rightarrow 0$ of the transfer payments in the selling mechanism that you obtained in part (iii).

7

DOMINANT STRATEGY MECHANISMS

7.1 INTRODUCTION

We next consider dominant strategy mechanisms in a more general setup than in Chapter 4. We shall in particular present results about dominant strategy mechanisms that achieve welfare maximization, namely the "Vickrey–Clarke–Groves" mechanisms. Of course, every dominant strategy mechanism is also Bayesian incentive-compatible for any prior over the type space. Therefore, the results of this chapter also have bearing on Bayesian incentive compatibility.

7.2 DOMINANT STRATEGY INCENTIVE COMPATIBILITY

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. If alternative a is chosen and agent i pays transfer t_i, agent i's utility 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 setup does include the case in which A is the set of lotteries over some set of nonrandom 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 7.1 *A direct mechanism $(q, t_1, t_2, \ldots, t_N)$ consists of a mapping*

$$q : \Theta \to A,$$

which 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." Note that we restrict ourselves to mechanisms where the payment is deterministic. Because we assume that agents are risk-neutral, 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 7.2 *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 begin by applying some of the results of Chapter 5 in a straightforward manner to dominant strategy mechanisms. We can adapt Rochet's theorem, Proposition 5.2 from Chapter 5, to the setting of the current chapter to obtain the following characterization of decision rules that are part of a dominant strategy incentive-compatible direct mechanism.

Proposition 7.1 *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 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* $a^\kappa \equiv q(\theta_i^\kappa, \theta_{-i})$.

We can also adapt Proposition 5.6 on incentive compatibility with one-dimensional type spaces to our setting here.

Proposition 7.2 *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 one-dimensional with respect to* R_i. *Let q be a decision rule. 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 if and only if for every* $i \in I$ *and* $\theta_{-i} \in \Theta_{-i}$ *the decision rule q is monotone in* θ_i *with respect to* R_i.

Finally, we can adapt Proposition 5.8 in Chapter 5 due to Krishna and Maenner (2001) to obtain the following result on revenue equivalence.

Proposition 7.3 *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 dominant strategy 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*

$$t_i'(\theta) = t_i(\theta) + \tau_i(\theta_{-i}) \qquad \textit{for all } \theta \in \Theta.$$

7.3 IMPLEMENTING EFFICIENT DECISION RULES

Often, we are interested in decision rules that maximize the utilitarian welfare function. Examples are the public goods model and the bilateral trade model in Chapter 4. We shall call decision rules that maximize utilitarian welfare "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 7.3 *An allocation rule* q^* *is called "efficient" if for every* $\theta \in \Theta$ *we have*

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

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 Chapter 4.

Definition 7.4 *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

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

The logic of the payments in a VCG mechanism is extraordinarily simple. In a VCG mechanism each agent i is paid the sum of the other agents' utility from the implemented alternative where utilities are calculated using the agents' reported types. This is the first term in the formula in Definition 7.4. This term aligns agent i's interests with utilitarian welfare. The second term is a constant that depends on the other agents' reported types and that does not affect agent i's incentives. This constant can be used to raise the overall revenue from the mechanism. The consequence of the payment formula is the following proposition. The proof of the Proposition essentially repeats the considerations that we have just mentioned.

Proposition 7.4 *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

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

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: $\theta_i' = \theta_i$. □

There is another way of looking at the formula in Definition 7.4 that is insightful. Consider the difference in the payment that agent i receives when reporting his true type, θ_i, and some other type θ_i'. Obviously, the second term in the formula drops out when we take that difference, and we obtain

$$-\left(\sum_{j\neq i} u_j(q(\theta),\theta_j) - \sum_{j\neq i} u_j(q(\theta_i',\theta_{-i}),\theta_j)\right). \tag{7.1}$$

This expression is the difference between (a) the sum of the other agents' utilities if agent i reports his type truthfully and (b) the sum of the other agents' utilities if agent i reports type θ_i' and thus potentially changes the alternative that is implemented. Thus, if agent i lies about his type, his payment is reduced by the externality that he imposes on the other agents.

Proposition 7.3 and its proof show that efficient decision rules can be implemented by a dominant strategy incentive-compatible mechanism. Instead of providing explicit mechanisms that do the job, we could have applied Proposition 7.1. In problem (a) for this chapter, you are asked to provide a proof along that route.

We can combine this result with the payoff equivalence result in Proposition 7.3 to obtain conditions under which VCG mechanisms are the only mechanisms that make efficient decision rules dominant strategy incentive-compatible.

Corollary 7.1 *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, and suppose that q is efficient. Then $(q, t_1, t_2, \ldots t_N)$ is a VCG mechanism.*

Proof

By Proposition 7.3, 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. □

7.4 POSITIVE ASSOCIATION OF DIFFERENCES

We now consider more general, but not necessarily efficient, decision rules that can be implemented in dominant strategies. We show that a surprisingly weak monotonicity condition, "positive association of differences" (PAD), together with conditions on the domain and range of the decision rule, is necessary and sufficient for dominant strategy implementability.

Definition 7.5 *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 all other alternatives are larger for all agents.

Proposition 7.5 *If q is weakly monotone in every θ_i, 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

$$u_i(a, \theta_i') - u_i(b, \theta_i') > u_i(a, \theta_i) - u_i(b, \theta_i) \Leftrightarrow$$
$$u_i(b, \theta_i') - u_i(a, \theta_i') < u_i(b, \theta_i) - u_i(a, \theta_i).$$

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 7.6 *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 7.6 *Suppose A is finite, and suppose for every $i \in I$ and $v \in \mathbb{R}^{\#A}$ there is a $\theta_i \in \Theta_i$ such that $(u_i(a, \theta_i))_{a \in A} = v$. 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 7.6. Although we shall not provide a proof of this characterization, we mention it and show how it implies Proposition 7.6.

Proposition 7.7 *Suppose A is finite, and suppose for every $i \in I$ and $v \in \mathbb{R}^{\#A}$ there is a $\theta_i \in \Theta_i$ such that $(u_i(a,\theta_i))_{a \in A} = v$. 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 $\theta \in \Theta$ we have*

$$\sum_{i=1}^{N} k_i u_i(q(\theta),\theta_i) + F(q(\theta)) \geq \sum_{i=1}^{N} k_i u_i(a,\theta_i) + F(a)$$

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 could set $F(\bar{a}) = z > 0$ and $F(a) = 0$ for all $a \in A$ with $a \neq \bar{a}$. 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 7.7 implies Proposition 7.6. It is obvious from Propositions 5.1 and 7.5 that PAD is necessary for implementability. All that is needed to derive Proposition 7.6 from Proposition 7.7 is therefore to show the following:

Proposition 7.8 *Suppose q satisfies the characterization in Proposition 7.7. 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:

$$t_i(\theta) = -\frac{1}{k_i}\left(\sum_{j \neq i} k_j u_j(q(\theta),\theta_j) + F(q(\theta))\right). \tag{7.2}$$

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

$$u_i(q(\theta_i', \theta_{-i}), \theta_i) + \frac{1}{k_i}\left(\sum_{j \neq i} k_j u_j(q(\theta_i', \theta_{-i}), \theta_j) + F(q(\theta_i', \theta_{-i}))\right). \tag{7.3}$$

Maximizing this expression is equivalent to maximizing the product of this expression and k_i. This product is

$$\begin{aligned} &k_i u_i(q(\theta_i', \theta_{-i}), \theta_i) + \sum_{j \neq i} k_j u_j(q(\theta_i', \theta_{-i}), \theta_j) + F(q(\theta_i', \theta_{-i})) \\ &\quad = \sum_{j=1}^{N} k_j u_j(q(\theta_i', \theta_{-i}), \theta_j) + F(q(\theta_i', \theta_{-i})) \end{aligned} \tag{7.4}$$

and hence agent i chooses his report to maximize the same function that q maximizes. Therefore, reporting θ_i truthfully is optimal. □

7.5 INDIVIDUAL RATIONALITY AND BUDGET BALANCE

We now enrich our framework to bring in individual rationality and budget balance. We begin with individual rationality. We model individual rationality in the same way as in the two previous chapters, except that we now require it ex post.

Definition 7.7 *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 Chapters 2–4, 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 need to restrict our attention to single-dimensionality of the type space. Proposition 5.9 immediately implies the following:

Proposition 7.9 *Consider any agent $i \in I$, and let R_i be an order of A. Suppose that the type set Θ_i is one-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; that is, $\theta_i \succ_{R_i} \underline{\theta}_i$ for all $\theta_i \in \Theta_i$ such that $\theta_i \neq \underline{\theta}_i$. Finally, assume that $\underline{a}_i$ is the lowest element of A in the order R_i; that is, $bR_i\underline{a}_i$ for every*

$b \in A$ with $b \neq \underline{a}_i$. Then a dominant strategy incentive-compatible mechanism satisfies the ex post individual rationality constraint for agent i if and only if for every $\theta_{-i} \in \Theta_{-i}$ we have

$$u_i(q(\underline{\theta}_i, \theta_{-i}), \underline{\theta}_i) - t_i(\underline{\theta}_i, \theta_{-i}) \geq u_i(\underline{a}_i, \underline{\theta}_i).$$

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 7.8 *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 7.10 *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*

$$\sum_{i=1}^{N} u_i(q(\theta), \theta_i) = \sum_{i=1}^{N} f_i(\theta_{-i}) \qquad \text{for all } \theta \in \Theta.$$

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 7.4. The mechanism is budget balanced if for all $\theta \in \Theta$ we have

$$\sum_{i=1}^{N} \left(-\sum_{j \neq i} u_j(q(\theta), \theta_j) + \tau_i(\theta_{-i}) \right) = 0 \Leftrightarrow$$

$$(N-1) \sum_{i=1}^{N} u_i(q(\theta), \theta_i) = \sum_{i=1}^{N} \tau_i(\theta_{-i}) \Leftrightarrow$$

$$\sum_{i=1}^{N} u_i(q(\theta), \theta_i) = \sum_{i=1}^{N} \frac{\tau_i(\theta_{-i})}{N-1}. \qquad (7.5)$$

Hence, if we set for every $i \in I$ and $\theta_{-i} \in \Theta_{-i}$

$$f_i(\theta_{-i}) \equiv \frac{\tau_i(\theta_{-i})}{N-1}, \tag{7.6}$$

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

$$\tau_i(\theta_{-i}) \equiv (N-1) f_i(\theta_{-i}). \tag{7.7}$$

Then for every $\theta \in \Theta$ the sum of agents' payments is

$$\begin{aligned}&\sum_{i=1}^{N}\left(-\sum_{j \neq i} u_j(q(\theta), \theta_j) + (N-1) f(\theta_{-i})\right)\\ &\quad = -(N-1)\sum_{i=1}^{N} u_i(q(\theta), \theta_i) + (N-1)\sum_{i=1}^{N} f(\theta_{-i}),\end{aligned} \tag{7.8}$$

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 nonempty interior, and suppose that θ, θ' are two different elements of that interior with $\theta < \theta'$. Then the necessary and sufficient condition in Proposition 7.10 requires

$$\begin{aligned}\theta &= \max\{\theta, \theta\} = f_B(\theta) + f_S(\theta),\\ \theta' &= \max\{\theta, \theta'\} = f_B(\theta) + f_S(\theta'),\\ \theta' &= \max\{\theta', \theta\} = f_B(\theta') + f_S(\theta),\\ \theta' &= \max\{\theta', \theta'\} = f_B(\theta') + f_S(\theta').\end{aligned} \tag{7.9}$$

Subtracting the third from the fourth equality in (7.9), we find that $f_S(\theta') = f_S(\theta)$; and subtracting the second from the fourth equality in (7.9), we find that $f_B(\theta') = f_B(\theta)$. But then the first right-hand side in (7.9) 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 7.1 holds, then Proposition 7.10 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 7.1. This observation applies to the bilateral trade model.

Note that the impossibility result that follows from Proposition 7.10 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 in Chapter 4, 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.

7.6 REMARKS ON THE LITERATURE

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 concept of "positive association of differences" and its characterization are in Roberts (1979). Finally, I have taken Proposition 7.10 from Milgrom (2004, p. 54), where this result is attributed to Bengt Holmström's 1977 Stanford PhD thesis.

7.7 PROBLEMS

(a) Give a direct proof that efficient decision rules are cyclically monotone. (Don't take the detour of first appealing to the fact that efficient decision rules are part of ex post incentive-compatible mechanisms and then applying Rochet's theorem. This would make this problem pointless.)

(b) Suppose every agent i's preferences were given by $u_i(a, \theta_i) - \mu_i(t_i)$, where $\mu_i : \mathbb{R} \to \mathbb{R}$ is strictly increasing. Let efficient allocation rules be defined as in Definition 7.3. Can you find a modification of the payments in VCG mechanisms that makes efficient decision rules dominant strategy incentive-compatible?

(c) Consider the problem of allocating a single indivisible good to one of two agents. Use the setup of Section 4.2. Is there a budget-balanced dominant strategy mechanism that allocates the good to the agent with the higher value? Note that we don't require the mechanism to be individually rational.

8

NONTRANSFERRABLE UTILITY

8.1 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. These are very restrictive assumptions. It is easy to imagine situations in which these assumptions are violated. The simplest case is that agents are not risk neutral in money. Another way in which the assumptions might be violated is that there are interactions between money and the allocative decision, for example because the value of different allocative decisions depends on how much money agents have left after paying transfers. 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 "nontransferrable utility."

On the spectrum 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 some results from 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.

8.2 THE GIBBARD–SATTERTHWAITE THEOREM

8.2.1 Setup

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 and transitive, and the only indifference is among identical alternatives. 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: $R = (R_1, R_2, \ldots, R_N)$. We also 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. The revelation principle then applies, and we can restrict attention to direct mechanisms.

Definition 8.1 *A "direct mechanism" is a function $f : \mathcal{R}^N \to A$.*

In the literature, direct mechanisms in the sense of Definition 8.1 are sometimes also called "social choice functions." To maintain consistency with earlier parts of this book, we shall speak of "direct mechanisms."

Definition 8.2 *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 setup 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 the 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 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 we can then restrict our attention to mechanisms that perform as we want them to perform only for some, but not for all preference profiles. Natural restrictions may be that either (a) all agents rank all elements of some subset $\hat{A}$ of A higher than all elements of the complement $A \setminus \hat{A}$ or (b) they give them 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.

8.2.2 Statement of the Result and Outline of the Proof

Definition 8.3 *A direct mechanism f is called "dictatorial" if there is some individual $i \in I$ such that for all $R \in \mathcal{R}^N$:*

$$f(R) R_i a \qquad \text{for all } a \in A.$$

Proposition 8.1 *Suppose that A has at least three elements and that the range of a direct mechanism f is A. Then 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 an impossibility result. It shows that the requirements of dominant strategy incentive compatibility and unlimited domain together are too strong. Our earlier discussion of dominant strategy mechanisms with quasi-linear preferences showed that this particular restriction for the space of admissible preferences may lead to more positive results. In Section 8.3 we shall discuss other restrictions of the preference domain that lead to more positive results, and in Section 8.4 we shall consider relaxations of the requirement of dominant strategy incentive compatibility.

In Gibbard and Satterthwaite's 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 redefine 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. Suppose A has only 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 8.1 is obvious. We therefore focus on proving the "necessity part." Our presentation of this proof is based on Reny (2001).[1] 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 8.4 *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:*

$$f(R) = a \text{ and for all } i \in I: \qquad [aR_i'b \text{ for all } b \in A \text{ such that } aR_ib]$$
$$\Rightarrow f(R') = a.$$

Proposition 8.2 *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 8.4 differed only in the ith 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 construct a sequence of preference profiles with the property that we start with R and end up with R', and each element of the sequence differs from the previous element of the sequence in only one component i. We can then apply the above argument successively to each element in the sequence, concluding in each step that the chosen alternative must be a, and thus we obtain the conclusion $f(R') = a$. □

Now we introduce two simple implications of monotonicity.

Definition 8.5 *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:*

$f(R) \in B$ *and* $[aR_i'a' \Leftrightarrow aR_ia'$ *for all* $a, a' \in A$ *such that at least one of* a, a' *is not in* $B] \Rightarrow f(R') \in B$.

Proposition 8.3 *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 $f(R') \notin B$. Then monotonicity implies that $f(R) = f(R')$, which contradicts $f(R) \in B$. □

Definition 8.6 *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:*

$$aR_ib \qquad \text{for all } i \in I \text{ and } b \in A \Rightarrow f(R) = a.$$

Proposition 8.4 *If f is monotone and the range of f is 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 reorder alternatives below a in arbitrary ways. Again, by monotonicity, the choice must remain a. This proves that f respects unanimity. □

8.2.3 Every Monotone Direct Mechanism Is Dictatorial

Proposition 8.5 *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"; that is, 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 8.1. In this figure, and in subsequent figures, each column corresponds to one agent's ranking of alternatives in A, with the highest ranked alternative at the top. Figure 8.1 shows a specific ranking, involving alternatives other than a, but for the moment this is of no relevance to the argument. In Figure 8.1, 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 8.1 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 there exists some preference profile R and some agent n such that according to R agent n ranks a highest whereas all other agents rank a lowest, as in Figure 8.1, and $f(R) = a$ through a sequence of steps starting with the profile shown in Figure 8.2. In the profile in Figure 8.2 the social choice has to be a because f respects unanimity. Now suppose that we move b up in

R_1	$\ldots$	R_{n-1}	R_n	R_{n+1}	$\ldots$	R_N
.		.	a	.		.
.		.	c	.		.
.		.	b	.		.
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
c	$\ldots$	c	.	c	$\ldots$	c
b	$\ldots$	b	.	b	$\ldots$	b
a	$\ldots$	a	.	a	$\ldots$	a

Figure 8.1 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 8.2 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 8.3 Social Choice Is a.

agent 1's ranking, until it is just below a, as shown in Figure 8.3. 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 8.4. 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 8.4, then we set $n = 1$. However, if the social choice remains a in Figure 8.4, then we repeat the same procedure for agent 2, and so on. 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 8.5 and 8.6 we show the generic situation for agent n before and after the social choice switches from a to b.

Figures 8.7 and 8.8 show the same profiles as Figures 8.5 and 8.6 except that we have moved alternative a to the bottom of the ranking for agents $i < n$, and we have

R_1	...	R_{n-1}	R_n	R_{n+1}	...	R_N
b	...	a	a	a	...	a
a	...	.	.	.	...	.
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
.		b	b	b	...	b

Figure 8.4 Social Choice Is a or b.

R_1	$\dots$	R_{n-1}	R_n	R_{n+1}	$\dots$	R_N
b	$\dots$	b	a	a	$\dots$	a
a	$\dots$	a	b	.	$\dots$	.
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
.		.	.	b	$\dots$	b

Figure 8.5 Social Choice Is a.

R_1	$\dots$	R_{n-1}	R_n	R_{n+1}	$\dots$	R_N
b	$\dots$	b	b	a	$\dots$	a
a	$\dots$	a	a	.	$\dots$	.
.		.	.	.		.
.		.	.	.		.
.		.	.	.		.
.		.	.	b	$\dots$	b

Figure 8.6 Social Choice Is b.

R_1	$\dots$	R_{n-1}	R_n	R_{n+1}	$\dots$	R_N
b	$\dots$	b	a	.	$\dots$	.
.	$\dots$	.	b	.	$\dots$	.
.		.	.	.		.
.		.	.	.		.
.		.	.	a		a
a		a	.	b	$\dots$	b

Figure 8.7 Social Choice Is a.

R_1	$\dots$	R_{n-1}	R_n	R_{n+1}	$\dots$	R_N
b	$\dots$	b	b	.	$\dots$	.
.	$\dots$	.	a	.	$\dots$	.
.		.	.	.		.
.		.	.	.		.
.		.	.	a		a
a		a	.	b	$\dots$	b

Figure 8.8 Social Choice Is b.

moved a 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 8.5 and 8.6. For Figure 8.8 it follows from Figure 8.6 and monotonicity that the choice has to be b. Now consider the profile in Figure 8.7. Comparing Figure 8.7 and Figure 8.8 we can conclude from set monotonicity (with $B \equiv \{a, b\}$) that the social choice in Figure 8.7 has to be either a or b. But if the alternative chosen were b, then it would also have to be b in Figure 8.5, by monotonicity. Therefore, the choice has to be a in Figure 8.7.

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 8.9 Social Choice Is a.

Now consider the preferences in Figure 8.9. The position of alternative a has not changed relative to other alternatives in comparison to Figure 8.7, and therefore the social choice has to be a.

Finally, compare Figure 8.9 to Figure 8.1. The choice in Figure 8.1 has to be a or b, by set monotonicity, setting $B = \{a, b\}$. But if the choice were b in Figure 8.1, 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 8.1 has to be a. This concludes the proof. $\square$

8.3 DOMINANT STRATEGY INCENTIVE COMPATIBILITY ON RESTRICTED DOMAINS

Two natural ways of relaxing the stringent requirements of Proposition 8.1 so that more positive results obtain are, first, to consider a more restricted domain of preferences and, second, 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: $k(i)R_i a$ for all $a \in A$ and (ii) agent i's preferences decline monotonically "to the left" and "to the right" of $k(i)$—that is, if $\ell \geq k(i)$, then $\ell R_i(\ell + 1)$; and if $\ell \leq k(i)$, 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 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 8.6 *Suppose that A has at least three elements and that the range of f is A. If preferences are single-peaked, then there are dominant strategy incentive-compatible 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. We shall call this rule the *median voting mechanism.* 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 $\left(\frac{N-1}{2}\right)$th largest and the $\left(\frac{N+1}{2}\right)$th largest alternative among the most preferred alternatives of the $N-1$ agents other than i. Denote these by m_- and m_+. The median will be between 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 all cases, 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 nondictatorial.

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. □

Other restricted domains on which dominant strategy incentive compatibility does not imply dictatorships are, of course, those that we studied in Chapters 4 and 7. In those environments we assumed that agents' utility functions were of a particular form, for example, satisfying additive separability and risk neutrality. This domain restriction leads 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 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; that is, the alternative that is chosen maximizes this individual's preference relation. A formal statement of Hylland's theorem appears in Chapter 10 where it is Proposition 10.12.

8.4 BAYESIAN INCENTIVE COMPATIBILITY

Although much of the literature on mechanism design with nontransferrable utility has focused on dominant strategy incentive compatibility, there is no fundamental reason why we should not study Bayesian incentive compatibility as well in this setting. Among the papers that explore this approach, I shall, in a self-centered way, briefly discuss Börgers and Postl (2009). Other papers on this subject are mentioned in the final subsection.

Postl and I consider in our paper 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 (2009) interpret the alternative b as a "compromise" because 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 a Bayesian 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 each yield a utility sum 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 (2009) 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 some upper bound of probability of their preferred alternative.

Using arguments similar to our arguments in Section 3.3, Börgers and Postl (2009) show that no Bayesian incentive-compatible mechanism implements the first best—that is, choice of the compromise with probability 1 if and only if $\theta_1 + \theta_2 \geq 1$. The problem of finding a second best mechanism 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 (2009) have analytical results only for small parameterized 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.

8.5 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 so far 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 8.6 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 voting mechanism if agents are restricted to report only their top preference. Gershkov et al. (2013b) have shown in a setting with cardinal utility and single-peaked preferences that a mechanism designer maximizing utilitarian welfare will choose a similar generalized version of the median voting mechanism. They do not assume that agents are restricted to reporting only their top preference.

Our discussion of Hylland's result is based on Dutta et al. (2007) and Dutta et al. (2008), 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, 1978).

Other papers that investigate Bayesian incentive compatibility without transferrable utility include Miralles (2012). He characterizes the ex ante expected utilitarian welfare-maximizing mechanism for allocating two indivisible goods among a set of agents. In case all agents prefer the same good, the optimal mechanism involves an auction in which agents use probabilities of their preferred good as "bidding currency." This is related to the interpretation of the probability with which agents give up their preferred alternative as contributions to a public good in Börgers and Postl (2009).

A Bayesian perspective on voting is also pursued in Jackson and Sonnenschein (2007). They consider Bayesian incentive-compatible mechanisms for a large set of independent decision problems, and they explain how a mechanism designer who is concerned with ex ante Pareto efficiency can elicit strength of preferences by linking the decision problems. Very informally speaking, agents in Jackson and Sonnenschein's mechanism can announce large intensities of preference for some alternative only for a limited number of decisions problems.

A large proportion of the literature on the design of matching markets is also in a setting without transferrable utility. When strategic properties of mechanisms are considered in this literature, the focus is more often on dominant strategies than on Bayesian incentive compatibility. The literature is too large to be even touched upon here; for a survey see Roth (2008).

8.6 PROBLEMS

(a) Do Propositions 8.2, 8.3, and 8.4 remain true when the domain of f is not the set of *all* possible preference profiles, but some subset of this set, and when the range of f is not necessarily A?

(b) Explain where and how we used in the proof of Proposition 8.5 the assumption that A has three elements.

(c) Prove that if f is monotone and the range of f is A, then there cannot be a preference profile R and an alternative $a \in A$ such that $aP_i f(R)$ for all $i \in I$.

(d) Give an example of a dominant strategy incentive-compatible direct mechanism on the set of all single-peaked preferences that is not dictatorial and different from the direct mechanism described in the proof of Proposition 8.6.

(e) In the setting of Section 8.4 prove that a decision rule that picks the compromise whenever that is efficient, and otherwise picks A and C with equal probability, is not Bayesian incentive-compatible.

9

INFORMATIONAL INTERDEPENDENCE

9.1 INTRODUCTION

In all previous chapters of this book, we have assumed that each agent's private information is all that matters for that agent's preferences over group choices. In this chapter, 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 private information of other agents. We shall refer to this as the case of "informational interdependence." By contrast, we shall 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 the radio spectrum for operating 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 this book except Chapter 8. This will make it easier to contrast the results of this section with those of previous sections.

For the case of informational interdependence, 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 spaces will play an important role in this chapter. 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 of 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 was independent of the dimensions of the signal spaces. Efficient decision rules could be implemented using Vickrey–Clarke–Groves mechanisms.

9.2 AN EXAMPLE

We start off with an extremely simple example. Suppose that there are two 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. If alternative a is chosen and agent i pays transfer t_i, agent i's utility is $\theta_i - t_i$. If alternative b is chosen and agent i pays transfer t_i, agent i's utility 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 $q : [0, 1]^2 \to [0, 1]$ maps agent 1's private signal into the probability with which alternative a is chosen. A transfer rule $t : [0, 1]^2 \to \mathbb{R}$ 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 satisfies $q^*(\theta_1, \theta_2) = 0$ 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.[1]

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:

$$\begin{aligned} &\theta_1 - t_a = 0.5 - t_b \Leftrightarrow \\ &t_a - t_b = \theta_1 - 0.5. \end{aligned} \tag{9.1}$$

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.

Jehiel and Moldovanu (2001) have generalized the above example in two ways. First, 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 in this environment. Their argument for this more general case is slightly different from the above argument. Second, 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 and Moldovanu's second extension of the example.

9.3 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: $\theta^i = (\theta^i_1, \theta^i_2, \ldots, \theta^i_K) \in [0, 1]^K$.[2] The signal θ^i has a distribution with density f^i that is positive everywhere. Different agents' signals are independent.

If alternative a_k is chosen and if agent i has to pay transfer t_i, agent i's von Neumann–Morgenstern utility is.

$$\sum_{j=1}^{N} \alpha^j_{ki} \theta^j_k - t_i. \tag{9.2}$$

Thus, agent i's utility from alternative a_k is a linear function of the kth component of all agents' signals. The factor in front of the kth 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, first, of a mapping $q : [0, 1]^{IK} \to \Delta(A)$ that assigns to each vector of reported private signals $\theta \equiv (\theta^1, \theta^2, \ldots, \theta^N)$ 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 $Q^i_a : [0, 1]^K \to [0, 1]$ 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 Chapter 7 we defined agent i's transfer rule in a VCG mechanism by

$$t_i(\theta) = -\sum_{j\neq i} u_j(q(\theta), \theta_j) + \tau_i(\theta_{-i}) \qquad \text{for all } \theta \in \Theta, \tag{9.3}$$

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

$$u_i(q(\theta_i', \theta_{-i}), \theta_i) + \sum_{j\neq i} u_j(q(\theta_i', \theta_{-i}), \theta_j) - \tau_i(\theta_{-i})$$

$$= \sum_{j=1}^{N} u_j(q(\theta_i', \theta_{-i}), \theta_j) - \tau_i(\theta_{-i}). \tag{9.4}$$

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

$$t_i(\theta) = -\sum_{j\neq i} u_j(q(\theta), \theta) + \tau_i(\theta_{-i}) \qquad \text{for all } \theta \in \Theta. \tag{9.5}$$

Now, if agent i is of type θ_i and reports that he is of type θ_i', then his utility becomes

$$u_i(q(\theta_i', \theta_{-i}), \theta) + \sum_{j\neq i} u_j(q(\theta_i', \theta_{-i}), (\theta_i', \theta_{-i})) - \tau_i(\theta_{-i}). \tag{9.6}$$

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 under some conditions no direct mechanism implements a first best decision rule.

Proposition 9.1 *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

$$\frac{\alpha^i_{ai}}{\alpha^i_{bi}} \neq \frac{\sum_{j=1}^N \alpha^i_{aj}}{\sum_{j=1}^N \alpha^i_{bj}}.$$

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 9.1 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 9.1 presents a generic impossibility result.

Note the contrast between Proposition 9.1 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, Proposition 9.1 refers 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 Envelope Theorem implies for every agent i and every $a \in A$:

$$\frac{\partial U^i}{\partial \theta^i_a} = \alpha^i_{ai} Q^i_a(\theta^i). \tag{9.7}$$

Now suppose we differentiate this expression again, this time with respect to θ^i_b for some $b \in A$. Then we obtain

$$\frac{\partial^2 U^i}{\partial \theta^i_a \partial \theta^i_b} = \alpha^i_{ai} \frac{\partial Q^i_a(\theta^i)}{\partial \theta^i_b}. \tag{9.8}$$

If we change the order of differentiation, we obtain similarly:

$$\frac{\partial^2 U^i}{\partial \theta^i_b \partial \theta^i_a} = \alpha^i_{bi} \frac{\partial Q^i_b(\theta^i)}{\partial \theta^i_a}. \tag{9.9}$$

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.

$$\alpha^i_{ai} \frac{\partial Q^i_a(\theta^i)}{\partial \theta^i_b} = \alpha^i_{bi} \frac{\partial Q^i_b(\theta^i)}{\partial \theta^i_a} \Leftrightarrow$$

$$\frac{\alpha^i_{ai}}{\alpha^i_{bi}} = \frac{\dfrac{\partial Q^i_b(\theta^i)}{\partial \theta^i_a}}{\dfrac{\partial Q^i_a(\theta^i)}{\partial \theta^i_b}}. \tag{9.10}$$

Here, we could divide by the partial derivative of the interim probability of a with respect to θ^i_b because, by the first assumption in Proposition 9.1, this derivative is 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:

$$\frac{\sum_{j=1}^N \alpha^i_{aj}}{\sum_{j=1}^N \alpha^i_{bj}} = \frac{\dfrac{\partial Q^i_b(\theta^i)}{\partial \theta^i_a}}{\dfrac{\partial Q^i_a(\theta^i)}{\partial \theta^i_b}}. \tag{9.11}$$

We can now deduce from our results so far that

$$\frac{\alpha^i_{ai}}{\alpha^i_{bi}} = \frac{\sum_{j=1}^N \alpha^i_{aj}}{\sum_{j=1}^N \alpha^i_{bj}}. \tag{9.12}$$

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 9.1. □

9.4 CHARACTERIZING ALL INCENTIVE-COMPATIBLE MECHANISMS

The impossibility result presented in the previous section is remarkable because it holds even if neither individual rationality nor budget balance is required. 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 subgradient 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 his report. We then say that truth telling is an "ex post Bayesian equilibrium" of the direct mechanism.[3] If truth telling is an ex post Bayesian 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 is, rules that choose the same alternative for every type realization—are part of an incentive-compatible direct mechanism if each agent's signal space is at least two-dimensional. Note that this result is extraordinarily strong. In previous chapters, we have seen several contexts in which a very limited class of decision rules can be part of a dominant strategy incentive-compatible direct mechanism. But none of these results is as restrictive as that of Jehiel et al. (2006). In that result, the only decision rules that can be implemented in ex post equilibrium are those where the decision rule does not respond at all to agents' private information.

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: If all agents other than i revealed their private information (and they would be willing to do so, as they cannot affect the collective decision), agent i will, in some cases, have an ex post incentive to deviate, and to change his decision, based on what he has learned from the other agents' signals. Therefore, agent i picking the alternative that is best given her own information is not ex post incentive-compatible.

The example of the previous paragraph is revealing. It clarifies what ex post incentive compatibility really requires: Each agents' choice must be optimal, given their signal regardless of what the agent believes about the other agents' signals. Note that ex post equilibrium is still an equilibrium concept: Agents are supposed to fully understand the other agents' strategies (typically truth telling) that map the other agents' signals into their choices. But, given knowledge of the strategy, optimality of each agent's strategy with respect to all beliefs about the other agents' types is required. The result of Jehiel et al. shows that this requirement is so strong that it does not allow the implementation of any nontrivial decision rule.

9.5 REMARKS ON THE LITERATURE

The example in Section 9.2 is a special case of an example that appears in Jehiel and Moldovanu (2001). This paper is also the main reference for Proposition 9.1. However, the short proof of Proposition 9.1 that I have given is taken from Jehiel and Moldovanu (2005). Subsection 9.4 is based on Jehiel et al. (2006).

9.6 PROBLEMS

(a) For the example in Section 9.2, determine the direct mechanism that maximizes expected welfare among all Bayesian incentive-compatible direct mechanisms.

(b) Suppose a single indivisible object has to be allocated to one of two agents. Each agent observes a signal θ_i. The sum of the signals, $\theta_1 + \theta_2$ constitutes agent i's valuation of the good. Note that both agents have the same valuation. The agents have quasi-linear utility. A second price auction is conducted with minimum bid zero to allocate the good. Find all ex post equilibria of this auction.

(c) There are two alternatives: a and b and two agents: $i = 1, 2$. Each agent i observes $\theta_a^i \in [0, 1]$ and $\theta_b^i \in [0, 1]$. Agent i's utility from a is given by $\theta_a^1 + \theta_a^2$, whereas agent i's utility from b is given by θ_b^i. Assume that there is no money and no transferrable utility. Consider only deterministic direct mechanisms. For which direct mechanisms is truth telling an ex post equilibrium?

10

ROBUST MECHANISM DESIGN

10.1 INTRODUCTION

The name of the area of research that we cover in this chapter, "robust mechanism design," can mislead. One might expect to find results about the robustness of optimal mechanisms to *small* changes in the environment. This is not the focus of "robust mechanism design." Instead, "robust mechanism design" refers to the robustness of optimal mechanisms to *large* changes in the environment. In other words, we shall assume that the mechanism designer has a high degree of uncertainty about the environment that she faces. The mechanism designer does not know the agents' utility functions, and she does not know what agents believe about each others' utility functions, what they believe about each others' beliefs, and so on. We thus study a mechanism designer who cannot rule out a large set of possible circumstances in which she might find herself and who needs to design the mechanism in the face of this wide ranging uncertainty.

An example where such uncertainty is realistic is the design of a country's constitution. Constitutions are typically written when it is not possible to foresee all the circumstances in which the constitution will apply. For example the Constitutional

Convention that drafted between 1787 and 1789 the relatively short document that is still the Constitution of the United States of America (although 27 amendments have since been added) could not foresee that they were choosing the rules by which 221 years later decisions about a federal act regarding health care reform were made.

How does robust mechanism design differ from the theory of mechanism design that we have discussed earlier? To explain, we need to be more systematic about the uncertainty that a mechanism designer faces. If we think of agents in the classical way as rational, Bayesian agents,[1] then there are two such types of uncertainties: First, there is uncertainty about the observations that agents make and that enter directly their own or other agents' payoff functions. Second, there is the uncertainty about agents' beliefs about each others' observations, their beliefs about others' beliefs, their beliefs about others' beliefs about others' beliefs, and so on.

Uncertainty of the first type is already captured by the Bayesian models that we have discussed earlier. One may argue that the specifications of payoff relevant observations that we have considered imply unrealistically small domains of possible preferences of the agents. Allowing larger domains of preferences has been studied by a few authors (Bierbrauer et al., 2014), but there is no well-developed theory on this subject yet.

Traditional models in mechanism design make very restrictive assumptions regarding the second type of uncertainty that we mentioned above—that is, about agents' hierarchies of beliefs about each other. Models with independent types assume that agents' beliefs about other agents are common knowledge and are also known by the mechanism designer. Models with correlated types that satisfy Crémer and McLean's condition assume that a one-to-one relation between agents' payoff relevant observations and beliefs about other agents is common knowledge among the agents and is known by the mechanism designer. Relaxing these assumptions has been the focus on the recent literature on robust mechanism design. We shall follow the literature in this chapter, and we shall also focus on this form of uncertainty of the mechanism designer.

There is an additional agenda implicit in much of the literature on mechanism design, although it is not formalized explicitly. Mechanisms that are optimal for large type spaces may be very complicated. They are not tailored to particular assumptions about the details of the environment, but, as we will see, they are complicated because they are designed to function differently for each possible realization of these details. Robust mechanism design seems more attractive if one can find not just *any* mechanism, but a *simple* mechanism, that can be used in a large variety of situations. But it is hard to formalize the notion of "simplicity." We shall incorporate simplicity considerations from time to time below, but our discussion of such considerations will be largely informal.

This chapter seeks to describe the agenda of "robust mechanism design" and to report some existing results. Two words of warning in advance: First, the description of the research agenda in this chapter reflects my personal views on robust mechanism design. This chapter is more subjective than any of the previous chapters. I am, of course, not entirely sure that my personal perspective on this subject is reasonable or useful. Other approaches of interest will be explained in Section 10.12, which discusses related literature. The second word of warning is that in this chapter the description of the research agenda is more weighty than the presentation of results. This reflects that there is much more work to be done. Perhaps, this also reflects how difficult it is to do more.

10.2 AN EXAMPLE

To begin with, we consider a simple example that falls into none of the special categories of Bayesian mechanism design that we developed earlier in this book. The example shows that there is in principle no conceptual difficulty in considering type spaces that embody more complicated hierarchies of beliefs than the ones that we have considered so far. The example also provides a first hint that optimal mechanisms for large type spaces may be counterintuitive and complicated. The conceptual, as well as some technical, issues raised by the example will constitute a focus of the remainder of this chapter.

We consider a seller who sells a single indivisible good to two potential buyers. The buyers have additively separable, quasi-linear utility. Buyer 1's value for the good is 1, 4, or 5. Buyer 2's value for the good is either 0 or 2. The joint distribution of the two values is shown in Figure 10.1 where rows correspond to buyer 1 and columns correspond to buyer 2.

Observe that types are not independent. On the other hand, buyer 1 has two types with the same beliefs: the type with value 4 and the type with value 5. Therefore, the Crémer– McLean condition is trivially violated for buyer 1. For buyer 2, by contrast, this condition is satisfied. Because types are not independent, but also Crémer and McLean's condition does not hold, we have an example that is outside of the framework considered so far. Of course, the example is simple, and the range of uncertainty that we allow for the mechanism designer is small. Nonetheless, we can illustrate some important issues in the theory of robust mechanism design in this example.

The seller seeks to find an expected revenue-maximizing mechanism subject to incentive compatibility and individual rationality constraints. Without loss of

	0	2
1	$\frac{1}{11}$	$\frac{2}{11}$
4	$\frac{2}{11}$	$\frac{2}{11}$
5	$\frac{2}{11}$	$\frac{2}{11}$

Figure 10.1 A Type Space.

generality, we restrict our attention to direct mechanisms. We proceed in two steps. We first determine the mechanism that would be expected revenue maximizing if the only relevant incentive compatibility was that the two types of buyer 1 that have identical beliefs do not have an incentive to imitate each other. We then use the argument of Crémer and McLean to show that this mechanism can be modified so that all other incentive compatibility constraints are satisfied, too. The modified mechanism then maximizes expected revenue among the mechanisms that satisfy all constraints.

The reason why we can proceed in this way is that Crémer and McLean's argument allows the mechanism designer to extract buyers' beliefs at no cost. Therefore, only incentive constraints that apply to buyers with identical beliefs bind. In our example, only the types 4 and 5 of buyer 1 have identical beliefs. Therefore, the only incentive constraint that we need to consider initially, before appealing to Crémer and McLean's argument, concerns these two types.

Let p_4 be the interim probability with which type 4 of buyer 1 obtains the good, and let t_4 denote this type's interim expected payment. Let p_5 be the interim probability with which type 5 of buyer 1 obtains the good, and let t_5 denote this type's interim expected payment. As usual, a necessary condition for types 4 and 5 not to want to imitate each other is that $p_5 \geq p_4$. If p_4 and p_5 have been chosen, then the optimal interim expected payments are $t_4 = 4p_4$, which ensures that type 4's participation constraint is satisfied as an equality, and $t_5 = 5p_5 - p_4$, which ensures that type 5's incentive constraint holds as an equality. We omit the argument that shows that these payments also satisfy type 5's individual rationality constraint and type 4's incentive compatibility constraint and that these payments are optimal. We shall now determine the optimal values of p_4 and p_5.

Suppose first p_4 is given, and the seller chooses only p_5. The seller's expected revenue from type 5, conditional on buyer 1 being type 5, is $5p_5 - p_4$. If the seller does not sell to buyer 1, then either she sells to buyer 2, which gives her at most a revenue of 2, or she does not sell at all, earning a revenue of 0. The seller will therefore want to choose $p_5 = 1$, that is, sell to type 5 with certainty. The incentive compatibility constraint $p_5 \geq p_4$ is then obviously satisfied irrespective of what p_4 is.

We now consider the optimal choice of p_4. The seller's expected revenue from type 4, conditional on buyer 1 being type 4, is $4p_4$. Raising p_4 also reduces what she obtains from type 5, as determined above: For each unit increase in p_4, she loses one unit in revenue from type 5, which is less than the four units that she gains from type 4. Because, moreover, types 4 and 5 are equally likely, and the alternative of selling to buyer 2, or not selling, is clearly inferior, the seller's optimal choice is $p_4 = 1$. This gives us the allocation rule indicated in Figure 10.2, where, in case that buyer 1 is of type 1, the good is simply allocated to the buyer type with the higher value, as no incentive compatibility constraints have to be respected.

	0	2
1	$(1,0)$	$(0,1)$
4	$(1,0)$	$(1,0)$
5	$(1,0)$	$(1,0)$

Figure 10.2 An Optimal Mechanism's Allocation Rule.

The corresponding interim expected payments of types 4 and 5 are $t_4 = t_5 = 4$. For simplicity we assume that these interim expected payments are implemented with zero variance; that is, buyer 1 has to pay them regardless of buyer 2's type. In the case when buyer 1 is of type 1, we only have to respect individual rationality constraints, and not incentive constraints, and therefore we assume that whoever gets the object pays their true value. This leads to the payment rules indicated in Figure 10.3.

The mechanism that we have constructed so far violates exactly two of the omitted incentive constraints: Types 4 and 5 have an incentive to pretend to be type 1. To satisfy these types' incentive constraints, we modify the payment rule for buyer 1 so that no type's interim expected payments are changed if they report their beliefs truthfully, but so that the losses from misreporting beliefs outweigh any gains that could be attained. To achieve this, we apply the argument of Crémer and McLean. Figure 10.4 shows payments for player 1 that have conditional expected value zero for each type of player 1, such that no player has an incentive to misreport their type. Negative numbers are payments that buyer 1 receives from the auctioneer, while positive numbers are payments that buyer 1 makes to the auctioneer. In fact, these payments have also been chosen so that the losses from misreporting a type are sufficiently large to remedy all incentive problems in our mechanism. We thus add the payments from Figure 10.4 to the payments in Figure 10.3, and we obtain the payment rule in Figure 10.5.

The allocation rule in Figure 10.2 and the payment rule in Figure 10.5 together describe an expected revenue maximizing selling mechanism. The payment rule for buyer 1 is quite counterintuitive. In some cases, buyer 1 obtains the good for free. In

	0	2
1	$(1,0)$	$(0,2)$
4	$(4,0)$	$(4,0)$
5	$(4,0)$	$(4,0)$

Figure 10.3 An Optimal Payment Rule that Provides Incentives to Buyer 1's Types 4 and 5 to Reveal Their Payoff Types.

	0	2
1	8	−4
4	−4	4
5	−4	4

Figure 10.4 Revealing Player 1's Beliefs.

	0	2
1	$(9, 0)$	$(-4, 2)$
4	$(0, 0)$	$(8, 0)$
5	$(0, 0)$	$(8, 0)$

Figure 10.5 An Optimal Mechanism's Payment Rule.

other cases, buyer 1 has to pay far above her willingness to pay. On average, she pays exactly the same as in the mechanism in Figure 10.3.

Two observations are important. The first is technical: To maximize expected revenue, we initially only considered the incentive constraints of types with identical beliefs but different valuations. The remaining incentive constraints could then be satisfied using a Crémer–McLean transformation. This reflects a general result that we state later in this chapter. The second observation is conceptual: The optimal mechanism is not particularly intuitive, but it is tailored to a very specific environment. We extended in this example the range of uncertainty about beliefs that we consider in comparison to earlier sections. But we should be reluctant to describe the optimal mechanism that we obtained as a "robustly optimal mechanism." We shall later consider much larger type spaces than we did in this example. But the problem that the optimal mechanisms seem very complicated does not go away. This seems to indicate that optimal mechanisms in the context of robust mechanism design need not only be designed having in mind a variety of beliefs that agents might hold, but also need to satisfy other conditions—in particular, simplicity. But such requirements are hard to capture formally. We shall return to this point later, but in the next section we first focus on agents' beliefs.

10.3 MODELING INCOMPLETE INFORMATION

To fully explain how the details of our assumptions about agents' information and beliefs affect the study of mechanism design, we describe in this section a very general model of incomplete information. This framework, which is due to Bergemann and Morris (2005), differs from the standard framework used in game theory to model incomplete information due to Harsanyi (1967–1968) only in minor ways.

The space over which agents are uncertain is modeled in the same way as in previous chapters. We shall slightly change the terminology in comparison to previous chapters, however. We shall split up what the previous chapters referred to as agents' "types" into two components: agents' "payoff types" and agents' "belief types." We denote "payoff types" using the same notation that we previously used to denote "types." Thus, a payoff type of an agent $i \in \{1, 2, \ldots, N\}$ will be denoted by θ_i, and the set of all payoff types of agent i will be Θ_i. The set of payoff types of agent i will be interpreted as the sets of potentially *payoff relevant* observations that agent i might make.

Utility functions are only based on payoff types and actions and are thus of the form $u_i : A \times \Theta \to \mathbb{R}$. Belief types, which will be introduced below, do not enter the utility functions. This setup reflects the assumption that we discussed in the introduction to this chapter—that is, that the mechanism designer knows which payoff relevant observations each agent might make, and how those payoff relevant observations jointly determine each agent's von Neumann–Morgenstern utility, and that the mechanism designer is certain that this information is common certainty among agents.

Note that we allow nontransferrable utility, as well as interdependence of preferences. We shall think of the sets Θ_i as finite, or as products of intervals in finite-dimensional Euclidean space. This allows us to define probability measures for these sets. We denote the set of all probability measures on any set X for which it is obvious how to define probability measures on that set by $\Delta(X)$.

10.3.1 Hierarchies of Beliefs

We assume that each agent i knows their own type θ_i. We focus on the mechanism designer's uncertainty about what agents believe about each other. To begin with, therefore, we describe formally agents' beliefs about each other. In an environment in which several people are interacting with each other, and everyone's utility potentially depends on other agents' actions, it seems compelling that we need to consider not only an agent's beliefs about other agents' types, but also the agent's beliefs about other agents' beliefs, his beliefs about other agents' beliefs about his own beliefs, and so on. Indeed, one is led to consider an infinite number of layers of such beliefs. There never seems to be a good reason to stop constructing further layers of belief. Wherever incomplete information of some form or another is involved in the theory of games, infinite hierarchies of beliefs are either implicitly or explicitly assumed. We shall begin by formally constructing infinite hierarchy of beliefs. The construction that we describe is a standard construction of game theory (Harsanyi, 1967–1968; Mertens and Zamir, 1985; Brandenburger and Dekel, 1993) adapted to the current framework. It was also sketched in Bergemann and Morris (2001), which is an early discussion paper version of Bergemann and Morris (2005).

Agent i's beliefs about other agents' payoff types is a measure:

$$b_{i,1} \in B_{i,1} = \Delta\left(\prod_{j \neq i} \Theta_j\right). \tag{10.1}$$

This is called agent i's "first-order belief." We also want to describe agent i's beliefs about other agents' beliefs. These will be called agent i's "second-order beliefs." Agent i might think that there is a connection between other agents' payoff types and other agents' beliefs. Therefore, agent i's second-order beliefs are beliefs about the Cartesian product of the sets of other agents' payoff types and other agents' first-order beliefs:

$$b_{i,2} \in B_{i,2} = \Delta\left(\prod_{j \neq i} \left(\Theta_j \times B_{j,1}\right)\right). \tag{10.2}$$

From $b_{i,2}$ we can derive a marginal distribution on $\prod_{j \neq i} \Theta_j$. For agent i's beliefs to be coherent, this marginal distribution must coincide with agent i's first-order beliefs $b_{i,1}$. We shall denote the subset of belief pairs in $B_{i,1} \times B_{i,2}$ that satisfy this consistency condition by $\mathcal{B}_{i,2}$.

Next, we construct agent i's third-order beliefs. These are beliefs about other agents' payoff types, as well as their first- and second-order beliefs. We shall assume that agent i is certain that the other agents' first- and second-order beliefs are coherent. Therefore, agent i's third-order beliefs satisfy

$$b_{i,3} \in B_{i,3} = \Delta\left(\prod_{j \neq i} \left(\Theta_j \times \mathcal{B}_{j,2}\right)\right). \tag{10.3}$$

From $b_{i,3}$ we can derive a marginal distribution on $\prod_{j \neq i} \left(\Theta_j \times B_{j,1}\right)$ that should coincide with $b_{i,2}$. We shall call the subset of belief triples in $\mathcal{B}_{i,2} \times B_{i,3}$ that satisfy this consistency condition $\mathcal{B}_{i,3}$.

We can now generalize. Suppose we had defined all coherent hierarchies of beliefs up to level n, and we had called the set of all such hierarchies of beliefs $\mathcal{B}_n$. Agent i's $(n+1)$th-order belief will satisfy

$$b_{i,n+1} \in B_{i,n+1} = \Delta\left(\prod_{j \neq i} \left(\Theta_j \times \mathcal{B}_{j,n}\right)\right). \tag{10.4}$$

From $b_{i,n+1}$ we can derive a marginal distribution on $\prod_{j \neq i} \left(\Theta_j \times \mathcal{B}_{j,n-1}\right)$. Consistency of i's beliefs require that this marginal distribution is the same as agent i's nth order belief $b_{i,n}$ We shall denote the sets of all belief sequences in $\mathcal{B}_{j,n} \times B_{i,n+1}$ that satisfy this consistency condition by $\mathcal{B}_{i,n+1}$.

We now complete our construction:

Definition 10.1 *An* infinite hierarchy of beliefs *of agent i is a sequence of beliefs $b_i = (b_{i,n})_{n \in \mathbb{N}}$ such that for every $n \in \mathbb{N}$ we have $(b_{i,1}, b_{i,2}, \ldots, b_{i,n}) \in \mathcal{B}_{i,n}$. We denote the set of all such sequences by $\mathcal{B}_i$.*

One might think that we are still not done once we have constructed these infinite hierarchies of beliefs. Shouldn't we also consider agents' beliefs over $\prod_{j \neq i} \left(\Theta_j \times \mathcal{B}_j\right)$?

The answer to this question depends on details that we have left out of our presentation. It is sufficient for our purposes to say that one can specify these details in a reasonable way so that the infinite hierarchy of beliefs that we have constructed already implies what all further beliefs that one might wish to consider have to be. In this sense, the countably infinite hierarchy of beliefs provides all information about an agent's beliefs that might be relevant.[2]

The uncertainty of the mechanism designer that we want to capture in this chapter can now be described as follows: The mechanism designer does not know which hierarchies of beliefs each agent might hold, and she thinks that large sets of hierarchies of beliefs—that is, belief hierarchies in a large subset of $\mathcal{B}_1 \times \mathcal{B}_2 \times \ldots \times \mathcal{B}_N$—are possible. The framework introduced so far is, however, cumbersome, and we next present a modeling language that makes it easier to deal with infinite hierarchies of beliefs.

10.3.2 Type Spaces

We return to the notion of an agent's "type." We used this expression earlier, but now we use it in a slightly different sense. We distinguish, as indicated before, an agent's payoff type from an agent's type. An agent's type, as we use the expression in this section, represents the agent's payoff type, and the agent's infinite hierarchy of beliefs, to which we shall refer as the agent's "belief type." Describing payoff types and infinite hierarchies of beliefs using the notion of a "type" is the method for modeling incomplete information in games that was first introduced by Harsanyi (1967–1968). Game theory textbooks often don't even introduce the language of hierarchies of beliefs explicitly, but instead use the language of type spaces from the outset. The set of all vectors of types, one for each agent, is called a "type space." The definition of an agent's type space is then completed by describing the agent's beliefs about other agents' types.

Definition 10.2 *A* type space *is a list* $\mathcal{T} = \left(T_i, \hat{\theta}_i, \hat{\beta}_i\right)_{i \in I}$ *where for each agent* $i \in I$ T_i *is some nonempty set, and* $\hat{\theta}_i$ *and* $\hat{\beta}_i$ *are functions of the form*[3]*:*

$$\hat{\theta}_i : T_i \to \Theta_i \quad \textit{and} \quad \hat{\beta}_i : T_i \to \Delta(T_{-i}). \tag{10.5}$$

Note that in this definition we have not referred to a prior probability distribution over the set of type vectors. We shall address the special case in which types' beliefs are derived from a common prior in the next subsection.

For each type $\tau_i \in T_i$, we can construct a hierarchy of beliefs that agent i holds when her type is τ_i. Agent i's first-order belief, with type τ_i, is defined as follows: We know for each measurable subset of T_{-i} the subjective probability that i attaches to that set. Moreover, because we know the functions $\hat{\theta}_j, j \neq i$, we know for each element of T_{-i} the corresponding payoff type θ_{-i}. Combining these, we obtain agent i's belief about the other agents' payoff types, that is, agent i's first-order belief. We can do this for every agent i, for every $\tau_i \in T_i$, and thus obtain for every agent i a function $\hat{b}_{i,1} : T_i \rightarrow B_{i,1}$. But now we can construct agent i's second-order belief: We know for each measurable subset of T_{-i} the subjective probability that i attaches to that set. Moreover, because we know the functions $\hat{\theta}_j, j \neq i$, and $\hat{b}_{j,1}, j \neq i$, we know for each element of T_{-i} the corresponding payoff types θ_{-i} as well as the corresponding first-order beliefs $b_{-i,1}$. Combining these, we obtain agent i's beliefs about the other agents' payoff types and their first-order beliefs, that is, agent i's second-order belief. This gives us a function $\hat{b}_{i,2} : T_i \rightarrow B_{i,2}$ for every agent i. Iterating this argument, we obtain for every $\tau_i \in T_i$ every agents' hierarchy of beliefs. These hierarchies of beliefs will also satisfy the consistency condition that we explained in the previous subsection. Thus, we obtain for each agent i a mapping $\hat{b}_i : T_i \rightarrow \mathcal{B}_i$.

We shall now clarify our terminology. We call τ_i agent i's "type" and T_i agent i's "type space." We call $\hat{\theta}_i(\tau_i)$ the "payoff type" of agent i, so that Θ_i is the set of agent i's payoff types. We call $\hat{\beta}_i(\tau_i)$ the "belief type" of agent i, so that $\Delta(T_{-i})$ is the set of agent i's belief types.

We shall occasionally break down large type spaces into smaller, self-contained subsets. We shall call these "belief closed subsets."

Definition 10.3 *Given a type space* $\mathcal{T} = \left(T_i, \hat{\theta}_i, \hat{\beta}_i\right)_{i \in I}$, *we shall call* $(\hat{T}_i)_{i \in I}$ *where* $\hat{T}_i \subseteq T_i$ *for each* $i \in I$ *a* belief closed subset of $\mathcal{T}$ *if for every* $i \in I$ *and* $\tau_i \in \hat{T}_i$ *we have* $\hat{\beta}_i(\tau_i)(\hat{T}_{-i}) = 1$.

Thus, when types are drawn from a belief closed subset, it is common certainty among agents that the types are in this belief closed subset; that is, everyone believes with probability 1 that all types are in this subset, and everyone believes with probability 1 that everyone believes with probability 1 that all types are in this subset, and so on.

10.3.3 Common Prior Type Spaces

In economics it has often been assumed that agents all agree about a prior over the type space and that their beliefs are derived from this common prior by Bayesian updating once agents have learned their own type.

Definition 10.4 *A* common prior type space *is a type space such that there is a probability measure* μ *on* $T_1 \times T_2 \times \ldots \times T_N$ *such that for every agent i and type* τ_i *the beliefs* $\hat{\beta}_i(\tau_i)$ *are equal to the conditional probability measure that one obtains by conditioning* μ *on* τ_i.

We are imprecise here because we don't say exactly what we mean by conditional probabilities in case the type spaces are infinite. This is of not much consequence as we shall restrict our attention to finite type spaces whenever we are considering common prior type spaces.

It is important to emphasize that the restrictive assumptions about beliefs that are often made in the classic theory of mechanism design, and that motivate our efforts in this section to develop a more general theory, are not implied by a common prior. Even with a common prior, these assumptions can be relaxed, and far more general models can be considered than have been traditionally. In other words, robust mechanism design does not require giving up the common prior assumption. If one wishes, one can study robust mechanism design maintaining the common prior assumption. We shall return to this point later.

It is instructive to study the structure of common prior type spaces in more detail. The following proposition shows for finite common prior type spaces that the conditional distribution of payoff types, conditional on all belief types, is a product distribution; that is, payoff types are conditionally independent. Note that if we know all agents' beliefs about the other agents' types, we don't necessarily yet know their payoff types. The assertion of the proposition is, however, that payoff types are independent, so that if one of the agents in addition to his beliefs also told us her payoff type, we would not learn anything new about the other agents' payoff types. Looking at things from this perspective is useful because it allows us to break down the revelation of types in a common prior type spaces into two steps: First, we need to know agents' belief types. In some settings, we can elicit these through a Crémer–McLean style construction. Second, once we know agents' belief types, we need to elicit their payoff types. But these payoff types are stochastically independent, and therefore the incentive problems that we face when trying to elicit the payoff types are similar to the incentive problems studied in the examples in Chapter 3. Note that the analysis of the example in Section 10.2 illustrates exactly this way of proceeding.

Proposition 10.1 *Suppose the type space is finite. Suppose beliefs are derived from a common prior μ with full support on the type space. Then for all $\beta \in \prod_{i\in I} \hat{\beta}(T_i)$ we have*

$$\mu((\theta_1, \theta_2, \ldots, \theta_N)|\beta) = \mu(\theta_1|\beta)\mu(\theta_2|\beta) \cdot \ldots \cdot \mu(\theta_N|\beta)$$

Proof

We first prove

$$\mu((\theta_1, \theta_2, \ldots, \theta_N)|\beta) = \mu(\theta_1|\beta)\mu(\theta_{-1}|\beta) \tag{10.6}$$

Suppose not. Then there are θ_1, θ_1' such that

$$\mu(\theta_{-1}|\beta, \theta_1) \neq \mu(\theta_{-1}|\beta, \theta_1') \tag{10.7}$$

Then there must be some $\tau, \tau' \in \mathcal{T}$ such that

$$\mu(\theta_{-1}|\tau) \neq \mu(\theta_{-1}|\tau') \quad \text{and} \quad \hat{\beta}(\tau) = \hat{\beta}(\tau') = \beta, \hat{\theta}_1(\tau_1) = \theta_1, \hat{\theta}_1(\tau_1') = \theta_1'. \tag{10.8}$$

But for every type vector for which player 1's belief type is given by some β_1, the conditional distribution of θ_{-1} conditional on τ is the one implied by β_1. Therefore, $\hat{\beta}_1(\tau_1) = \hat{\beta}_1(\tau_1')$ contradicts $\mu(\theta_{-1}|\tau) \neq \mu(\theta_{-1}|\tau')$. This proves (10.6). We can apply the same argument to prove that $\mu(\theta_{-1}|\beta) = \mu(\theta_2|\beta)\mu(\theta_{-1,2} \mid \beta)$. Iterating the argument, we obtain the proposition. □

To illustrate Proposition 10.1, we return to the example from Section 10.2. In that example, each player has two belief types. The types of player 1 corresponding to the two bottom rows in Figure 10.1 have identical belief types. There are four pairs of belief types to be considered. Conditional on those two pairs of belief types in which player 1's belief type corresponds to the first row of the table in Figure 10.1, the distribution of payoff types is deterministic; that is, each player has one payoff type with probability 1. Thus, the independence conclusion of Proposition 10.1 trivially holds. On the other hand, conditional on the other two pairs of belief types, the payoff type of player 2 is fixed, but the payoff type of player 1 is stochastic. Again, the independence conclusion of Proposition 10.1 is trivially satisfied. If we added a column to the table and introduced another type of player 2 that has the same belief type as one of the types of player 2 depicted in the table, then both players' belief types would be stochastic, and again they would obviously be independent.

We can use Proposition 10.1 to understand what precisely was special about the common prior type spaces that we considered earlier. When types are identical to payoff types and are independent, then the conditional distribution of payoff types is the same for all belief types. In fact, there is only one belief type. One could generalize this class of models slightly and construct a more general type space in which agents have some incomplete information about payoff irrelevant variables, so that there would be multiple belief types for at least some players, but still the conditional distribution of payoff types is the same for all belief types. If a type space satisfies the Crémer–McLean condition, then, by contrast, the conditional distribution of payoff types is different for every belief type, and moreover it is deterministic: It attaches probability 1 for each player to some payoff type. The example of Section 10.2 is intermediate between these two cases because conditional on some belief types the payoff type distribution is deterministic; and conditional on some others, it is stochastic. As discussed prior to the statement of Proposition 10.1, the proposition offers us a hint how we can deal with the more general case that Section 10.2 illustrates.

10.4 THE MECHANISM DESIGNER'S UNCERTAINTY

Type spaces are a simple way of describing agents' payoffs and beliefs. In mechanism design they are also part of the description of the mechanism designer's uncertainty. They describe the range of payoff and belief types that the mechanism designer thinks are possible. The starting point for this chapter was the observation that much of the literature on Bayesian mechanism design has focused on very special type spaces. The agenda of robustness can be interpreted as the agenda of considering mechanism design for larger type spaces. Of course, it is ambiguous what one means by a large type space. This section presents several formal definitions of what it means for a type space to be "large."

One large type space is particularly prominent in the literature. It is referred to as the "universal type space." In the universal type space, each agent i's type space, is the set $\Theta_i \times \mathcal{B}_i$—that is, the Cartesian product of the set of all payoff types of agent i and the set of all infinite hierarchies of beliefs of agent i. Recall that we constructed that set in Section 10.3.1. Now consider a type $\tau_i = (\theta_i, b_i)$ of player i. One can define this type's subjective beliefs in such a manner that the implied hierarchy of beliefs is b_i.[4] Let this belief on T_{-i} be denoted by $\tilde{\beta}_i(\tau_i)$. Then we can define the universal type space as follows:

Definition 10.5 *The* universal type space $\mathcal{T}^* \equiv (T_i^*, \hat{\theta}_i^*, \hat{\beta}_i^*)_{i \in I}$ *is defined by*

1. $T_i^* = \Theta_i \times \mathcal{B}_i$ *for all* $i \in I$.
2. $\hat{\theta}_i^*(\theta_i, b_i) = \theta_i$ *for all* $i \in I$ *and all* $(\theta_i, b_i) \in T_i^*$.
3. $\hat{\beta}_i^*(\theta_i, b_i) = \tilde{\beta}_i(\theta_i, b_i)$ *for all* $i \in I$ *and all* $(\theta_i, b_i) \in T_i^*$.

If one tries to study mechanism design on the universal type space, one encounters severe technical difficulties. One such difficulty is that even for finite mechanisms, the existence of Bayesian equilibria in mixed strategies is hard to establish if one considers the universal type space. This makes it attractive to consider smaller spaces than the universal type space that are still sufficiently large to capture a large variety of beliefs. Such a type space is the space of all finite types that we define next.[5] Informally speaking, a type is finite if it has a hierarchy of beliefs that can arise in a finite type space. The space of all finite types is itself, of course, infinite because there are infinitely many finite type spaces. We will show in Proposition

10.5, however, that every finite mechanism has at least one Bayesian equilibrium in mixed strategies on the space of all finite types.

Definition 10.6 *The* space of finite types $\mathcal{T}^+ \equiv (T_i^+, \hat{\theta}_i^+, \hat{\beta}_i^+)_{i\in I}$ *is defined by the following:*

1. $T_i^+ = \{(\theta_i, b_i) \in T_i^*$: *there are a finite type space* $\mathcal{T} = (T_i, \hat{\theta}_i, \hat{\beta}_i)_{i\in I}$ *and a* $\tau_i \in T_i$ *such that* $\hat{\theta}_i(\tau_i) = \theta_i$ *and* $\hat{\beta}_i(\tau_i) = b_i\}$.
2. $\hat{\theta}_i^+(\theta_i, b_i) = \theta_i$ *for all* $i \in I$ *and all* $(\theta_i, b_i) \in T_i^+$.
3. $\hat{\beta}_i^+(\theta_i, b_i) = \beta_i^*(\theta_i, b_i)$ *for all* $i \in I$ *and all* $(\theta_i, b_i) \in T_i^+$.

Our main objective is to consider type spaces that are "sufficiently large." For some results, if one examines the proof of these results, it is sufficient to assume that for certain probability 1 beliefs there are types who hold those beliefs. Specifically, for every player i and every vector of payoff types (θ_i, θ_{-i}), there must be a type of player i who has payoff type θ_i and believes with probability 1 that the other players have payoff types θ_{-i}. We shall say that type spaces with this property "have a large variety of certainties." The formal definition of this notion is as follows.

Definition 10.7 *A type space* $\mathcal{T} = (T_i, \hat{\theta}_i, \hat{\beta}_i)_{i\in I}$ has a large variety of certainties *if for all agents* $i \in I$ *and payoff types* $\theta_i \in \Theta_i, \theta_{-i} \in \Theta_{-i}$ *there are types* $\tau_i \in T_i, \tau_{-i} \in T_{-i}$ *such that* $\hat{\theta}_i(\tau_i) = \theta_i$, *and* $\hat{\beta}_i(\tau_i)$ *assigns probability 1 to* $\{\tau_{-i} \in T_{-i} : \hat{\theta}_{-i}(\tau_{-i}) = \theta_{-i}\}$.

The universal type space and the space of all finite type spaces obviously have large varieties of certainties. The universal type space, and the space of all finite types, do not have a common prior. By contrast, there are also common prior type spaces that have large varieties of certainties. For example, payoff types could be uniformly distributed, and realized payoff types could be common knowledge with probability 1.

10.5 MECHANISMS

We now recapitulate the definition of a mechanism:

Definition 10.8 *A* mechanism *is a list* $(S_1, S_2, \ldots, S_N, g)$ *where for each* $i \in I$ *the set* S_i *is a nonempty set of strategies and where* $g : S_1 \times S_2 \times \ldots \times S_N \to \Delta(A)$.

As before, direct mechanisms will play a special role.

Definition 10.9 *A mechanism* $(S_1, S_2, \ldots, S_N, g)$ *is a* direct mechanism *if* $S_i = T_i$ *for all* $i \in N$.

Note that in this definition we take the type space $\mathcal{T}$ as given and fixed. Of course, different type spaces will have different sets of direct mechanisms.

If the type space $\mathcal{T}$ is large, then a direct mechanism can be a very cumbersome object. Suppose, for example, the type space were the universal type space. Then a direct mechanism would invite agents to report not only their payoff types, but also their complete, infinite hierarchies of beliefs. This may not be feasible. We shall also consider a reduced version of a direct mechanism in which agents are only asked to report their payoff types, but not their infinite hierarchies of beliefs. There is no established terminology for such direct mechanisms. We shall refer to them in this chapter as "reduced direct mechanisms."

Definition 10.10 *A mechanism* $(S_1, S_2, \ldots, S_N, g)$ *is a* reduced direct mechanism *if* $S_i = \Theta_i$ *for all* $i \in N$.

In a reduced direct mechanism, agents simply report their payoff types. Reduced direct mechanisms might appeal because they are particularly simple mechanisms. As we discussed at the beginning of this chapter, implicitly the theory of robust mechanism design searches for optimal mechanisms that are not only robust but also simple. Another way of formalizing the simplicity requirement is to require that the strategy spaces be finite. This is a technical, but sometimes very effective, way of restricting the complexity of the mechanisms that one is willing to consider.

10.6 BAYESIAN EQUILIBRIA AND THE REVELATION PRINCIPLE

The mechanism designer is concerned with type spaces because she wants to analyze how agents will choose strategies in the mechanism that she proposes. In this section we shall postulate a mechanism designer who has identified the type space $\mathcal{T}$ that seems relevant to her. She also believes that the solution concept that she should use to predict agents' behavior is that of Bayesian equilibrium. We shall also consider certain refinements of Bayesian equilibrium that are potentially relevant in the current context.

We now define the standard concept of a Bayesian equilibrium, adapted to our setting.

Definition 10.11 *A* Bayesian equilibrium *of a mechanism* $(S_1, S_2, \ldots, S_N, g)$ *is a list of strategies* $\sigma^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$, *where for each* $i \in I$: $\sigma_i^* : T_i \to \Delta(S_i)$, *such that for every player* i *for every type* $\tau_i \in T_i$ *the choice* $\sigma_i^*(\tau_i)$ *maximizes type* τ_i*'s expected utility using type* τ_i*'s beliefs,* $\hat{\beta}_i(\tau_i)$.

This definition is slightly informal, hopefully with no danger of creating any important ambiguity. We are slightly informal because we have not explicitly introduced the measure theoretic structures that would be necessary to offer a completely formal definition.

It is immediate that the revelation principle applies. In fact, the revelation principle is simply a property of Bayesian equilibria, and it is not necessarily related to mechanism design at all. For the sake of clarity, we state the revelation principle explicitly:

Proposition 10.2 *Let $\sigma^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ be a Bayesian equilibrium of a mechanism $(S_1, S_2, \ldots, S_N, g)$. Construct a direct mechanism $(\tilde{S}_1, \tilde{S}_2, \ldots, \tilde{S}_N, \tilde{g})$ as follows:*

1. *$\tilde{S}_i = T_i$.*
2. *$\tilde{g}(\tau_1, \tau_2, \ldots, \tau_N) = g(\sigma_1^*(\tau_1), \sigma_2^*(\tau_2), \ldots, \sigma_N^*(\tau_N))$ for all $(\tau_1, \tau_2, \ldots, \tau_N) \in T$.*

Then in this direct mechanism the strategies given by $\tilde{\sigma}_i^*(\tau_i) = \tau_i$ for all $i \in N$ and $\tau_i \in T_i$ form a Bayesian equilibrium. Moreover, this equilibrium is outcome equivalent to the Bayesian equilibrium of the original mechanism; that is, for every vector of types, the outcome implied by the equilibrium is the same in both mechanisms and their equilibria.

The proof of Proposition 10.2 is the same as the proof of the standard revelation principle and therefore omitted. Note that the notation $g(\sigma_1^*(\tau_1), \sigma_2^*(\tau_2), \ldots, \sigma_N^*(\tau_N))$ that we use in Proposition 10.2 is slightly sloppy. If some of the strategies in brackets are mixed strategies, we simply mean by this notation the lottery over A that results from putting together the mixed strategies of the players with the (potentially also random) outcomes determined by g.

As we mentioned before, for large type spaces the use of direct mechanisms is very unrealistic. We view Proposition 10.2 therefore as a device for the modeler rather than as a result that can be used to construct real mechanisms. The use of reduced direct mechanisms is maybe more realistic. We therefore develop next a revelation principle in which direct mechanisms are replaced by reduced direct mechanisms. To obtain such a result, we need to focus on the case that a mechanism has a type of Bayesian equilibrium that has received particular attention in the literature on robust mechanism design.

Definition 10.12 *A Bayesian equilibrium $\sigma^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ of a mechanism $(S_1, S_2, \ldots, S_N, g)$ is* belief-independent *if for every agent $i \in I$ and every pair of types $\tau_i, \tau_i' \in T_i$ we have $\sigma_i^*(\tau_i) = \sigma_i^*(\tau_i')$ whenever $\hat{\theta}_i(\tau_i) = \hat{\theta}_i(\tau_i')$.*

Thus, a belief-independent Bayesian equilibrium is one in which agents' strategy choice depends only on their payoff types, not on their belief types. Most mechanisms don't have belief-independent Bayesian equilibria. Thus, implicitly, when we restrict attention to belief-independent equilibria, we also restrict attention to a subset of mechanisms.

It may be useful to think of belief-independent Bayesian equilibria as simple equilibria. They are at a minimum easy to implement, because agents need not consider their beliefs to choose optimally.

We now state a revelation principle for belief-independent Bayesian equilibria that refers to reduced direct mechanisms.

Proposition 10.3 *Let $\sigma^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ be a belief-independent Bayesian equilibrium of a mechanism $(S_1, S_2, \ldots, S_N, g)$. Construct a reduced direct mechanism $(\tilde{S}_1, \tilde{S}_2, \ldots, \tilde{S}_N, \tilde{g})$ as follows:*

1. *$\tilde{S}_i = \Theta_i$ for all $i \in N$.*
2. *$\tilde{g}(\theta_1, \theta_2, \ldots, \theta_N) = g(\sigma_1^*(\tau_1), \sigma_2^*(\tau_2), \ldots, \sigma_N^*(\tau_N))$ for all $(\theta_1, \theta_2, \ldots, \theta_N) \in \Theta$, where for all $i \in N$ τ_i is some type of player i such that $\hat{\theta}_i(\tau_i) = \theta_i$.*

Then in this direct mechanism the strategies given by $\tilde{\sigma}_i^*(\tau_i) = \hat{\theta}_i(\tau_i)$ for all $i \in N$ and $\tau_i \in T_i$ form a Bayesian equilibrium. Moreover, this equilibrium is outcome equivalent to the Bayesian equilibrium of the original mechanism; that is, for every vector of types, the outcome implied by the equilibrium is the same in both mechanisms and their equilibria.

Note that in item 2 in Proposition 10.3, any choice of $(\tau_1, \tau_2, \ldots, \tau_N)$ will lead to the same outcome precisely because the Bayesian equilibrium is belief-independent. The proof of Proposition 10.3 is the same as the proof of the standard revelation principle and is therefore omitted.

Thus, if equilibria are belief-independent, the direct mechanism that implements them is also simple: It only requires agents to report their payoff types. We obtain a much stronger result than Proposition 10.3 if we assume that the type space has a large variety of certainties. We shall use the following definition:

Definition 10.13 *A Bayesian equilibrium $\sigma^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ of a mechanism $(S_1, S_2, \ldots, S_N, g)$ is an* ex post Bayesian equilibrium *if for every agent $i \in I$, $\tau_i \in T_i$, and $\tau_{-i} \in T_{-i}$, the choice $\sigma_i^*(\tau_i)$ maximizes type τ_i's expected utility if τ_i's beliefs attach probability 1 to the other players' types being τ_{-i}.*

Proposition 10.4 *Let $\sigma^* = (\sigma_1^*, \sigma_2^*, \ldots, \sigma_N^*)$ be a belief-independent Bayesian equilibrium of a mechanism $(S_1, S_2, \ldots, S_N, g)$ for a type space with a large variety of* *certainties. Then σ^* is an ex post Bayesian equilibrium of this mechanism. Moreover, if we construct a reduced direct mechanism $(\tilde{S}_1, \tilde{S}_2, \ldots, \tilde{S}_N, \tilde{g})$ as follows:*

1. $\tilde{S}_i = \Theta_i$ *for all* $i \in N$.
2. $\tilde{g}(\theta_1, \theta_2, \ldots, \theta_N) = g(\sigma_1^*(\tau_1), \sigma_2^*(\tau_2), \ldots, \sigma_N^*(\tau_N))$ *for all* $(\theta_1, \theta_2, \ldots, \theta_N) \in \Theta$, *where, for all* $i \in N$, τ_i *is some type of player i such that* $\hat{\theta}_i(\tau_i) = \theta_i$.

then in this reduced direct mechanism the strategies given by $\tilde{\sigma}_i^*(\tau_i) = \hat{\theta}_i(\tau_i)$ for all $i \in N$ and $\tau_i \in T_i$ form an ex post Bayesian equilibrium. Moreover, this equilibrium is outcome equivalent to the Bayesian equilibrium of the original mechanism; that is, for every vector of types, the outcome implied by the equilibrium is the same in both mechanisms and their equilibria.

The reason that the original Bayesian equilibrium, as well as truthful strategies in the reduced direct mechanism, are ex post Bayesian equilibria is that the given strategy, or telling the truth in the direct mechanism, must be optimal in particular for types that believe with probability 1 that the vector of other players' payoff types is $\theta_{-i} \in \Theta_{-i}$, and this must be true for every $\theta_{-i} \in \Theta_{-i}$. Thus, even if a player learns ex post the types of all other players, the player must have no incentive to change his strategy. We omit the formalization of this argument.

Propositions 10.3 and 10.4 describe versions of the revelation principle that allow three important simplifications: First, we can choose a simple mechanism, namely a reduced direct mechanism. Second, we can suggest to agents a simple equilibrium: truth telling. Finally, the suggested equilibrium can be implemented by the agents in a simple way, as they need not consider their beliefs about other agents' strategies.

Finally, the following simple technical observation indicates why we consider the space of finite types.

Proposition 10.5 *Every finite mechanism has at least one Bayesian equilibrium in mixed strategies on the space of finite types.*

Proof[6]

We outline the steps of this proof in a slightly informal manner. In problem (a) for this chapter, you are asked to fill in the gaps and give a completely formal proof.

By definition of the space $\mathcal{T}^+$ of finite types, there is for every type vector in this space at least one finite belief closed subset to which this type vector belongs. If a type vector belongs to multiple finite belief closed subsets, it also belongs to the intersection of these belief closed subsets, and moreover this intersection also forms a belief closed subset. This intersection is the smallest belief closed subset to which the type vector belongs. A strategy vector forms a Bayesian equilibrium if and only if for every type vector it is a Bayesian equilibrium on the smallest belief closed subset to which this type vector belongs.

We now proceed by induction. Consider first all type vectors that belong to a belief closed subset of cardinality 1. On such a type space, the game defined by the mechanism is a finite game of complete information and, therefore, by Nash's existence theorem, has at least one Nash equilibrium in mixed strategies. We assign these equilibrium strategies to the types that make up the given type vector.

Now suppose we had defined Bayesian Nash equilibrium strategies for all types such that the smallest belief closed subset to which they belong has cardinality of at most $k \in \mathbb{N}$. Consider any type vector for which the smallest belief closed subset to which it belongs has cardinality $k + 1$. That subset may contain types that belong to belief closed subsets with cardinality less than $k + 1$. For those types we have already defined their equilibrium choices. We hold these choices fixed, and we consider the game played by the remaining type vectors. Nash's existence theorem, now applied to the game played by these remaining type vectors, implies that we can find an equilibrium in mixed strategies of this game. We assign to each type the strategy that he plays in this equilibrium. We can continue this construction ad infinitum, and we obtain a Bayesian equilibrium on the space of all finite types. □

10.7 WHAT CAN BE IMPLEMENTED?

Now consider a given type space, a mechanism, and a Bayesian equilibrium of that mechanism. The mechanism together with the equilibrium determine for each vector of types a corresponding outcome. When asking what we can implement in a robust setting, we can ask about characteristics of the mappings of type vectors into outcomes that can be derived from arbitrary mechanisms and their Bayesian equilibria. Another approach suggests itself sometimes. Because the mechanism designer's objectives may be independent of agents' beliefs, we might focus on the mapping that assigns to each vector of payoff types a set of outcomes. We obtain this set including all outcomes that are possible for the given vector of payoff types under some vector of belief types. Here, we discuss both approaches. Sections 10.7.1 and 10.7.2 are concerned with the first approach, and Section 10.7.3 is concerned with the second approach.

10.7.1 Belief Revelation

Consider first the mapping of types into outcomes, and focus on some agent i, and only on those types of agent i that have some particular belief type $\beta_i(\tau_i)$. These types thus only differ in their payoff type. For each of these types, we can determine the probability distributions over outcomes that they expect in equilibrium. The mapping from this particular subset of types of player i into probability distributions over outcomes has to satisfy the standard conditions for Bayesian incentive compatibility. Nothing additional needs to be said here about these conditions.

When we want to characterize the mappings from types into outcomes that can be implemented by some mechanism and some Bayesian equilibrium of that mechanism, we also obtain incentive compatibility restrictions that apply to agents with different beliefs. The further restrictions that follow from these incentive compatibility constraints are hard to characterize. To make progress, we focus on the case that agents have additively separable and utility functions that are linear in money—that is, the framework of earlier chapters, such as Chapters 6 and 7. Recall that in those chapters the set of alternatives denoted by A is not identical to the set of possible outcomes of a mechanism. Rather, the possible outcomes of a mechanism are a vector $(a, t_1, t_2, \ldots, t_N)$, where t_i is a transfer paid by agent i. Unlike in Chapters 6 and 7, we now allow for interdependent types, so that the utility functions are of the form $u_i(a, \theta) - t_i$. We obtain the following generalization of the theorem of Crémer and McLean that we presented earlier as Proposition 6.4.

Proposition 10.6 *Consider a finite type space $\mathcal{T}$ such that:*

(i) for every player i, the set of beliefs that player i may hold:

$$\left\{\hat{\beta}_i(\tau_i) : \tau_i \in T_i\right\}$$

has the property that no vector in this set can be written as a convex combination of all the other vectors,

and consider a direct mechanism such that:

(ii) for every player i, for every pair of types $\tau_i, \tau_i' \in T_i$ such that $\hat{\beta}_i(\tau_i) = \hat{\beta}_i(\tau_i')$, type τ_i has no incentive to pretend being type τ_i'.

Then there is another direct mechanism in which truth telling is a Bayesian equilibrium, such that:

(iii) for every type vector $\tau \in T$ the alternative in A that is implemented by the former mechanism is the same as the alternative that is implemented by the latter mechanism;
(iv) for all players i and all types $\tau_i \in T_i$, the interim expected payments of type τ_i in the two mechanisms are the same.

Condition (i) in this result is very similar to the Crémer–McLean condition that we studied in Chapter 6, but it is *not* the same condition. The Crémer–McLean condition is always violated if two different types have the same beliefs. That is not true for our condition (i). The set to which condition (i) refers is the collection of beliefs that any type of agent *i* may hold. By definition, no two elements of this set will be identical. But, of course, it may be that any one belief in this set is held by two or more different types of player *i*.

We don't give a formal proof of Proposition 10.6, but only sketch the argument. The precise formalization only repeats arguments that we have seen earlier. As in the proof of Crémer and McLean's result, we can use Farkas's lemma to construct for each player a payment scheme that gives the player strict incentives to honestly report his beliefs and where dishonesty is punished so severely that no change in the alternative implemented can compensate for the severity of the punishment. Moreover, if agents report their beliefs truthfully, their expected payment is zero.

As the expected payment is the same for all types that agent *i* may report, provided that they have the same beliefs, agent *i*'s secondary consideration will be which payoff type to report. By condition (ii), each agent has an incentive to report their payoff type truthfully, given that they reveal their beliefs truthfully. Thus, truthful reporting is a Bayesian equilibrium.

In Crémer and McLean's result, because no two types have the same beliefs, when agents report their beliefs, they automatically also report their types. In our result, because we have not ruled out that two types have different payoff types but identical belief types, we have to add to Crémer and McLean's original result a condition that ensures that agents don't have an incentive to misreport their payoff type.

The main message of this result is that when checking incentive compatibility, if we are only concerned with agents' interim expected utility, it is sufficient to check incentive compatibility with respect to payoff types. We don't have to check incentive compatibility with respect to belief types. We can always rearrange payments so that also incentive compatibility with respect to belief types is ensured and so that agents' interim expected utility does not change.

The role of this result for the design of optimal mechanisms is limited, for two reasons. The first is that the mechanism designer might not share the agents' interim expected beliefs, in particular if these are not derived from a prior that is common to all agents and to the mechanism designer. If the mechanism designer does not share the agents' interim expected beliefs, then rearrangements of payments that don't affect agents' interim expected payments may change the mechanism designer's expected value of those payments. Therefore, Proposition 10.6 will be invoked below when we discuss mechanism design in the case of a common prior, but not in the case of subjective, potentially inconsistent beliefs.

The second reason why Proposition 10.6 has only a limited role in the design of optimal mechanisms is that, even with a common prior assumption, it is by no means a "full extraction of the surplus" result, unlike Crémer and McLean's original result. Consider the case of a single object auction in which payoff types are private values. Suppose the mechanism designer wants to allocate the object to the bidder with the highest value and then extract from this bidder his value. If different types may have different private values but identical beliefs, then whatever rearrangement of payments is chosen, the types with high private values will always have an incentive to pretend to be types with the same beliefs but lower private values. Incentive constraints for agents with identical beliefs but different payoff types limit the auctioneer's ability to "extract the surplus."

Even so, Proposition 10.6 is a very powerful result that may significantly reduce the number of incentive constraints that we have to take into account in any given mechanism design problem. The mechanisms that are constructed to prove Proposition 10.6 may not be simple and thus may not be satisfactory if simplicity is included among the desiderata of robust mechanisms, a possibility that was mentioned in the introduction to this chapter. However, nothing in our formalism rules out the construction described here. Thus, Proposition 10.6 emphasizes the need to add to our formalism some precisely defined requirement of simplicity of mechanisms.

10.7.2 Betting

By a theorem of Milgrom and Stokey (1982), two rational risk-neutral agents whose beliefs are based on the same prior, and whose beliefs differ only because agents have received different information, will not bet with each other. The reason is that these agents have strictly opposed preferences over money, and, moreover, that each agent infers from the other agent's acceptance of the bet that the other agent holds information that makes the bet attractive to him or her and therefore unattractive to the original agent him- or herself. This theorem does not hold if agents' beliefs are not derived from the same prior. Intuitively, without a common prior, there is no guarantee that an agent would interpret the information that another agent holds in the same way as that agent. Thus, even if some information of yours may make a bet look attractive to you and if I infer from your acceptance of a bet that you have that particular information, it may be that I would interpret that information differently from you and that therefore the bet looks attractive to both you and me.

This possibility of betting among agents with beliefs not derived from a common prior ("inconsistent beliefs") affects the theory of mechanism design when agents' preferences are quasi-linear. This is what we demonstrate in this section in Proposition 10.7. This Proposition looks formidable, but what it expresses is almost trivial. I just could not think of a simpler way of writing the result down.

Because we have only modeled the agents' beliefs, but not the mechanism designer's beliefs, we cannot consider bets among the mechanism designer and the agents, but only bets that the agents make with each other. Within the confines of our model, the only event that two agents can bet on concern the types of agents. Agents' types is the only uncertain event in our model, and therefore bets have to be on types. If two agents were to bet on an event that is made public only through their own reports, we would obviously face new incentive issues. Therefore, we only consider bets among two agents about some other agents' types. This is why Proposition 10.7 requires the number of agents to be at least 3. The result then describes how agents i and j can bet on the event that the other agents' types are in some set $\widehat{T}_{-i,j}$. The bet is that agent i pays a dollar to agent j if the other agents' types are in $\widehat{T}_{-i,j}$ and that agent j pays c_i dollars to agent i otherwise. For this to be an acceptable bet, it is not sufficient that agents i and j have different subjective probabilities that the other agents' types are in $\widehat{T}_{-i,j}$. What is needed is that such inconsistencies exist once each agent has conditioned on the event that the other agent accepts the bet. This is the meaning of assumption (ii) in the proposition below, where we have introduced $\varepsilon > 0$ such that the inconsistency of beliefs is at least ε. If the beliefs are sufficiently inconsistent, we can pick c_i and c_j such that the bet described above is attractive for both sides.

We can add the payments resulting from these bets to any arbitrary Bayesian incentive-compatible mechanism and obtain another Bayesian incentive-compatible mechanism. This is what the following proposition says.

Proposition 10.7 *Suppose that $N \geq 3$ and that there are $i, j \in I$, $p, \varepsilon \in (0,1)$ and sets $\widehat{T}_i \subseteq T_i$, $\widehat{T}_j \subseteq T_j$, $\widehat{T}_{-i,j} \subsetneq T_{-i,j}$ such that:*

(i) $\hat{\beta}_i(\tau_i)[\hat{T}_j] > 0$ *for all* $\tau_i \in T_i$,
$\hat{\beta}_j(\tau_j)[\hat{T}_i] > 0$ *for all* $\tau_j \in T_j$.

(ii) $\hat{\beta}_i(\tau_i)[\hat{T}_{-i,j} \mid \hat{T}_j] \leq p - \varepsilon$ *for all* $\tau_i \in \hat{T}_i$,
$\hat{\beta}_i(\tau_i)[\hat{T}_{-i,j} \mid \hat{T}_j] \geq p + \varepsilon$ *for all* $\tau_i \notin \hat{T}_i$,
$\hat{\beta}_j(\tau_j)[\hat{T}_{-i,j} \mid \hat{T}_i] \geq p + \varepsilon$ *for all* $\tau_j \in \hat{T}_j$,
$\hat{\beta}_j(\tau_j)[\hat{T}_{-i,j} \mid \hat{T}_i] \leq p - \varepsilon$ *for all* $\tau_j \notin \hat{T}_j$.

Then, if (T, q, t) is a Bayesian incentive-compatible direct mechanism, the following direct mechanism $(T, \tilde{q}, \tilde{t})$ is also Bayesian incentive-compatible:

(iii) $\tilde{q} = q$;
(iv) $\tilde{t}_k = t_k$ *for all* $k \neq i, j$;

(v) If $\tau_i \in \hat{T}_i$ *and* $\tau_j \in \hat{T}_j$*:*

$$\begin{aligned}
\tilde{t}_i(\tau) &= t_i(\tau) - 1 && \text{if } \tau_{-i,j} \in \hat{T}_{-i,j},\\
\tilde{t}_i(\tau) &= t_i(\tau) + c_i && \text{if } \tau_{-i,j} \notin \hat{T}_{-i,j},\\
\tilde{t}_j(\tau) &= t_j(\tau) + 1 && \text{if } \tau_{-i,j} \in \hat{T}_{-i,j},\\
\tilde{t}_j(\tau) &= t_j(\tau) - c_i && \text{if } \tau_{-i,j} \notin \hat{T}_{-i,j}.
\end{aligned}$$

(vi) If $\tau_i \notin \hat{T}_i$ *or* $\tau_j \notin \hat{T}_j$*:*

$$\tilde{t}_i(\tau) = t_i(\tau) \quad \textit{and} \quad \tilde{t}_j(\tau) = t_j(\tau),$$

where $\frac{p-\varepsilon}{1-(p-\varepsilon)} < c_i < \frac{p+\varepsilon}{1-(p+\varepsilon)}$.

We don't prove the result. A simple calculation shows that the bet by itself, disregarding the rest of the mechanism, is attractive to types in $\widehat{T}_i$ and $\widehat{T}_j$, and not to other types, if c_i satisfies the inequalities described in the last line of the proposition. In addition to the bet, the mechanism that we are amending by introducing bets is unchanged. Agents will choose in this mechanism as they did before because the quasi-linear form of utility implies that incentives to bet don't interact at all with the incentives in the original mechanism.

Observe that in the mechanism $(T, \tilde{q}, \tilde{t})$ the size of the bets is chosen by the mechanism designer, not by the agents. The mechanism offers one and only one opportunity of entering a bet. As explained before, the bet is that agent i pays agent j one dollar if the other agents' types are in $\hat{T}_{-i,j}$ and that agent j pays agent i c_i dollars otherwise. This formulation is chosen here for simplicity. Observe that linearity of agents' utility in money implies that the same mechanism would be incentive-compatible if the payments were replaced by $k \cdot 1$ dollar and $k \cdot c_i$ dollars, where k is a positive constant. In particular, by letting k tend to infinity, we can make the payments resulting from the bet arbitrarily large. Agents with inconsistent beliefs are willing to enter arbitrarily large bets.

How frequently are the conditions of Proposition 10.7 satisfied? Note that the conditions are, for example, satisfied if two agents i and j regard the other agents' types as independent of the vector of their own types, but hold different beliefs about the other agents' types. In any large type space, this will typically occur unless there is a common prior.

This implications of Proposition 10.7 for optimal mechanism design with quasi-linear utility functions and subjective, inconsistent beliefs are significant. We won't anticipate them here, but shortly explain them in Section 10.10.

10.7.3 Equilibrium Outcomes and Payoff Types

We now consider the second perspective on the question "What can be implemented?" In this perspective the focus is on the correspondences mapping payoff types into sets of outcomes that a mechanism and a Bayesian equilibrium of that mechanism induce. For every payoff type vector θ, let $F(\theta)$ be the corresponding set of outcomes. Unlike in the discussion above, we consider not only environments in which agents' utility is quasi-linear, but also general environments. It is hard to obtain results for the case that there are θ such that $F(\theta)$ has more than one element. But if for every θ the set $F(\theta)$ is a singleton, the following simple result holds.

Proposition 10.8 *Suppose the type space has a large variety of certainties, and suppose that the mechanism designer has constructed a mechanism and a Bayesian equilibrium of this mechanism on the type space such that for every $\theta \in \Theta$ the set $F(\theta)$ contains exactly one element. Then in the reduced direct mechanism in which agents report their payoff type, and then the unique alternative in $F(\theta)$ is implemented, truth telling is an ex post Bayesian equilibrium.*

Proof

Consider some $i \in I$ and some $\theta \in \Theta$. Consider
the type $\tau_i \in T_i$ who attaches probability 1 to the event that the other players' payoff type is θ_{-i}. Because the type space has a large variety of certainties, such a type exists. This type obtains in equilibrium payoff $u_i(F(\theta), \theta)$. By choosing the strategy of some other type τ_i' with payoff type $\theta_i' \neq \theta_i$, but with the same beliefs as type θ_i (such a type exists because the type space has a large variety of certainties), player i can obtain payoff $u_i(F(\theta_i', \theta_{-i}), \theta)$. Because we are considering a Bayesian equilibrium, we must have $u_i(F(\theta), \theta) \geq u_i(F(\theta_i', \theta_{-i}), \theta)$. This inequality is true for all $\theta \in \Theta$ and all $\theta_i' \in \Theta_i, \theta_i' \neq \theta_i$. Therefore, truth telling is an ex post Bayesian equilibrium. □

Proposition 10.8 shows that if we want to implement a unique outcome for every payoff type vector, we can restrict out attention to (a) mechanisms that are "simple" in the sense of being reduced direct mechanisms and (b) equilibria that are simple, in the sense of being ex post Bayesian equilibria.

A result that is related to Proposition 10.8 holds in the quasi-linear framework of earlier chapters, such as Chapters 6 and 7. We use the notation of those chapters. Suppose we are given a type space, a mechanism, and a Bayesian equilibrium of the mechanism on the type space. We denote by $\hat{F}(\theta)$ the projection of $F(\theta)$ into A, that is, $\hat{F}(\theta) = \{a \in A : (a, t_1, t_2, \ldots, t_N) \in F(\theta) \text{ for some } (t_1, t_2 \ldots, t_N) \in \mathbb{R}^N\}$. This is the set of alternatives, with the transfer components omitted, that may be chosen in equilibrium for the given mechanism if agents play the given Bayesian equilibrium.

For the following result, $\hat{F}(\theta)$ is assumed to be a singleton for every θ. Note that this is weaker than requiring that $F(\theta)$ itself is a singleton. Even if $\hat{F}(\theta)$ is a singleton, $F(\theta)$ may not be, because a variety of transfers may be combined with the single element of $\hat{F}(\theta)$.

Proposition 10.9 *Suppose the type space has a large variety of certainties, and suppose that the mechanism designer has constructed a mechanism and a Bayesian equilibrium of this mechanism on the type space such that for every $\theta \in \Theta$ the set $\hat{F}(\theta)$ contains exactly one element. Then there is a reduced direct mechanism in which agents report their payoff type, the alternative a that is implemented if the reported type vector is θ is the unique element of $\hat{F}(\theta)$, and truth telling is an ex post Bayesian equilibrium.*

Proof

Consider some $i \in I$ and some $\theta \in \Theta$. The alternative attached to θ is obviously $\hat{F}(\theta)$. To construct the transfer payment by agent i, we pick a type $\tau_i \in T_i$ who attaches probability 1 to the event that the other players' payoff type is θ_{-i}, and we set t_i equal to the equilibrium expected transfer payment of that type. By repeating this construction for every $i \in I$ and $\theta \in \Theta$, we complete the construction of the reduced direct mechanism. The same argument as in the proof of Proposition 10.8 then implies that truth telling is an ex post Bayesian equilibrium. □

Note that Proposition 10.9 is silent about the transfers in the reduced direct mechanism. The proof shows that each agent i's transfer payment in the reduced direct mechanism when the payoff vector is θ is a transfer payment that agent i makes in the original mechanism for some payoff and belief type, where the payoff type can be taken to be θ_i. But that is all that can be said.

10.8 ROBUST MECHANISM DESIGN WITH A COMMON PRIOR

So far we have given a description of a mechanism designer who constructs a mechanism and then analyzes Bayesian equilibria of that mechanism on large type spaces, without considering how the mechanism designer chooses among these mechanisms. We now turn to the mechanism designer's choice problem. It turns out that for the solution to this choice problem it is crucial whether or not there is a common prior. In this section we deal with the case of a common prior, and in the next sections we deal with the case without a common prior.

The common prior assumption is important because it allows us to treat the mechanism designer as a Bayesian decision maker whose belief is the same as the common prior that the agents share with each other. This is not a logical consequence of the common prior assumption, but it is a very natural way of proceeding once the common prior assumption is made. We thus assume in this section that the mechanism designer is a Bayesian decision maker, maximizing an objective such as expected revenue, or expected welfare, using as her beliefs the agents' common prior.

A second distinction that matters is whether preferences are quasi-linear or whether utility is not transferrable. This is important because, as we showed in the previous section, if preferences are quasi-linear, we can extract players' beliefs at no cost, and we only have to take care of payoff type incentive constraints. Moreover, Proposition 10.1 shows that once we have elicited belief types, conditional on belief types, payoff types are independent. Proposition 10.1 assumes a finite type space with a full support prior. In this section, we restrict our attention to this case. An optimal mechanism can then be obtained as follows: We first construct for any given belief type an optimal mechanism using the methods that were illustrated in Chapter 3, and then we amend this mechanism by providing agents with incentives to reveal their beliefs using Crémer and McLean's construction as we described in Proposition 10.6. This was illustrated in the example in Section 10.2 where conditional on belief types, the only incentive constraints that an optimal mechanism had to take care of were those of player 1, for the two payoff types with identical beliefs which correspond to the two bottom rows in Figure 10.1.

The construction that we described can be applied in other auction settings if one is interested in expected revenue maximizing auctions. This is illustrated in Farinha Luz (2013). It can also, in principle, be used to study welfare-maximizing public goods mechanisms or welfare-maximizing bilateral trade mechanisms, although in these cases the budget balance conditions that we imposed earlier need careful study.

Note why it is important for the arguments just described that there is a common prior. The reason is that the Crémer–McLean construction leaves the mechanism designer's expected revenue unchanged only if the mechanism designer shares the agents' beliefs. We discussed this point already in Subsection 10.7.1.

In conclusion, remarkably, for type spaces with common priors, if preferences are quasi-linear, we have found a method that allows us to construct optimal mechanisms even if the type space is finite but large. The mechanisms that we obtain are not necessarily simple, however. They deal with a large set of circumstances by incorporating many bells and whistles that address these circumstances. As we indicated in the introduction to this chapter, this is probably not what the theory of robust mechanism design is looking for. We therefore interpret these results as indicating that the formalism that we have been using is lacking a notion of simplicity and is lacking a constraint that limits the complexity of the mechanisms that we consider. Therefore, this is a possible direction for future research.

10.9 ROBUST MECHANISM DESIGN WITHOUT A COMMON PRIOR

10.9.1 The Mechanism Designer's Objectives

We now turn to the mechanism designer's choice problem if there is no common prior. Even if there is no *common* prior, we could attribute to the mechanism designer a prior that is just hers and not shared by the other agents. Without a relation between the agents' beliefs and the mechanism designer's prior, it is unlikely that this approach would yield interesting results. In particular, as we have emphasized, the construction described in the previous section would not apply. Instead, in this section we shall take a "prior-free" approach to the mechanism designer's choice problem; that is, we shall study what can be said without attributing a specific prior belief to the mechanism designer.

Without a prior, we can no longer focus on the mechanism designer's maximization of the ex ante expected value of some objective function. We therefore have to consider more carefully the mechanism designer's choice criterion. First, we introduce an abstract formalization of the mechanism designer's objective. For simplicity, we refer to this objective as "welfare," even though our formalism allows the objective to be taken from a large class of functions, some of which cannot reasonably be interpreted as "welfare" in the conventional sense. For example, our analysis will cover the case that the mechanism designer seeks to maximize her own revenue. Because our definition will be general, it might at first sight appear a little abstract. We shall use examples to make it more concrete.

Definition 10.14 *A welfare function maps every type space $\mathcal{T}$, direct mechanism M, and type vector $\tau \in T$ into a vector $w(\mathcal{T}, M, \tau) \in \mathbb{R}^m$, where m is some natural number.*

The interpretation of $w(\mathcal{T}, M, \tau) \in \mathbb{R}^m$ is that it equals welfare when the type space is $\mathcal{T}$, when the direct mechanism M is played, where we assume that all agents report their types truthfully, and when the agents' type vector is τ. The welfare function is real-valued if $m = 1$, and it is more generally vector-valued if $m \geq 2$. We allow $m \geq 2$ so that we can capture situations in which the mechanism designer's notion of welfare allows only a partial order of potential outcomes of the mechanism. We endow the space $\mathbb{R}^m$ with the partial order according to which one vector is larger than another if and only if it is at least as large as the other in every component, and in some component larger. We denote this order with the ordinary "$>$" sign. We shall also use the "$\geq$" sign. By this we mean that $>$ is true or that the two vectors are equal.

To illustrate Definition 10.14, we define the three prominent cases that are all captured by the definition.

Definition 10.15

1. *The following welfare function is called "ex post Pareto welfare":*

   $$w(\mathcal{T}, M, \tau) = (u_i(g(\tau))_{i \in I}.$$

   (For this welfare function we have $m = N$.)
2. *The following welfare function is called "interim Pareto welfare":*

   $$w(\mathcal{T}, M, \tau) = \left(\int_{T_{-i}} u_i(g(\tau))\, d\hat{\beta}_i(\tau_i) \right)_{i \in I}.$$

   (For this welfare function we have $m = N$.)
3. *In the case of additively separable, quasi-linear utility functions, the following welfare function is called "revenue":*

   $$w(\mathcal{T}, M, \tau) = \sum_{i \in I} t_i(g(\tau)),$$

   where we write $t_i(g(\tau))$ for the transfer that agent i pays to the mechanism designer if the outcome is $g(\tau)$. (For this welfare function we have $m = 1$.)

10.9.2 Undominated Mechanisms

We are now ready to formulate the mechanism designer's decision problem. To recapitulate, the mechanism designer has in mind a large type space that describes agents' incomplete information. The extent to which the mechanism designer's welfare objective will be achieved by any particular mechanism and a Bayesian equilibrium of that mechanism will potentially be different for different realized type vectors. What remains to be discussed is how the mechanism designer handles her own uncertainty about which type vector will be realized. We seek results that are independent of the mechanism designer's belief. The following definition shows how one can order mechanisms and their equilibria without specifying the mechanism designer's belief.

Definition 10.16 *The mechanism designer* ranks one mechanism and its Bayesian equilibrium above another mechanism and its Bayesian equilibrium *if for every vector of types $\tau \in T$ the welfare implied by the former equilibrium is at least as large as the welfare implied by the second equilibrium, and sometimes larger. We then also say that the former mechanism and its Bayesian equilibrium* dominates *the latter mechanism and its Bayesian equilibrium.*

If the mechanism designer ranks one mechanism above another, then she will prefer the former mechanism over the latter irrespective of what her beliefs about type vectors are, provided that she is not willing to ignore any type vector. This latter provision is just a reflection of the assumption that the type space precisely delineates the extent of the mechanism designer's uncertainty.

We have constructed a possibly incomplete ordering of mechanisms. There are two potential sources of incompleteness. One is that the mechanism designer's welfare objective may be an incomplete ordering, for example, if it only relies on Pareto comparisons. The second source of incompleteness is that the comparison of different mechanisms may not be uniform across the type space. One mechanism may perform better than another for some type vectors, whereas the reverse is the case for some other type vectors.

As our order is incomplete, when "solving" the mechanism designer's problem we shall not look for a class of mechanisms that are optimal and among which the mechanism designer is indifferent, or at least a member of such a class. Instead we shall look for the set of mechanisms such that the designer ranks no mechanism in this set, or outside of this set, above the given mechanism. The mechanism designer need not be indifferent between the members of this set. Typically, instead, she will be unable to order any two elements of this set without committing first to a particular belief. By studying this set, we hopefully obtain results about robust mechanism design that are independent of the details of the mechanism designer's subjective beliefs about her environment. This set is analogous, for example, to the "Pareto-frontier" of feasible allocations in an exchange economy. The set will be referred to as the set of "undominated" mechanisms.

10.10 CONCEPTUAL PROBLEMS IN THE QUASI-LINEAR CASE

In this and the next section we focus on the case of no common prior. We divide our study of the dominance relation described in the previous section into two parts. The first part, this section, deals with the case of quasi-linear preferences. The second part, the next section, gives an example in which preferences are not quasi-linear.

This section will be about the implications of Proposition 10.7 for mechanism design. Consider a benevolent mechanism designer who evaluates mechanisms on the basis of agents' expectations—that is, at the interim level, where these expectations may be based on subjective and inconsistent beliefs. Then the mechanism designer can build bets of arbitrary size into the mechanism and therefore increase agents' interim expected utility without limit. An optimal mechanism will, in general, not exist. Intuitively, this mechanism designer, for example, may appreciate public lotteries, even if with hindsight almost all participants lose money. The reason is that agents may hold subjective beliefs that attach much higher weight to certain numbers as outcomes of the lottery than to others. Then all agents may have positive expected value at the stage of submitting their bets but before seeing the realized numbers. If the mechanism designer is nonpaternalistic to the extreme and accepts these somewhat crazy beliefs of the agents, then the lottery is a Pareto improvement. Because the size of bets is unbounded, this leads to the nonexistence of optimal mechanisms. We state an example of a formal result that is based on this intuition:

Proposition 10.10 *In the single unit auction problem, if the mechanism designer's objective function is interim Pareto welfare, if* $N \geq 3$*, and if the type space is the set of all finite types, then the set of undominated mechanisms is empty.*

A similar result will be true for interim Pareto welfare in other problems with quasi-linear utility, such as the public goods or the bilateral trade problem. By focusing on the space of all finite types, we make sure in this result that sufficiently inconsistent beliefs are included that allow us to apply Proposition 10.7.

If the mechanism designer seeks to maximize revenue, then she might arrange bets for agents and charge agents a fee for those bets. Given that in Proposition 10.7 both i and j strictly prefer betting over not betting if their types are in $\widehat{T}_i$ or $\widehat{T}_j$, they will even be willing to pay a small fee to the auctioneer for their bets. Just as the size of bets may tend to infinity, also the size of the charge may tend to infinity. Thus, the mechanism designer has possibly arbitrarily large fees for arranging bets. One might object that agents may bet with each other even without telling the mechanism designer and without paying the mechanism designer's fees. But this plausible argument is outside of our model. In our model, the mechanism designer sets the rules of a game, and this game is the only game that agents play. They have no other opportunity to get together and enter agreements. This is in the nature of the theory of mechanism design as explained in this book, but of course it is a drawback of the theory. Mechanisms in practice organize only a small part of an agent's economic life, and there are many other parts of an agent's economic life in which there are no well-defined and chosen rules of the game. But in our formal setting, we immediately obtain the following result:

Proposition 10.11 *In the single unit auction problem, if the mechanism designer's objective function is ex post revenue, if* $N \geq 3$*, and if the type space is the set of all finite types, then the set of undominated mechanisms is empty.*

The betting arguments that we explained in this section do not, however, necessarily affect the case that the mechanism designer's objective is an ex post welfare criterion. Suppose, for example, that the mechanism designer's definition of welfare is the ex post sum of agents' utilities. This sum is not affected by monetary transfers. For this case, there are therefore no analogues of the two nonexistence results listed above.

10.11 VOTING REVISITED

We now revisit the voting problem from Section 8.2. We thus assume non-transferrable utility. We also allow subjective beliefs. There does not seem to be a characterization of undominated mechanisms in this case in the literature. Our objective here is more modest. It is to show that the theory of robust mechanism design as developed here allows one to overcome the dead end for research that was apparently created by impossibility results such as the Gibbard–Satterthwaite theorem, Proposition 8.1.

To make things as simple as possible, we assume that there are two agents, $i \in \{1, 2\}$, and three candidates; $x \in C = \{a, b, c\}$. All results reported in this chapter generalize to the case that there are arbitrarily many agents. Propositions 10.13 and 10.14 also generalize to the case of arbitrarily finite sets of candidates. However, for Proposition 10.15 I do not know whether it generalizes to more than three candidates.

The set of alternatives A is the set of all probability distributions over $\{a, b, c\}$: $A = \Delta(\{a, b, c\})$. We assume private values: For every $i \in \{1, 2\}$, $\theta \in \Theta$, and $\delta \in \Delta(\{a, b, c\})$, we assume that $u_i(\theta, \delta)$ is a function of θ_i only, not of θ_{-i}. We also assume that for every $i \in \{1, 2\}$ and $\theta \in \Theta$, $u_i(\theta)$ satisfies the expected utility axioms and that $u_i(\theta)$ does not assign the same value to two different candidates in C, so that no agent is ever indifferent between two candidates. Finally, we assume that the space of payoff types is large: For every utility function $\hat{u}_i$ that satisfies the expected utility axioms and that does not assign the same value to two different candidates in C, there is a $\theta \in \Theta$ such that $u_i(\theta)$ is $\hat{u}_i$.

We shall focus on finite mechanisms (S_1, S_2, g). The finiteness restriction can be interpreted as a very elementary way of modeling simplicity. We also focus on mechanisms that have two desirable properties listed in the next definition. These two conditions are implied by, but weaker than, ex post Pareto efficiency. The main reason that we impose these conditions is that they are included in the conditions of Hylland's theorem, which we shall state below and which will be the starting point of our analysis. We mentioned Hylland's theorem already in Section 8.3.

Definition 10.17 *A finite mechanism (S_1, S_2, g) and a Bayesian equilibrium σ^* of this mechanism on some type space $\mathcal{T}$ satisfy* *(i)* positive unanimity *if whenever some candidate $x \in C$ is most preferred by both agents (given their payoff types), then*

$$\sum_{s \in S} \sigma_1^*(\tau_1)[s_1] \cdot \sigma_2^*(\tau_2)[s_2] \cdot g(s)[x] = 1; \tag{10.9}$$

(ii) negative unanimity *if whenever some candidate $x \in C$ is least preferred by both agents (given their payoff types), then*

$$\sum_{s \in S} \sigma_1^*(\tau_1)[s_1] \cdot \sigma_2^*(\tau_2)[s_2] \cdot g(s)[x] = 0. \tag{10.10}$$

The two sums in this definition add up for every strategy combination s the product of the probability with which this strategy combination is chosen under σ^* and the probability with which candidate x is chosen if the strategy vector s is played. I hope that the notation is self-explanatory. The first condition then says that a candidate ranked top by both individuals is chosen with probability 1, and the second condition says that a candidate ranked bottom by both individuals is chosen with probability 0.

The first result that we present is an impossibility result. We need for this result the following definition:

Definition 10.18 *A mechanism (S_1, S_2, g) and a Bayesian equilibrium σ^* of (S_1, S_2, g) for some type space $\mathcal{T}$ are a* random dictatorship *if there is some $p \in [0, 1]$ such that for every $\tau \in T$ and $x \in C$ we have*

$$\sum_{s \in S} \sigma_1^*(\tau_1)[s_1] \cdot \sigma_2^*(\tau_2)[s_2] \cdot g(s)[x] = \sum_{i=1,2} \mathbb{1}_i(x) p_i, \tag{10.11}$$

where $\mathbb{1}(x)$ is an indicator which is 1 if x is agent i's most preferred candidates, given her payoff type $\hat{\theta}_i(\tau_i)$, and 0 otherwise, and where $p_1 \equiv p$ and $p_2 \equiv 1 - p$.

In a random dictatorship, thus, agent 1's preferred candidate is chosen with probability p, and agent 2's preferred candidate is chosen with probability $1 - p$, regardless of the agents' type vectors.

The following is a distant relative of Gibbard and Satterthwaite's impossibility result. It is a very close relative of a theorem due to Hylland.[7] We don't prove the result here.

Proposition 10.12 *A mechanism (S_1, S_2, g) and a belief-independent Bayesian equilibrium σ^* of G for a type space with a large variety of certainties $\mathcal{T}$ satisfy positive and negative unanimity if and only if they are a random dictatorship.*

From now on, when we refer to random dictatorship, it is useful to have in mind a specific mechanism and a specific Bayesian equilibrium of this mechanism.

Definition 10.19 *For any $p \in [0,1]$ and type space $\mathcal{T}$ the following mechanism (S_1, S_2, g) and equilibrium σ^* of (S_1, S_2, g) on $\mathcal{T}$ will be referred to as p-*random dictatorship*:*

(i) $S_i = C$ *for* $i = 1, 2$,
(ii) $g(s, x) = \sum_{i=1,2} \mathbb{1}_{\{i: x = s_i\}} p_i$,
(iii) $\sigma_i^*(\tau_i)(\arg\max u_i(\hat{\theta}(\tau_i))) = 1$ *for* $i = 1, 2$ *and* $\tau_i \in T_i$,

where $p_1 \equiv p$ *and* $p_2 \equiv 1 - p$.

There are other mechanisms and equilibria that are random dictatorships in the sense of Definition 10.18, but it is without loss of generality to only consider the ones described in Definition 10.19.

We we will consider three possible welfare functions for the mechanism designer. The first two will be the interim and the ex post Pareto welfare function that were already defined above. The third will be an equity focused welfare function that we shall introduce later. Our first result says that for every $p \in (0,1)$ there are a mechanism and a Bayesian equilibrium of this mechanism that are ranked higher than p-random dictatorship if the mechanism designer's objective is interim Pareto welfare.[8] We refer to the higher ranked mechanism as *p-random dictatorship with compromise*:

Definition 10.20 *For every $p \in [0,1]$ the following mechanism is called p-*random dictatorship with compromise.

(i) *For* $i = 1, 2$:

$$S_i = \mathcal{C} \times \mathcal{R},$$

where $\mathcal{C}$ is the set of all nonempty subsets of C, and $\mathcal{R}$ is the set of all complete strict ordinal rankings of C; we write $s_i = (C_i, R_i) \in S_i$ for a strategy for agent i.

(ii) *If* $C_1 \bigcap C_2 = \emptyset$, *then for all* $x \in C$ *we have*

$$g(s, x) = \sum_{\{i \in I : x R_i x' \ \forall x' \in C\}} p_i.$$

(iii) *If* $C_1 \bigcap C_2 \neq \emptyset$, *then for all* $x \in C_1 \bigcap C_2$ *we have*

$$g(s, x) = \sum_{\{i \in I : x R_i x' \ \forall x' \in C_1 \bigcap C_2\}} p_i,$$

where $p_1 \equiv p$ *and* $p_2 \equiv 1 - p$.

In words, this mechanism offers each agent i the opportunity to provide a complete ranking R_i of the candidates, along with a set C_i of "acceptable" alternatives. If there is at least one common element among the sets of acceptable alternatives for all agents, then the mechanism implements random dictatorship (with the preferences described by the R_i) but with the restriction that the dictator can only choose an outcome from the unanimously acceptable alternatives. Otherwise, the mechanism reverts to random dictatorship (with outcomes determined by the highest ranked elements of the R_i). We refer to this mechanism as *p-random dictatorship with compromise* because it offers agents the opportunity to replace the outcome of p-random dictatorship by a compromise on a mutually acceptable alternative.

What are reasonable strategy choices in this mechanism? It is elementary to verify that a strategy of player i such that for some type τ_i, her most preferred candidate is not included in her set C_i of acceptable candidates, is weakly dominated by the same strategy in which that most preferred candidate is added to C_i. Moreover, any strategy of some player i such that for some type τ_i we have that R_i is not type τ_i's true preference over C_i, as described by $u_i(\hat{\theta}(\tau_i))$, is weakly dominated by a strategy such that C_i is left unchanged, but R_i is replaced by a preference ordering that reflects τ_i's true preference over C. These considerations motivate us to restrict our attention to "truthful strategies" which we define to be strategies such that C_i includes i's most preferred candidate (given her payoff type) and such that R_i is the true preference according to $u_i(\hat{\theta}(\tau_i))$, for all types τ_i. Note that we have ruled out some, but not necessarily all, weakly dominated strategies. In any case, it seems eminently plausible that all players will choose truthful strategies.

In a Bayesian equilibrium on any type space, in which all players choose truthful strategies, any type's interim expected utility is not smaller than the interim expected utility from p-random dictatorship. This is because a type can always force an outcome that gives at least as high interim expected utility as p-random dictatorship by choosing the truthful strategy for which C_i consists of her most preferred candidate only. Note also that all players choosing this strategy for all types will always be a Bayesian equilibrium, whatever the type space is.

We now show that *p-random dictatorship with compromise* also has a Bayesian equilibrium for the space of all finite types in which all players choose truthful strategies, and that interim Pareto dominates p-random dictatorship. We further show that this equilibrium respects positive and negative unanimity. The latter observation clarifies that our result is indeed a consequence of weakening the belief independence requirement, and not of weakening any other property listed in Proposition 10.12. We restrict our attention in the following result to the space $\mathcal{T}^+$ of all finite types. This is to avoid technical issues of existence of Bayesian equilibrium. Note that $\mathcal{T}^+$ has a large variety of certainties, so that Hylland's impossibility result holds for this type space.

Proposition 10.13 *For every $p \in (0, 1)$, p-random dictatorship with compromise has a Bayesian equilibrium for the space $\mathcal{T}^+$ of all finite types that interim Pareto dominates p-random dictatorship and satisfies positive and negative unanimity.*

Proof

For every η satisfying $0.5 < \eta < 1$, denote by $\theta_1^*(\eta)$ the payoff type of player 1 who has utility function $u_1(\theta_1^*(\eta), a) = 1, u_1(\theta_1^*(\eta), b) = \eta, u_1(\theta_1^*(\eta), c) = 0$ and denote by $\theta_2^*(\eta)$ the payoff type of player 2 who has utility function $u_2(\theta_2^*(\eta), a) = 0$, $u_1(\theta_2^*(\eta), b) = \eta, u_2(\theta_2^*(\eta), c) = 1$. Denote by b_1^* and b_2^* the infinite hierarchies of beliefs according to which these payoff types are common certainty. Note that $T_1 = \{(\theta_1^*, b_1^*)\}$ and $T_2 = \{(\theta_2^*, b_2^*)\}$ form a minimal belief-closed subset of the type space $\mathcal{T}^+$. We define the equilibrium strategies of these types to be $C_1 = \{a, b\}$ and $C_2 = \{c, b\}$, with the players reporting their true ordinal preferences. These strategies result in outcome b being realized with probability 1, which both players strictly prefer to the lottery implied by random dictatorship. For all other minimal belief-closed subsets of $\mathcal{T}^+$, we can assume that players play any arbitrary Bayesian Nash equilibrium in truthful strategies. No player's interim expected utility can be less than it would be under random dictatorship because every player can always force a reversion of the mechanism to random dictatorship by choosing $C_i = \{\arg\max u_i(\hat{\theta}(\tau_i))\}$.

□

We next show that no mechanism ex post Pareto dominates p-random dictatorship.

Proposition 10.14 *For every $p \in [0, 1]$ there is no mechanism M with a Bayesian equilibrium σ^* for the space $\mathcal{T}^+$ of all finite types that ex post Pareto dominates p-random dictatorship.*

Proof

Indirect. Suppose for some $p \in [0, 1]$ there were a mechanism M and a Bayesian equilibrium of M for the space of all finite types that ex post Pareto dominated p-random dictatorship. For the outcome resulting from M and σ^* to be different from p-random dictatorship, there must be some vector of types $(\hat{\tau}_1, \hat{\tau}_2) \in \mathcal{T}^+$ such that the probability with which one agent's preferred candidate is chosen is less than it would be under random dictatorship. Without loss of generality, assume that the agent is agent 1 and that the candidate that is agent 1's preferred candidate is a. Suppose that $\hat{\tau}_1$ ranks candidate b second. Denote by $\tilde{\tau}_2$ the type of player 2 that gives b utility 1, c utility $1 - \varepsilon$ for $\varepsilon > 0$ arbitrarily close to zero, and a utility 0 and that attaches probability 1 to player 1 being of type $\hat{\tau}_1$. Consider the equilibrium outcome for type vector $(\hat{\tau}_1, \tilde{\tau}_2)$. If ε is sufficiently close to zero, then the probability of a at this type vector must be less than p, because otherwise agent 2 would deviate and choose the strategy of type $\hat{\tau}_2$. The joint probability of b and c must therefore be more than p. But then ex post agent 1 prefers at this type vector random dictatorship, because random dictatorship would yield a with probability p and his second most preferred candidate b with probability $1 - p$. This contradicts the assumption that M and equilibrium σ^* ex post Pareto dominate p-random dictatorship. □

We now introduce another possible objective for the mechanism designer: We are interested in a designer who prefers agents to reach a compromise not so much for Pareto efficiency but for ethical reasons. We give the mechanism designer the following welfare function:

Definition 10.21 *The following welfare function is called "ex post equity:" If $g(\tau) \in C$, then:*

$$w(\mathcal{T}, M, \tau) = \begin{cases} 1 & \text{if } g(\tau) \in \arg\max u_i(\hat{\theta}(\tau_i)) \text{ for } i = 1, 2, \\ 0 & \text{if } g(\tau) \in \arg\min u_i(\hat{\theta}(\tau_i)) \text{ for } i = 1 \text{ or } i = 2, \\ 0.5 & \text{otherwise.} \end{cases}$$

If $g(\tau) \in \Delta(C)$, then $w(\mathcal{T}, M, \tau)$ is the expected value of the function defined above.

Proposition 10.15 *For every $p \in [0, 1]$, p-random dictatorship with compromise has a Bayesian equilibrium in truthful strategies for the space $\mathcal{T}^+$ of all finite types that ex post equity dominates p-random dictatorship.*

Proof

Consider the belief closed type space (T_1, T_2) that we constructed in the proof of Proposition 10.13. As in the proof of Proposition 10.13, we fix for player 1's type in this type space the strategy $C_1 = \{a, b\}$ and for player 2 the strategy $C_2 = \{a, b\}$. The resulting outcome, b, gives the mechanism designer ex post welfare of 0.5, whereas the outcome of random dictatorship would have given the mechanism designer ex post welfare of 1.

For all other type spaces we pick some arbitrary Bayesian Nash equilibrium in truthful strategies taking as given and fixed the strategies that the types in (T_1, T_2) play. We can complete the proof by showing that for all type vectors (τ_1, τ_2) if players play truthful strategies, the equilibrium outcome of p-random dictatorship does not yield lower expected welfare than the equilibrium outcome of random dictatorship.

We distinguish two cases regarding the payoff type vector $(\hat{\theta}(\tau_1), \hat{\theta}(\tau_2))$. First, suppose that both agents prefer the same candidate in $\{a, b, c\}$. Then, both p-random dictatorship and the equilibrium of p-random dictatorship with compromise pick this candidate with probability 1. There is no welfare difference. Now suppose that both agents rank the same candidate bottom. Then, both p-random dictatorship and the equilibrium of p-random dictatorship with compromise pick this candidate with probability 0 and the other two candidates with probability 0.5. There is again no welfare difference. It remains the case that the two agents differ both in their top ranked and in their bottom ranked candidate. Then the preferred candidate of one agent must be the lowest ranked of the other. Therefore, random dictatorship yields expected welfare zero. The Bayesian equilibrium in truthful strategies of random dictatorship with compromise cannot do worse. □

10.12 REMARKS ON THE LITERATURE

A seminal paper on the theory of robust mechanism design to which the presentation in this chapter owes much is Bergemann and Morris (2005). Bergemann and Morris focus on the question of whether a mechanism designer can find a direct mechanism and a Bayesian equilibrium of this direct mechanism that implement a given social choice correspondence when the type space is large. Here, a social choice correspondence maps payoff types into outcomes. Bergemann and Morris have also extensively worked on robust mechanism design when the mechanism designer cannot select the Bayesian equilibrium of any proposed mechanism, but when the mechanism designer needs to take all Bayesian equilibria of the mechanism into account. Bergemann and Morris's work on robustness is surveyed in Bergemann and Morris (2013). The focus in this chapter is different from that of Bergemann and Morris's work in that we seek an explicit formulation of the decision-maker's decision problem. Another seminal paper in this area is Neeman (2004), to which we already referred in the literature review in Chapter 6.

The example in Section 10.2 is inspired by similar, but different, examples in Neeman (2004). The framework developed in Section 10.3 is due to Bergemann and Morris (2001) and Bergemann and Morris (2005). Proposition 10.1 is Lemma 3.3 in Bergemann and Morris (2001). It is also, in an application, Lemma 1 in Neeman (2004). It is also implicit in the proof of Proposition 1 in Bergemann et al. (2012).

Proposition 10.6 is Proposition 4.5 in Bergemann and Morris (2001). It is implicit in Bergemann et al. (2012) in the proof of Proposition 1, and it is explicitly mentioned in the text of that paper. Lemma 3 in Farinha Luz (2013) is also similar to Proposition 10.6. Farinha Luz presents the result in an auction setting for a large type space with a common prior and private values.

Propositions 10.8 and 10.9 are both implied by Proposition 2 in Bergemann and Morris (2005). Their result is stronger than the combination of Propositions 10.8 and 10.9, but Propositions 10.8 and 10.9 illustrate the most interesting cases to which Bergemann and Morris' Proposition 2 applies. Bergemann and Morris (2005) provide other sufficient conditions for ex post implementability through reduced direct mechanisms. They also provide conditions under which more can be said about transfers in the reduced direct mechanism—in particular, conditions under which transfers can be chosen to be ex post budget balanced. They also provide examples of social choice correspondences that can be implemented in Bayesian equilibrium on large type spaces, but that cannot be implemented through ex post Bayesian equilibria.

The idea of Section 10.8 to build on Propositions 10.6 and 10.1 and to consider optimal mechanism design conditional on each vector of belief types is used by Farinha Luz (2013) to determine an expected revenue-maximizing auction of a single unit. Farinha Luz uses a stronger condition than Proposition 10.6, namely that each player's belief vectors are linearly independent. As a result, he obtains an optimal auction in which truthful reporting of a bidder's type is a dominant strategy.[9]

The approach to comparing different mechanisms without a common prior described in Subsection 10.9.2 is based on Smith (2010), who studies the design of a mechanism for public goods. Smith considers the performance of different mechanisms in a Bayesian equilibrium on all type spaces. He focuses on an ex post efficiency and demonstrates that a mechanism designer can improve efficiency using a more flexible mechanism than a dominant strategy mechanism. Smith also identifies one mechanism for the public goods problem such that the mechanism designer ranks no other mechanism higher.

Smith (2010) and Bergemann et al. (2012) have studied the construction of ex post efficient mechanisms in the case of quasi-linear preferences when agents don't necessarily have a common prior. Ex post efficiency is in both papers defined as the sum of ex post utilities, so that monetary transfers don't affect efficiency. Bergemann et al. (2012) use the fact that then also with subjective priors belief elicitation is costless. They then adapt Proposition 5.2 to provide conditions under which ex post efficient outcomes can be implemented. Smith studies the public goods problem. He requires ex post budget balance, which makes belief elicitation impossible. He gives one example of an undominated mechanism.

Section 10.11 is a shortened version of material in Börgers and Smith (2014) and in Börgers and Smith (2012). The p-random dictatorship with compromise mechanism was inspired by *Approval Voting* [see Brams and Fishburn (2007)], which, like our mechanism, allows voters to indicate "acceptable" alternatives. However, in approval voting the alternative that the largest number of agents regards as acceptable is selected, whereas our mechanism requires unanimity. Moreover, our mechanism uses random dictatorship as a fallback, whereas approval voting does not have any such fallback. When $p = 1/2$ the mechanism that we consider is closely related to the *Full Consensus or Random Ballot Fall-Back* mechanism that Heitzig and Simmons (2012) introduced. Heitzig and Simmons require the sets C_i to be singletons.

Chung and Ely (2007) describe an auctioneer of a single object who designs an auction to maximize expected revenues. The auctioneer considers equilibria of different auction mechanisms on the universal type space, and he evaluates different mechanisms using a maximin criterion: Taking the distribution of the agents' valuations but not the agents' beliefs, as given, for each mechanism the auctioneer determines the probability distribution on the universal type space for which that mechanism yields the lowest expected revenue. The auctioneer then chooses a mechanism that maximizes the lowest expected revenue. Whereas in this chapter the mechanism designer has only a partial order of mechanisms, Chung and Ely's mechanism designer has a complete order. Chung and Ely allow subjective beliefs. The observations in Section 10.7.2 apply to their model. However, their maxmin approach implies that the mechanism designer focuses on type spaces in which profitable betting games cannot be arranged. This issue is discussed in more detail in Börgers (2013).

Yamashita (2012) considers another approach to the choice problem of a mechanism designer without a common prior. Yamashita considers a bilateral trade setting, and he evaluates mechanisms on the basis of the lowest expected welfare among all outcomes that can result if agents use strategies that are not weakly dominated. Expected welfare is calculated on the basis of the mechanism designer's subjective prior over agents' types.

10.13 PROBLEMS

(a) Give a precise formal proof of Proposition 10.5. Follow the outline given in the text.

(b) Consider the bilateral trade problem of Section 3.4. Assume that the seller's type is commonly known to be zero and that the buyer's type is drawn from the set $[0, 1]$. Consider the following fixed price mechanism: The buyer reports her type. If the reported type is above 0.5, then the object is transferred from the seller to the buyer, and the buyer pays to the seller 0.5 dollars. Otherwise, the seller keeps the object and no payments are made. Now consider the alternative mechanism in which the seller can choose either price 0.4 or price 0.5. The buyer reports her type. If her type is above the price that the seller chose, then the object is transferred from the seller to the buyer, and the buyer pays to the seller the price that the seller chose. Otherwise, the seller keeps the object and no payments are made. Prove that this mechanism has a Bayesian equilibrium on the space of all finite types that ex post Pareto dominates the truthful equilibrium of the fixed price mechanism.

(c) In the setting of problem (c) in Chapter 9, find a mechanism and a Bayesian equilibrium of that mechanism that interim Pareto-dominate on all finite type spaces the mechanism that chooses alternative a regardless of agents' types. Do the mechanism and Bayesian equilibrium that you find ex post Pareto-dominate the constant mechanism?

11

DYNAMIC MECHANISM DESIGN

Daniel Krähmer and Roland Strausz

11.1 INTRODUCTION

The classic mechanism design approach considers a static framework, where agents observe private information only once, and the mechanism implements a single allocation. This chapter considers two different approaches toward introducing dynamics into the mechanism design problem: dynamic private information and dynamic allocations.

In order to isolate the effects of dynamic private information, the chapter first analyzes a framework in which the allocation itself—the sale of a single indivisible good—is still static, but the agents' private information changes over time. It is shown that a revenue-maximizing seller benefits from screening an agent sequentially by offering a menu of option contracts. In a large class of problems, optimal sequential screening enables the seller to extract at zero costs all the private information that agents receive after accepting the mechanism. An insight of this dynamic extension is therefore that private information which is received before the agent accepts the mechanism has a much stronger economic impact than private information which the agent receives after accepting the mechanism.

The chapter next studies a framework of dynamic allocations, where the mechanism picks an allocation for multiple periods but private information does not change over time. It is shown that this dynamic framework does not lead to any meaningful dynamics in the sense that the optimal dynamic mechanism is static and a simple repetition of the optimal static mechanism. The result, however, depends crucially on the full commitment assumption, which underlies the traditional theory of mechanism design.

11.2 DYNAMIC PRIVATE INFORMATION

11.2.1 Sequential Screening

In order to study dynamic private information, we revisit the seller of Section 2.2 who seeks to sell a single indivisible good to a buyer with some valuation $\theta \geq 0$. We change the setup and explore the idea that upon meeting the seller, the buyer is better but not perfectly informed about her valuation θ and that she learns her true valuation only later. To model this idea, we assume that prior to meeting the seller, the buyer privately observes a signal τ which is distributed with the cumulative distribution $G(\tau)$ and density $g(\tau) > 0$ on the support $[\underline{\tau}, \bar{\tau}]$. Only after the buyer accepts the seller's mechanism, she privately observes her true valuation θ. To capture that the signal τ provides information about the buyer's true valuation θ, we assume that τ and θ are correlated; that is, conditional on τ, the valuation θ is distributed with the cumulative distribution function $F(\theta|\tau)$ and density $f(\theta|\tau)$. We refer to τ as the buyer's ex ante type and refer to θ as her ex post type.

We maintain the assumptions of Section 2.2 that the support of $F(\theta|\tau)$ is an interval $[\underline{\theta}, \bar{\theta}]$ for all $\tau \in [\underline{\tau}, \bar{\tau}]$, where $0 \leq \underline{\theta} < \bar{\theta}$, and that $f(\theta|\tau) > 0$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and for all $\tau \in [\underline{\tau}, \bar{\tau}]$. Consequently, the support of $F(\theta|\tau)$ is independent of τ. In addition, we assume that $F(\theta|\tau)$ and $f(\theta|\tau)$ are continuously differentiable in τ and the derivatives are bounded, i.e., there exists a $K > 0$ such that $|\partial F(\theta|\tau)/\partial\tau| < K$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$ and for all $\tau \in [\underline{\tau}, \bar{\tau}]$. Moreover, we assume that the family $F(\cdot|\tau)$, $\tau \in [\underline{\tau}, \bar{\tau}]$, is ordered in the sense of first-order stochastic dominance (FOSD) so that $\partial F(\theta|\tau)/\partial\tau < 0$ for all $\theta \in (\underline{\theta}, \bar{\theta})$. FOSD captures the idea that a higher value of τ is good news about the buyer's valuation θ.

In this formulation of dynamic private information, the initial information τ is not directly payoff relevant. To see that this assumption is without loss of generality, suppose that the buyer observes an initial signal $\tau \in [\underline{\tau}, \bar{\tau}]$ and a subsequent signal $\gamma \in [\underline{\gamma}, \bar{\gamma}]$ with joint cumulative distribution $H(\tau, \gamma)$, which together determine the buyer's valuation $\theta = \theta(\tau, \gamma) \in [\underline{\theta}, \bar{\theta}]$. In this formulation, the signal τ is payoff relevant in the sense that it influences the buyer's valuation directly. We can, however, reinterpret this model as one in which the buyer first observes τ and subsequently learns his payoff type θ. We only have to define $G(\tau) = H(\tau, \bar{\gamma})$ and $F(\theta|\tau) = \text{Prob}[\gamma \in \{\gamma : \theta(\tau, \gamma) \leq \theta\}|\tau]$. We will come back to this equivalence later.

Despite the payoff irrelevance of τ, the private information of the buyer is multidimensional. In Section 2.4 we saw that, in a static environment, multidimensional private information may render the analysis of the monopolistic screening model intractable. As we will see, this is not the case in the sequential screening model. In this sense, the sequential screening model is "nice" and an attractive first step toward a tractable analysis of dynamic mechanism design. We keep the analysis as close as possible to the analysis of the static screening model presented in Chapter 2. Hence, we begin with considering direct mechanisms.

Definition 11.1 *A (dynamic) "direct mechanism" consists of functions q and t where*

$$q : [\underline{\tau}, \bar{\tau}] \times [\underline{\theta}, \bar{\theta}] \to [0, 1]$$

and

$$t : [\underline{\tau}, \bar{\tau}] \times [\underline{\theta}, \bar{\theta}] \to \mathbb{R}.$$

As before, the interpretation is that a (dynamic) direct mechanism requires the buyer to report her private information and therefore seems a straightforward extension of static mechanisms. There are, however, two subtle differences to note. First, the fact that we define the (dynamic) direct mechanism on the Cartesian product $[\underline{\tau}, \bar{\tau}] \times [\underline{\theta}, \bar{\theta}]$ is appropriate only because the support of $F(\theta|\tau)$ is independent of τ. If it were not, then not all type combinations $(\tau, \theta) \in [\underline{\tau}, \bar{\tau}] \times [\underline{\theta}, \bar{\theta}]$ are possible so that a direct mechanism defined on the Cartesian product $[\underline{\tau}, \bar{\tau}] \times [\underline{\theta}, \bar{\theta}]$ would specify too many contingencies. In other words, if we take seriously the idea that a direct mechanism requires the buyer to report *only her private information and nothing more*, a direct mechanism specifies $q(\tau, \theta)$ and $t(\tau, \theta)$ only if $f(\theta|\tau) > 0$. The second subtlety regards the timing of the direct mechanism. Crucially, the direct mechanism is dynamic in that the buyer reports her ex ante private information τ before she learns her ex post private information θ. Hence, the direct mechanism distinguishes between a first reporting stage, where the buyer reports τ, and a second reporting stage, where she reports θ.

A direct mechanism induces a dynamic single-person decision problem with initial uncertainty about θ. We can represent the buyer's optimal decision rule as a pair $\sigma = (\sigma_1, \sigma_2)$, where σ_1 is a mapping $\sigma_1 : [\underline{\tau}, \bar{\tau}] \to [\underline{\tau}, \bar{\tau}]$ and σ_2 is a mapping $\sigma_2 : [\underline{\tau}, \bar{\tau}] \times [\underline{\theta}, \bar{\theta}] \times [\underline{\tau}, \bar{\tau}] \to [\underline{\theta}, \bar{\theta}]$. The mapping σ_1 is the buyer's report of her ex ante type τ. It conditions only on the ex ante type τ, because the buyer has to report τ before learning θ. The mapping σ_2 is the buyer's report of her ex post type θ. It conditions on both the ex ante type τ and the ex post type θ. In addition, it conditions on the report which the buyer has sent about her type τ.

Following the approach in earlier chapters, the next proposition derives a revelation principle by considering general, not necessarily direct, dynamic mechanisms Γ without giving a formal definition of such mechanisms. It suffices for our purposes to keep in mind that a general mechanism is any game form where a strategy of the buyer is a mapping from the buyer's type in the action space specified by the mechanism.

Proposition 11.1 (*Dynamic Revelation Principle*) *For every dynamic mechanism Γ and every optimal buyer strategy σ in Γ, there is a direct mechanism Γ' and an optimal buyer strategy $\sigma' = (\sigma_1', \sigma_2')$ such that:*

(i) The strategy σ' satisfies:

$$\sigma_1'(\tau) = \tau \qquad \text{for every } \tau \in [\underline{\tau}, \bar{\tau}],$$
$$\sigma_2'(\tau, \theta, \tau) = \theta \qquad \text{for every } \theta \in [\underline{\theta}, \bar{\theta}], \tau \in [\underline{\tau}, \bar{\tau}];$$

that is, σ' prescribes telling the truth about τ and, after the buyer reported τ truthfully, telling the truth about θ.

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

Proof

For every $(\tau, \theta) \in [\underline{\tau}, \bar{\tau}] \times [\underline{\theta}, \bar{\theta}]$ define $q(\tau, \theta)$ and $t(\tau, \theta)$ as required by (ii) in Proposition 11.1. We prove the result by showing that for this direct mechanism the strategy σ' is optimal for the buyer. Note that under this strategy, for every (τ, θ) the buyer with type (τ, θ) obtains in the mechanism Γ' the same expected utility as in the mechanism Γ when choosing strategy σ. Moreover, when pretending to be some type τ' and subsequently reporting type $\theta'(\theta, \tau')$ when learning θ, the buyer obtains the same expected utility that she would have obtained had she played the strategy of type τ' before learning θ and after learning θ had followed the strategy of type $(\tau', \theta'(\theta, \tau'))$ in Γ. The optimality of playing $\sigma_1(\tau) = \tau$ and $\sigma_1(\tau, \theta, \tau) = \theta$ in Γ' for any $\tau \in [\underline{\tau}, \bar{\tau}]$ and $\theta \in [\underline{\theta}, \bar{\theta}]$ then follows immediately from the optimality of σ in Γ. □

Hence, the revelation principle extends to dynamic environments in that there is no loss in focusing on direct mechanisms that induce truth telling on the equilibrium path. A subtlety is, however, that the dynamic revelation principle does not say anything about the reporting behavior off the equilibrium path—that is, after the buyer has not told the truth about her ex ante type. Hence, by the dynamic revelation principle alone we cannot focus on direct mechanisms that also induce truthful reporting off the equilibrium path. Our formal definition of an incentive-compatible direct mechanism will have to reflect this feature. Consequently, we define

$$\theta^r : [\underline{\theta}, \bar{\theta}] \rightarrow [\underline{\theta}, \bar{\theta}]$$

as a reporting function. Clearly, buyers with different ex ante types or different reports about their ex ante type may use different reporting functions. Moreover, given a direct mechanism, let $u(\tau, \theta) = \theta q(\tau, \theta) - t(\tau, \theta)$,

$$\hat{U}(\tau'|\tau) = \int_{\underline{\theta}}^{\bar{\theta}} u(\tau', \hat{\theta}) f(\hat{\theta}|\tau)\, d\hat{\theta},$$

and $U(\tau) = \hat{U}(\tau|\tau)$.

Equipped with this additional notation, we now define incentive compatibility and individual rationality in our dynamic context.

Definition 11.2 *A direct mechanism is "incentive-compatible" if:*

(i) It is "incentive-compatible with respect to the ex post type θ," meaning that truth telling about θ is optimal for every ex ante type τ and every ex post type θ:

$$u(\tau, \theta) \geq \theta q(\tau, \theta') - t(\tau, \theta') \text{ for all } \tau \in [\underline{\tau}, \bar{\tau}] \text{ and } \theta, \theta' \in [\underline{\theta}, \bar{\theta}]. \tag{11.1}$$

(ii) It is "incentive-compatible with respect to the ex ante type τ," meaning that truth telling about τ is optimal for every ex ante type τ and any subsequent report about θ:

$$U(\tau) \geq \int_{\underline{\theta}}^{\bar{\theta}} [\hat{\theta} q(\tau', \theta^r(\hat{\theta})) - t(\tau', \theta^r(\hat{\theta}))] f(\hat{\theta}|\tau)\, d\hat{\theta}$$

for all $\tau, \tau' \in [\underline{\tau}, \bar{\tau}]$ and all $\theta^r : [\underline{\theta}, \bar{\theta}] \rightarrow [\underline{\theta}, \bar{\theta}]$.

Because we consider a framework in which the seller offers the buyer the mechanism after she learns her ex ante type τ but before she learns her final valuation θ, we require that, conditional on her ex ante type τ (but not her final type θ), the buyer is voluntarily willing to participate. This leads to the following definition of individual rationality.

Definition 11.3 *A direct mechanism is "individually rational" if*

$$U(\tau) \geq 0 \qquad \textit{for all } \tau \in [\underline{\tau}, \bar{\tau}].$$

We next characterize incentive compatibility and individual rationality. The following proposition shows that even though truth telling off the equilibrium path does not follow directly from the dynamic revelation principle, it nevertheless obtains, because in our formulation the agent's ex post type is equal to his payoff type. This insight will enable us to simplify the conditions under which a (dynamic) direct mechanism is incentive-compatible.

Proposition 11.2 *A direct mechanism is incentive-compatible if and only if it satisfies (11.1) and*

$$U(\tau) \geq \hat{U}(\tau'|\tau) \qquad \textit{for all } \tau, \tau' \in [\underline{\tau}, \bar{\tau}]. \tag{11.2}$$

Proof

"$\Leftarrow$" is obvious.

To show "$\Rightarrow$", note that (11.1) implies $\theta q(\tau', \theta) - t(\tau', \theta) \geq \theta q(\tau', \theta') - t(\tau', \theta')$ for all θ, τ', θ' so that for any $\theta^r : [\underline{\theta}, \bar{\theta}] \to [\underline{\theta}, \bar{\theta}]$ we have

$$\int_{\underline{\theta}}^{\bar{\theta}} [\hat{\theta} q(\tau', \hat{\theta}) - t(\tau', \hat{\theta})] f(\hat{\theta}|\tau)\, d\hat{\theta} \geq \int_{\underline{\theta}}^{\bar{\theta}} [\hat{\theta} q(\tau', \theta^r(\hat{\theta})) - t(\tau', \theta^r(\hat{\theta}))] f(\hat{\theta}|\tau)\, d\hat{\theta}.$$

Hence, if

$$U(\tau) \geq \hat{U}(\tau'|\tau) = \int_{\underline{\theta}}^{\bar{\theta}} [\hat{\theta} q(\tau', \hat{\theta}) - t(\tau', \hat{\theta})] f(\hat{\theta}|\tau)\, d\hat{\theta},$$

then also

$$U(\tau) \geq \int_{\underline{\theta}}^{\bar{\theta}} [\hat{\theta} q(\tau', \theta^r(\hat{\theta})) - t(\tau', \theta^r(\hat{\theta}))] f(\hat{\theta}|\tau)\, d\hat{\theta}$$

for all $\tau, \tau' \in [\underline{\tau}, \bar{\tau}]$ and all $\theta^r : [\underline{\theta}, \bar{\theta}] \to [\underline{\theta}, \bar{\theta}]$. □

In the static screening model, we have seen that incentive compatibility is fully characterized by monotonicity of the allocation rule and the revenue equivalence property that the allocation rule pins down the transfers up to a constant. This was useful in determining the optimal selling mechanism because it allowed us to eliminate the transfers from the seller's optimization problem.

We now investigate whether and to what extent this characterization extends to the sequential screening model. We will, in particular, see that incentive compatibility with respect to the ex ante type does, in general, not imply that the allocation rule is monotone in the ex ante type. This marks an important difference between the sequential and the static screening model. However, the static characterization of incentive compatibility carries over to incentive compatibility with respect to the ex post type θ. By following the analytical steps of Section 2.2, we can prove the following proposition.

Proposition 11.3 *A direct mechanism is incentive-compatible with respect to the ex post type θ if and only if:*

(i) For every ex ante type τ the function $q(\tau,\theta)$ is increasing in θ.
(ii) For every ex ante type τ, the function $u(\tau,\theta)$ is absolutely continuous in θ. In particular, it is differentiable except in at most countably many points, and whenever it is differentiable in θ: $\partial u(\tau,\theta)/\partial\theta = q(\tau,\theta)$.
(iii) For every ex ante type τ and ex post type θ:

$$t(\tau,\theta) = t(\tau,\underline{\theta}) + (\theta q(\tau,\theta) - \underline{\theta} q(\tau,\underline{\theta})) - \int_{\underline{\theta}}^{\theta} q(\tau,\hat{\theta})\, d\hat{\theta}.$$

Recall that the monotonicity of the allocation rule is a consequence of the single-crossing property which implies that a high valuation type gains more from consuming a high quantity than a low valuation type. Therefore, if the allocation rule was not monotone and a low valuation type has an incentive to announce her own type over announcing a high valuation type who is assigned a lower quantity, then this high valuation type would have an even higher incentive to announce the low valuation type than her own type.

When we now consider incentive compatibility with respect to the buyer's ex ante type, this argument fails. The reason is that the buyer's utility in the first period is given by an expectation over her second period utility and so depends on the whole schedule of quantities instead of a single, type-specific quantity only. This leaves the single-crossing property without bite, and we cannot, in general, infer monotonicity of the allocation rule in the ex ante type from incentive compatibility.

Having lost the monotonicity property of the allocation rule q with respect to the buyer's ex ante type, we next ask whether at least the revenue equivalence property of the static framework extends to our dynamic setup. In the static framework, the key step in proving this property was to show that incentive compatibility pins down the agent's marginal utility. If, in addition, the agent's utility function can be recovered from its derivative by integration, in other words, if the agent's utility function is absolutely continuous, then one can directly establish the revenue equivalence property.

We first demonstrate that incentive compatibility with respect to the ex ante type implies that $U(\tau)$ is increasing and absolutely continuous.

Lemma 11.1 *If a direct mechanism is incentive-compatible, then $U(\tau)$ is increasing and absolutely continuous.*

Proof

By part (ii) of Proposition 11.3, $u(\tau,\theta)$ is absolutely continuous in θ. Because also F is absolutely continuous in θ, we can apply integration by parts and obtain

$$\hat{U}(\tau'|\tau) = \int_{\underline{\theta}}^{\bar{\theta}} u(\tau',\theta)f(\theta|\tau)d\theta = \int_{\underline{\theta}}^{\bar{\theta}} q(\tau',\theta)[1-F(\theta|\tau)]d\theta, \tag{11.3}$$

where the second equality uses $\partial u(\tau,\theta)/\partial\theta = q(\tau,\theta)$ from Proposition 11.3(ii). Hence, for $\tau_2 > \tau_1$ it follows that

$$U(\tau_2) - U(\tau_1) \geq \hat{U}(\tau_1|\tau_2) - \hat{U}(\tau_1|\tau_1) = \int_{\underline{\theta}}^{\bar{\theta}} q(\tau_1,\theta)[F(\theta|\tau_1) - F(\theta|\tau_2)]d\theta \geq 0,$$

where the first inequality follows from incentive compatibility and the second inequality follows because of first order stochastic dominance. This establishes that $U(\tau)$ is increasing in τ.

Incentive compatibility with respect to the ex ante type τ implies moreover that

$$U(\tau) = \sup_{\tau'} \hat{U}(\tau'|\tau) = \hat{U}(\tau|\tau).$$

To show absolute continuity of U, we demonstrate that U is actually Lipschitz continuous, which implies absolute continuity. By definition, $U(\tau)$ is Lipschitz continuous if there is an $L > 0$ so that for all τ_1, τ_2:

$$|U(\tau_1) - U(\tau_2)| \leq L|\tau_1 - \tau_2|. \tag{11.4}$$

Indeed, observe first that for $U(\tau_1) \geq U(\tau_2)$,

$$\begin{aligned} U(\tau_1) - U(\tau_2) &\leq U(\tau_1) - \hat{U}(\tau_1|\tau_2) \\ &= \hat{U}(\tau_1|\tau_1) - \hat{U}(\tau_1|\tau_2) \\ &\leq \sup_{\tau'}\{\hat{U}(\tau'|\tau_1) - \hat{U}(\tau'|\tau_2)\}. \end{aligned}$$

A similar argument implies that for $U(\tau_1) < U(\tau_2)$, we have
$0 > U(\tau_1) - U(\tau_2) \geq \inf_{\tau'}\{\hat{U}(\tau'|\tau_1) - \hat{U}(\tau'|\tau_2)\}$. Therefore, for all τ_1, τ_2:

$$|U(\tau_1) - U(\tau_2)| \leq \sup_{\tau'} |\hat{U}(\tau' | \tau_1) - \hat{U}(\tau' | \tau_2)|.$$

Moreover, since by assumption $F(\theta|\tau)$ is continuously differentiable in τ, the mean value theorem applies so that there is a $\tilde{\tau} \in [\underline{\tau}, \bar{\tau}]$ with

$$F(\theta|\tau_2) - F(\theta|\tau_1) = \frac{\partial F(\theta|\tilde{\tau})}{\partial \tau}(\tau_2 - \tau_1).$$

Hence, from the inequality above and because $q \leq 1$,

$$\begin{aligned}
|U(\tau_1) - U(\tau_2)| &\leq \sup_{\tau'} |\hat{U}(\tau'|\tau_1) - \hat{U}(\tau'|\tau_2)| \\
&= \sup_{\tau'} |\int_{\underline{\theta}}^{\bar{\theta}} q(\tau', \theta)[F(\theta|\tau_2) - F(\theta|\tau_1)]d\theta| \\
&\leq \int_{\underline{\theta}}^{\bar{\theta}} |F(\theta|\tau_2) - F(\theta|\tau_1)|d\theta \\
&\leq \int_{\underline{\theta}}^{\bar{\theta}} \left|\frac{\partial F(\theta|\tilde{\tau})}{\partial \tau}\right| |\tau_2 - \tau_1|d\theta \\
&\leq K|\bar{\theta} - \underline{\theta}||\tau_2 - \tau_1|,
\end{aligned}$$

where in the final line we have used the assumption that $|\partial F(\theta|\tau)/\partial\tau| < K$. This shows (11.4) for $L = K|\bar{\theta} - \underline{\theta}|$. □

With the help of the previous lemma, we next obtain an envelope theorem with respect to the ex ante type, which enables us to show that the revenue equivalence property extends to our dynamic setting.

Proposition 11.4 *If a direct mechanism is incentive-compatible, then:*

(i) U is differentiable except in at most countably many points. At any point τ of differentiability:

$$U'(\tau) = \frac{\partial \hat{U}(\tau|\tau)}{\partial \tau} = -\int_{\underline{\theta}}^{\bar{\theta}} q(\tau, \hat{\theta}) \frac{\partial F(\hat{\theta}|\tau)}{\partial \tau} d\hat{\theta};$$

(ii) For every ex ante type τ we have

$$\begin{aligned}\int_{\underline{\theta}}^{\bar{\theta}} t(\tau,\hat{\theta})f(\hat{\theta}|\tau)\,d\hat{\theta} &= \int_{\underline{\theta}}^{\bar{\theta}} \hat{\theta}q(\tau,\hat{\theta})f(\hat{\theta}|\tau)\,d\hat{\theta}\\ &+ \int_{\underline{\theta}}^{\bar{\theta}} \{t(\underline{\tau},\hat{\theta}) - \hat{\theta}q(\underline{\tau},\hat{\theta})\}f(\hat{\theta}|\underline{\tau})\,d\hat{\theta}\\ &+ \int_{\underline{\tau}}^{\tau}\int_{\underline{\theta}}^{\bar{\theta}} q(\hat{\tau},\hat{\theta})\frac{\partial F(\hat{\theta}|\hat{\tau})}{\partial \tau}\,d\hat{\theta}\,d\hat{\tau}.\end{aligned}$$

Proof

Lemma 11.1 established that U is increasing and therefore differentiable almost everywhere. To determine its derivative, observe that from (11.3) and the boundedness of $F(\theta|\tau)$, we can conclude that $\hat{U}(\tau'|\tau)$ is differentiable in τ, since Lebesque's convergence theorem implies that we can exchange the order of integration and differentiation to obtain for all τ, τ':

$$\frac{\partial \hat{U}(\tau'|\tau)}{\partial \tau} = \frac{\partial}{\partial \tau}\int_{\underline{\theta}}^{\bar{\theta}} q(\tau',\theta)[1 - F(\theta|\tau)]d\theta = -\int_{\underline{\theta}}^{\bar{\theta}} q(\tau',\theta)\frac{\partial F(\theta|\tau)}{\partial \tau}d\theta. \tag{11.5}$$

Moreover, by definition of U and $\hat{U}$, we have for all τ and $\delta > 0$:

$$1/\delta \cdot [U(\tau) - U(\tau+\delta)] \leq 1/\delta \cdot [\hat{U}(\tau|\tau) - \hat{U}(\tau|\tau+\delta)]; \tag{11.6}$$
$$1/\delta \cdot [U(\tau-\delta) - U(\tau)] \geq 1/\delta \cdot [\hat{U}(\tau|\tau-\delta) - \hat{U}(\tau|\tau)]. \tag{11.7}$$

For any τ, where the derivative of U exists, the left hand sides of the previous inequalities converge to $U'(\tau)$, and the right hand sides converge to $\partial\hat{U}(\tau|\tau)/\partial\tau$. Putting the two inequalities together and using (11.5) establishes (i).

In order to show (ii) note that by Lemma 11.1, $U(\tau)$ is absolutely continuous so that we can write it as the integral of its derivative. This yields

$$U(\tau) = U(\underline{\tau}) + \int_{\underline{\tau}}^{\tau} U'(\hat{\tau})d\hat{\tau} = U(\underline{\tau}) - \int_{\underline{\tau}}^{\tau}\int_{\underline{\theta}}^{\bar{\theta}} q(\hat{\tau},\hat{\theta})\frac{\partial F(\hat{\theta}|\hat{\tau})}{\partial \tau}d\hat{\theta}d\hat{\tau}.$$

To obtain (ii), we rewrite this relationship by substituting out $U(\tau)$ and $U(\underline{\tau})$, since by definition,

$$U(\tau) = \int_{\underline{\theta}}^{\bar{\theta}} \{q(\tau,\theta)\theta - t(\tau,\theta)\}f(\theta|\tau)d\theta.$$

□

While Proposition 11.3 shows that incentive compatibility with respect to the ex post type θ implies that the transfer schedule $t(\tau,\theta)$ has only one remaining degree of freedom $t(\tau,\underline{\theta})$, Proposition 11.4 shows that incentive compatibility with respect to the ex ante type τ implies further restrictions on the transfer schedule. One may wonder whether these two restrictions on the transfer schedule do not lead to a contradiction. The next proposition shows that this is not the case. Analogously to the revenue equivalence property (Lemma 2.4), the two propositions taken together imply that, given a schedule $q(\tau,\theta)$, incentive compatibility pins down the transfer schedule $t(\tau,\theta)$ for all $(\tau,\theta) \in (\underline{\tau},\bar{\tau}] \times (\underline{\theta},\bar{\theta}]$ with the only remaining degree of freedom for the transfer $t(\underline{\tau},\underline{\theta})$.

Proposition 11.5 *If a direct mechanism is incentive-compatible, then*

$$t(\tau,\theta) = t_0(\tau) + \theta q(\tau,\theta) - \int_{\underline{\theta}}^{\theta} q(\tau,\hat{\theta})\, d\hat{\theta},$$

where

$$\begin{aligned} t_0(\tau) &= t(\underline{\tau},\underline{\theta}) - \underline{\theta} q(\underline{\tau},\underline{\theta}) + \int_{\underline{\tau}}^{\tau} \int_{\underline{\theta}}^{\bar{\theta}} q(\hat{\tau},\hat{\theta}) \frac{\partial F(\hat{\theta}|\hat{\tau})}{\partial \tau}\, d\hat{\theta}\, d\hat{\tau} \\ &\quad + \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\hat{\theta}} [q(\tau,x) f(\hat{\theta}|\tau) - q(\underline{\tau},x) f(\hat{\theta}|\underline{\tau})]\, dx d\hat{\theta}. \end{aligned}$$

Proof

Using the expression in Proposition 11.3 to substitute out $t(\tau,\theta)$ and $t(\underline{\tau},\theta)$ in Proposition 11.4 yields, after rearranging terms,

$$\begin{aligned} t(\tau,\underline{\theta}) &= t(\underline{\tau},\underline{\theta}) + \underline{\theta} q(\tau,\underline{\theta}) - \underline{\theta} q(\underline{\tau},\underline{\theta}) + \int_{\underline{\tau}}^{\tau} \int_{\underline{\theta}}^{\bar{\theta}} q(\hat{\tau},\hat{\theta}) \frac{\partial F(\hat{\theta}|\hat{\tau})}{\partial \tau}\, d\hat{\theta}\, d\hat{\tau} \\ &\quad + \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\hat{\theta}} [q(\tau,x) f(\hat{\theta}|\tau) - q(\underline{\tau},x) f(\hat{\theta}|\underline{\tau})]\, dx d\hat{\theta}. \end{aligned}$$

Using this expression to substitute out $t(\tau,\underline{\theta})$ in the expression in Proposition 11.3(iii) yields the result. □

As we have emphasized, incentive compatibility does not necessarily imply that the schedule $q(\tau,\theta)$ is increasing in τ. However, the next proposition shows that the first order stochastic dominance ranking of the conditional distributions implies that an increasing schedule is nevertheless sufficient for incentive compatibility.

Proposition 11.6 *If $q(\tau,\theta)$ is increasing in τ and θ, then there exists a transfer schedule $t(\tau,\theta)$ such that the direct mechanism (q,t) is incentive-compatible.*

Proof

Given an increasing $q(\tau,\theta)$, define $t(\tau,\theta)$ as in Proposition 11.5. Proposition 11.3 implies that the direct mechanism (q,t) satisfies (11.1) and is therefore, by definition, incentive-compatible with respect to the ex post type θ. Hence, we only need to prove (11.2); that is, that for all $\tau,\tau' \in [\underline{\tau},\bar{\tau}]$ we have

$$\Delta U = U(\tau) - \hat{U}(\tau'|\tau) \geq 0.$$

Note that

$$\begin{aligned}\Delta U = U(\tau) - \hat{U}(\tau'|\tau) &= \hat{U}(\tau|\tau) - \hat{U}(\tau'|\tau') + \hat{U}(\tau'|\tau') - \hat{U}(\tau'|\tau)\\ &= \int_{\tau'}^{\tau}\left[U'(\hat{\tau}) - \frac{\partial\hat{U}(\tau'|\hat{\tau})}{\partial\tau}\right]d\hat{\tau}.\end{aligned}$$

By construction, the transfers $t(\tau,\theta)$ as defined by Proposition 11.5 together with the schedule $q(\tau,\theta)$ pin down $U'(\hat{\tau})$ according to Proposition 11.4(i). Using similar steps as in the proof of Proposition 11.4, we use the definition of $\hat{U}(\tau'|\tau)$ and a subsequent integration by parts to rewrite the second term $\partial\hat{U}(\tau'|\tau)/\partial\tau$ in the squared brackets as

$$\begin{aligned}\frac{\partial\hat{U}(\tau'|\tau)}{\partial\tau} &= \int_{\underline{\theta}}^{\bar{\theta}} u(\tau',\hat{\theta})\frac{\partial f(\hat{\theta}|\tau)}{\partial\tau}\,d\hat{\theta}\\ &= \left[u(\tau',\hat{\theta})\frac{\partial F(\hat{\theta}|\tau)}{\partial\tau}\right]_{\underline{\theta}}^{\bar{\theta}} - \int_{\underline{\theta}}^{\bar{\theta}}\frac{\partial u(\tau',\hat{\theta})}{\partial\theta}\frac{\partial F(\hat{\theta}|\tau)}{\partial\tau}\,d\hat{\theta}\\ &= -\int_{\underline{\theta}}^{\bar{\theta}} q(\tau',\hat{\theta})\frac{\partial F(\hat{\theta}|\tau)}{\partial\tau}\,d\hat{\theta}.\end{aligned}$$

As a result, ΔU can be rewritten as

$$\Delta U = \int_{\tau'}^{\tau}\int_{\underline{\theta}}^{\bar{\theta}}[q(\tau',\hat{\theta}) - q(\hat{\tau},\hat{\theta})]\frac{\partial F(\hat{\theta}|\hat{\tau})}{\partial\tau}\,d\hat{\theta}\,d\hat{\tau}.$$

For $\tau > \tau'$, it follows that the obtained expression of ΔU is nonnegative because, due to FOSD, the partial derivative $\partial F(\hat{\theta}|\hat{\tau})/\partial\tau$ is negative; and since $q(\tau,\theta)$ is increasing in τ, the difference in the square brackets is also negative for all $\hat{\tau} \in [\tau',\tau]$. For $\tau' > \tau$, the expression is also nonnegative, because in this case the difference in the square brackets is positive for all $\hat{\tau} \in [\tau,\tau']$. □

We next turn to a characterization of individual rationality.

Proposition 11.7 *An incentive-compatible direct mechanism is individually rational if and only if* $U(\underline{\tau}) \geq 0$.

Proof

"$\Rightarrow$" is obvious. To see "$\Leftarrow$", note that, due to FOSD, $\partial F(\theta|\tau)/\partial\tau \leq 0$ so that $U(\tau)$ is increasing by Proposition 11.4. Hence, $U(\tau) \geq 0$ for all $\tau \in [\underline{\tau}, \bar{\tau}]$ if and only if $U(\underline{\tau}) \geq 0$.

□

We can now determine the optimal selling mechanism. Because of revenue equivalence, the seller's objective is a function of the allocation rule only. More precisely, given an incentive-compatible direct mechanism and a type (τ, θ), the seller's revenue is the difference between the generated surplus $\theta q(\tau, \theta)$ and the buyer's utility $u(\tau, \theta)$. The seller's expected revenue is therefore

$$\begin{aligned}&\int_{\underline{\tau}}^{\bar{\tau}} \int_{\underline{\theta}}^{\bar{\theta}} [\hat{\theta} q(\hat{\tau}, \hat{\theta}) - u(\hat{\tau}, \hat{\theta})] f(\hat{\theta}|\hat{\tau}) g(\hat{\tau})\, d\hat{\theta} d\hat{\tau} \\ &\quad = \int_{\underline{\tau}}^{\bar{\tau}} \int_{\underline{\theta}}^{\bar{\theta}} \hat{\theta} q(\hat{\tau}, \hat{\theta}) f(\hat{\theta}|\hat{\tau}) g(\hat{\tau})\, d\hat{\theta} d\hat{\tau} - \int_{\underline{\tau}}^{\bar{\tau}} U(\hat{\tau}) g(\hat{\tau})\, d\hat{\tau}.\end{aligned}$$

Integration by parts and Proposition 11.4 imply that the buyer's expected utility is

$$\begin{aligned}\int_{\underline{\tau}}^{\bar{\tau}} U(\hat{\tau}) g(\hat{\tau}) d\hat{\tau} &= -\left[(1 - G(\hat{\tau})) U(\hat{\tau})\right]_{\underline{\tau}}^{\bar{\tau}} + \int_{\underline{\tau}}^{\bar{\tau}} (1 - G(\hat{\tau})) U'(\hat{\tau})\, d\hat{\tau} \\ &= U(\underline{\tau}) - \int_{\underline{\tau}}^{\bar{\tau}} \int_{\underline{\theta}}^{\bar{\theta}} (1 - G(\hat{\tau})) q(\hat{\tau}, \hat{\theta}) \frac{\partial F(\hat{\theta}|\hat{\tau})}{\partial \tau}\, d\hat{\theta}\, d\hat{\tau}.\end{aligned}$$

We therefore can rewrite the seller's expected revenue as

$$\int_{\underline{\tau}}^{\bar{\tau}} \int_{\underline{\theta}}^{\bar{\theta}} \left[\hat{\theta} + \frac{1 - G(\hat{\tau})}{g(\hat{\tau})} \frac{\partial F(\hat{\theta}|\hat{\tau})/\partial\tau}{f(\hat{\theta}|\hat{\tau})}\right] q(\hat{\tau}, \hat{\theta}) f(\hat{\theta}|\hat{\tau}) g(\hat{\tau})\, d\hat{\theta} d\hat{\tau} - U(\underline{\tau}). \qquad (11.8)$$

When the seller picks an incentive-compatible and individually rational direct mechanism (q, t), her expected revenue coincides with expression (11.8). This means that if we find a schedule q together with a value $U(\underline{\tau})$ that maximizes the expression and, in addition, a transfer schedule t such that the associated direct mechanism (q, t) is incentive-compatible and individually rational, then the direct mechanism (q, t) is an optimal mechanism for the seller. Maximizing (11.8) while guaranteeing individual rationality clearly requires $U(\underline{\tau}) = 0$. Because (11.8) is linear in $q(\tau, \theta)$, a schedule maximizes the expression only if $q(\tau, \theta) = 0$ when the term in the square brackets is (strictly) negative and $q(\tau,\theta) = 1$ when the term is (strictly) positive. Hence, if we define

$$\psi(\tau,\theta) = \theta + \frac{1 - G(\tau)}{g(\tau)} \frac{\partial F(\theta|\tau)/\partial\tau}{f(\theta|\tau)} \qquad \text{for all } \tau \in [\underline{\tau},\bar{\tau}],\ \theta \in [\underline{\theta},\bar{\theta}],$$

then the argument above leads us to consider the following allocation rule:

$$q(\tau,\theta) = \begin{cases} 1 & \text{if } \psi(\tau,\theta) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{11.9}$$

According to Propositions 11.3–11.6 and by choosing $t(\underline{\tau},\underline{\theta})$ so that $U(\underline{\tau}) = 0$, there exists a transfer t that makes the allocation rule incentive-compatible if the allocation rule is increasing in both arguments. In the spirit of Assumption 2.1, we make a regularity assumption on the distribution functions $G(\tau)$ and $F(\theta|\tau)$, which ensures that (11.9) is indeed increasing:

Assumption 11.1 *$\psi(\tau,\theta)$ is increasing in τ and θ.*

Assumption 11.1 is the dynamic counterpart of the regularity Assumption 2.1 in the static screening context of Chapter 2. The modified regularity assumption reflects the dynamic revelation of private information: The hazard rate of the ex ante types, $(1 - G(\tau))/g(\tau)$, is now multiplied by the fraction

$$-\frac{\partial F(\theta|\tau)/\partial\tau}{f(\theta|\tau)},$$

which is sometimes referred to as an "informativeness measure." It captures how the buyer's private knowledge about the distribution of his valuation changes with his ex ante type.

Our assumptions guarantee that $\psi(\tau,\theta)$ is continuous in (τ,θ) so that

$$p(\tau) = \min\{\hat{\theta} \in [\underline{\theta},\bar{\theta}] \mid \psi(\hat{\theta},\tau) \geq 0\}$$

is well-defined. Then under Assumption 11.1, $p(\tau)$ is decreasing in τ and, moreover, $\psi(\tau,\theta) \geq 0$ for all $\theta \geq p(\tau)$ and $\psi(\tau,\theta) < 0$ for all $\theta < p(\tau)$.

Proposition 11.8 *Suppose Assumption 11.1 holds. Then an incentive-compatible and individually rational direct mechanism (q,t) is optimal if and only if*

$$q(\tau,\theta) = \begin{cases} 1 & \textit{if } \theta \geq p(\tau), \\ 0 & \textit{otherwise} \end{cases} \tag{11.10}$$

and

$$t(\tau,\theta) = \begin{cases} t_0(\tau) + p(\tau) & \text{if } \theta \geq p(\tau), \\ t_0(\tau) & \text{otherwise,} \end{cases} \tag{11.11}$$

where $t_0(\tau)$ is given by Proposition 11.5 and

$$t(\underline{\tau},\underline{\theta}) = \int_{p(\underline{\tau})}^{\bar{\theta}} \hat{\theta} f(\hat{\theta}|\underline{\tau})\, d\hat{\theta} - p(\underline{\tau})[1 - F(p(\underline{\tau})|\underline{\tau})] + \underline{\theta} q(\underline{\tau},\underline{\theta}). \tag{11.12}$$

Proof

Consider a pair (τ,θ) such that $\theta < p(\tau)$; then $q(\tau,\hat{\theta}) = 0$ for all $\hat{\theta} \in [\underline{\theta},\theta]$. Hence, from Proposition 11.5 we obtain

$$t(\tau,\theta) = t_0(\tau) + \theta q(\tau,\theta) - \int_{\underline{\theta}}^{\theta} q(\tau,\hat{\theta})\, d\hat{\theta} = t_0(\tau).$$

Consider a pair (τ,θ) such that $\theta \geq p(\tau)$; then $q(\tau,\hat{\theta}) = 1$ for all $\hat{\theta} \in [p(\tau),\theta]$. Hence, from Proposition 11.5 it follows that

$$t(\tau,\theta) = t_0(\tau) + \theta - \int_{p(\tau)}^{\theta} 1\, d\hat{\theta} = t_0(\tau) + \theta - (\theta - p(\tau)) = t_0(\tau) + p(\tau).$$

Finally note that expressions (11.10) and (11.11) imply

$$\begin{aligned} U(\underline{\tau}) &= \int_{\underline{\theta}}^{\bar{\theta}} [\hat{\theta} q(\underline{\tau},\hat{\theta}) - t(\underline{\tau},\hat{\theta})] f(\hat{\theta}|\underline{\tau})\, d\hat{\theta} \\ &= \int_{p(\underline{\tau})}^{\bar{\theta}} \hat{\theta} f(\hat{\theta}|\underline{\tau})\, d\hat{\theta} - p(\underline{\tau})[1 - F(p(\underline{\tau})|\underline{\tau})] - t_0(\underline{\tau}). \end{aligned}$$

Hence, $U(\underline{\tau}) = 0$, and the expression for $t_0(\underline{\tau})$ in Proposition 11.5 imply (11.12). □

The optimal selling mechanism can be implemented indirectly, without asking the buyer to announce her type, by a menu of buy option contracts. Under a buy option contract, the buyer acquires in period 1 for a fee the option to buy the good in period 2 at a prespecified "exercise" price. A menu of option contracts is a set of option contracts indexed by τ.

The menu of option contracts $(t_0(\tau), p(\tau))_{\tau \in [\underline{\tau},\bar{\tau}]}$ implements the optimal selling mechanism from Proposition 11.8, because, first, if the buyer has picked an option contract in period 1, she exercises the option in period 2 if and only if her type θ exceeds the exercise price. Thus, the optimal trading rule is implemented. Second, in period 1, the option contract $(t_0(\hat{\tau}), p(\hat{\tau}))$ gives the buyer the same expected utility as announcing $\hat{\tau}$ under the optimal, direct selling mechanism. Therefore, the menu of option contracts implements the same outcome as the optimal, direct selling mechanism.

11.2.2 The Role of Private Information

The optimal sequential selling mechanism leaves information rents to the buyer. Since, in a sequential framework, the buyer observes two pieces of private information, one may ask to what extent the buyer's ex ante and ex post private information contribute to her information rents. We will see that information rents accrue only from the buyer's ex ante private information. In contrast, any additional private information that the buyer receives once the contract is signed can be extracted by the seller at no cost.

To make this claim more precise, we first formalize what we mean by additional information. In the model considered so far, the buyer is in period 2 informed about her payoff type. The payoff type contains the total information the buyer receives and, therefore, contains more than only the additional information the buyer receives in period 2. Intuitively, the buyer's additional information is the "difference" in the information contained in the buyer's payoff type and her ex ante signal. This is literally the case if the buyer's initial signal is her true payoff type plus noise. In this case, the additional information the buyer observes in period 2 would correspond to the noise term.

We now show that, more generally, we can write any model in such a way that the buyer's payoff type is a compound of a first signal and an additional second and independent signal. Indeed, suppose that instead of observing her payoff type θ directly in period 2, the buyer with ex ante signal τ observes the realization of the signal $\gamma = F(\theta|\tau)$ in period 2. Since $F(\theta|\tau)$ is strictly increasing in θ, the buyer's payoff type can be backed out through $\theta = F^{-1}(\gamma|\tau)$, and therefore the new model formulation in terms of (τ, γ) is isomorphic to the formulation in terms of (τ, θ).

By construction, the signal γ is uniformly distributed on the unit interval for all τ and hence stochastically independent of τ. To see this, observe that $Pr(\gamma < \gamma_0|\tau) = Pr(\theta < F^{-1}(\gamma_0|\tau)|\tau) = F(F^{-1}(\gamma_0|\tau)|\tau) = \gamma_0$. Because of independence, the signal γ captures the buyer's additional ex post information that she receives beyond any information already contained in her first signal. A mechanism in this alternative representation of the model is a quantity transfer schedule $(\tilde{q}(\tau, \gamma), \tilde{t}(\tau, \gamma))$ and requires the buyer to report τ and γ rather than τ and θ.

We will now show that the seller can extract the signal γ at no cost. For this, we compare the environment in which γ is the buyer's private information to the hypothetical situation in which γ is publicly observable and thus need not be elicited from the buyer. We show that the seller attains the same payoff in both situations. This leads us to conclude that the privacy of the additional information γ does not reduce the seller's profit and does therefore not constitute a source of information rents for the buyer.

We begin by considering the model in which γ is private information. This amounts to considering a simple transformation of our original model formulated in terms of (τ, θ). Under Assumption 11.1, an optimal direct mechanism in the (τ, θ)-formulation is such that $q(\tau, \theta)$ maximizes the seller's objective

$$\int_{\underline{\tau}}^{\bar{\tau}} \int_{\underline{\theta}}^{\bar{\theta}} \psi(\tau, \theta) q(\tau, \theta) f(\theta | \tau) g(\tau) \, d\theta \, d\tau. \tag{11.13}$$

Because $\gamma = F(\theta|\tau)$, we have that $f(\theta|\tau) \, d\theta = d\gamma$; and after a change of variables, (11.13) becomes

$$\int_{\underline{\tau}}^{\bar{\tau}} \int_{0}^{1} \psi(\tau, F^{-1}(\gamma|\tau)) q(\tau, F^{-1}(\gamma|\tau)) g(\tau) \, d\gamma \, d\tau. \tag{11.14}$$

Setting $\tilde{q}(\tau, \gamma) = q(\tau, F^{-1}(\gamma|\tau))$ and

$$\tilde{\psi}(\tau, \gamma) = \psi(\tau, F^{-1}(\gamma|\tau)), \tag{11.15}$$

we can rewrite the seller's objective as

$$\int_{\underline{\tau}}^{\bar{\tau}} \int_{0}^{1} \tilde{\psi}(\tau, \gamma) \tilde{q}(\tau, \gamma) g(\tau) \, d\gamma \, d\tau. \tag{11.16}$$

Clearly, if a schedule $q^*(\tau, \theta)$ maximizes (11.13), then the schedule $\tilde{q}^*(\tau, \gamma) = q^*(\tau, F^{-1}(\gamma|\tau))$ maximizes (11.16). Conversely, if a schedule $\tilde{q}^*(\tau, \gamma)$ maximizes (11.16) and the schedule $q^*(\tau, \theta)$ is such that $\tilde{q}^*(\tau, \gamma) = q^*(\tau, F^{-1}(\gamma|\tau))$, then $q^*(\tau, \theta)$ maximizes (11.13).

For future reference, we compute $\tilde{\psi}$ explicitly. By definition,

$$\tilde{\psi}(\tau, \gamma) = F^{-1}(\gamma|\tau) + \frac{1 - G(\tau)}{g(\tau)} \frac{\partial F(F^{-1}(\gamma|\tau)|\tau)/\partial \tau}{f(F^{-1}(\gamma|\tau)|\tau)}.$$

We can simplify the second ratio by computing the *total* derivative

$$\begin{aligned} \frac{d}{d\tau} F(F^{-1}(\gamma|\tau)|\tau) = & f(F^{-1}(\gamma|\tau)|\tau) \cdot \frac{\partial F^{-1}(\gamma|\tau)}{\partial \tau} \\ & + \frac{\partial F}{\partial \tau}(F^{-1}(\gamma|\tau)|\tau). \end{aligned}$$

On the other hand, since $F(F^{-1}(\gamma|\tau)|\tau) = \gamma$ does not depend on τ, the previous derivative is zero. Hence,

$$\frac{\partial F(F^{-1}(\gamma|\tau)|\tau)/\partial\tau}{f(F^{-1}(\gamma|\tau)|\tau)} = -\frac{\partial F^{-1}(\gamma|\tau)}{\partial\tau},$$

and we obtain that

$$\tilde{\psi}(\tau,\gamma) = F^{-1}(\gamma|\tau) - \frac{1 - G(\tau)}{g(\tau)}\frac{\partial F^{-1}(\gamma|\tau)}{\partial\tau}. \tag{11.17}$$

We now consider the model mentioned earlier where the additional information γ is public information. More precisely, we assume that the buyer, before contracting, observes τ privately; but after the contract is signed, the signal γ is revealed publicly. For this alternative information structure, we want to derive the seller's optimal selling mechanism and compare it to the one in the sequential screening model, where the buyer learns γ privately.

When γ is observable, a selling mechanism can condition directly on γ. Following the revelation principle, the seller's optimal mechanism is therefore in the class of direct mechanisms $(\tilde{q}(\tau,\gamma), \tilde{t}(\tau,\gamma))$ which induce the buyer to truthfully reveal τ and condition directly on the true γ. We call such a mechanism *incentive-compatible with observable* γ. Recall that in the (γ,τ)-formulation of the model, the buyer's payoff type is $\theta = F^{-1}(\gamma|\tau)$ and the signal γ is uniformly distributed. Therefore, if the buyer type τ reveals her type truthfully, she obtains expected utility

$$U(\tau) = \int_0^1 [F^{-1}(\gamma|\tau)\tilde{q}(\tau,\gamma) - \tilde{t}(\tau,\gamma)]\, d\gamma.$$

Consequently, the selling mechanism $(\tilde{q}(\tau,\gamma), \tilde{t}(\tau,\gamma))$ is incentive-compatible with observable γ if for all $\tau' \in [\underline{\tau}, \bar{\tau}]$ we have

$$U(\tau) \geq \int_0^1 [F^{-1}(\gamma|\tau)\tilde{q}(\tau',\gamma) - \tilde{t}(\tau',\gamma)]\, d\gamma.$$

The incentive compatibility constraints pin down the seller's utility as stated in the next proposition.

Proposition 11.9 *If a direct mechanism* $(\tilde{q}(\tau,\gamma), \tilde{t}(\tau,\gamma))$ *is incentive-compatible with observable* γ*, then*

$$U(\tau) = U(\underline{\tau}) + \int_{\underline{\tau}}^{\tau}\int_0^1 \frac{\partial F^{-1}(\gamma|\hat{\tau})}{\partial\tau}\tilde{q}(\hat{\tau},\gamma)\, d\gamma\, d\hat{\tau}.$$

Proof

Incentive compatibility implies that a buyer of type $\tau + \delta$ prefers truth telling over reporting type τ. Hence, for $\delta > 0$, we have

$$\lim_{\delta \to 0} \frac{U(\tau + \delta) - U(\tau)}{\delta} \geq \lim_{\delta \to 0} \left(\int_0^1 \frac{F^{-1}(\gamma|\tau + \delta) - F^{-1}(\gamma|\tau)}{\delta} \tilde{q}(\tau, \gamma)\, d\gamma \right)$$
$$= \int_0^1 \frac{\partial F^{-1}(\gamma|\tau)}{\partial \tau} \tilde{q}(\tau, \gamma)\, d\gamma.$$

Similarly for $\delta > 0$, we have

$$\lim_{\delta \to 0} \frac{U(\tau) - U(\tau - \delta)}{\delta} \leq \int_0^1 \frac{\partial F^{-1}(\gamma|\tau)}{\partial \tau} \tilde{q}(\tau, \gamma)\, d\gamma.$$

Hence, $U(\cdot)$ is differentiable and

$$U'(\tau) = \int_0^1 \frac{\partial F^{-1}(\gamma|\tau)}{\partial \tau} \tilde{q}(\tau, \gamma)\, d\gamma.$$

The claim now follows by integration. □

Equipped with this result, we can now determine the optimal mechanism for the case that γ is observable. Recall that the seller's payoff is the difference between the surplus and the buyer's rents:

$$\int_{\underline{\tau}}^{\bar{\tau}} \int_0^1 F^{-1}(\gamma|\tau)\tilde{q}(\tau, \gamma) g(\tau)\, d\gamma\, d\tau - \int_{\underline{\tau}}^{\bar{\tau}} U(\tau) g(\tau)\, d\tau. \tag{11.18}$$

We can rewrite this expression by applying the same steps as in the derivation of expression (11.8) where, instead of Proposition 11.4, we now use Proposition 11.9. The seller's payoff thus becomes

$$\int_{\underline{\tau}}^{\bar{\tau}} \int_0^1 \left[F^{-1}(\gamma|\tau) - \frac{1 - G(\tau)}{g(\tau)} \frac{\partial F^{-1}(\gamma|\tau)}{\partial \tau} \right] \tilde{q}(\tau, \gamma) g(\tau)\, d\gamma\, d\tau - U(\underline{\tau}),$$

which, by (11.17), is equal to

$$\int_{\underline{\tau}}^{\bar{\tau}} \int_0^1 \tilde{\psi}(\tau, \gamma)\tilde{q}(\tau, \gamma) g(\tau)\, d\gamma\, d\tau - U(\underline{\tau}), \tag{11.19}$$

Hence, if $\tilde{q}(\tau, \gamma)$ together with $U(\underline{\tau}) = 0$ maximizes this expression and there exists a transfer schedule $\tilde{t}(\tau, \gamma)$ such that the direct mechanism is incentive-compatible with observable γ and satisfies $U(\underline{\tau}) = 0$, then it is optimal.

Note, however, that (for $U(\underline{\tau}) = 0$), expression (11.19) coincides with (11.16), the seller's objective when γ is private information. We know from our earlier considerations that under Assumption 11.1, there are transfers $\tilde{t}^*(\tau, \gamma)$ which together with the maximizer $\tilde{q}^*(\tau, \gamma)$ of (11.16) form the optimal selling mechanism when γ is private information. A fortiori, this mechanism is incentive-compatible when γ is public information. Since the seller's objective is the same when γ is private and when it is public information and since, in addition, the objective is independent of transfers, it follows that $(\tilde{q}^*(\tau, \gamma), \tilde{t}^*(\tau, \gamma))$ is also an optimal selling mechanism when γ is public information. We thus have proven the next proposition.

Proposition 11.10 *Suppose Assumption 11.1 holds. Then if the mechanism* $(\tilde{q}(\tau, \gamma), \tilde{t}(\tau, \gamma))$ *is optimal when* γ *is privately observable, it is also optimal when* γ *is publicly observable.*

Proposition 11.10 implies that the seller and the buyer obtain the same expected payoff irrespective of whether γ is private or public information. As a result, the seller can extract all ex post information that the buyer receives beyond what she knows at the contracting stage. In a static environment, this is a well-known result: If the buyer does not have private information at the contracting stage and only receives private information ex post, the seller does not need to concede any information rents to the buyer and can extract all gains from trade. Proposition 11.10 shows that this insight extends to a framework of dynamic private information, provided that Assumption 11.1 holds.

We point out that the extension does not hold in general and may fail when Assumption 11.1 is violated. In particular, in a setting where the optimal schedule $\tilde{q}(\tau, \gamma)$ with publicly observable γ is not monotone in γ so that the associated schedule $q(\tau, \theta) = \tilde{q}(\tau, F(\theta|\tau))$ is not monotone in θ, Proposition 11.3(i) implies that it is not implementable when γ is privately observed ex post. In this case, the optimal mechanism when γ is publicly observable differs from the optimal mechanism when γ is private information. The extension may also fail when ex ante types are discrete (see Problem 11.5,(c)).

Moreover, the reverse of Proposition 11.10 is generally not true, because when γ is publicly observable, incentive compatibility pins down the transfers only in expectation with respect to γ. This means that there are transfer schedules that are part of an optimal mechanism when γ is observable, which fail to be optimal when γ is private, because they lead to a violation of the ex post incentive compatibility constraints. Intuitively, incentive compatibility is less demanding with publicly observable γ, as it leaves an extra degree of freedom, but, under Assumption 11.1, this degree of freedom does not translate in a larger payoff for the principal.

Proposition 11.10 has an important implication for the seller's incentives to disclose additional information. Consider a setting where the seller can commit to take measures that improve the buyer's ex post private information, for example by disclosing information about the good whose value for the buyer only the buyer can assess. Then Proposition 11.10 implies that the seller wants to disclose as much information as possible. This is so because the seller can at no cost fully extract the gain in surplus that the availability of new information entails.

Proposition 11.10 is an illustration of the general insight that the seller would like to contract as early as possible with the buyer when the buyer still has relatively little private information. In this way the seller can reduce information rents because when contracting early the incentive compatibility and participation constraints need to hold only on average rather than for any realization of the buyer's ex post information. Under the conditions of Proposition 11.10, this effect is so strong that by contracting before the buyer observes additional information the seller can eliminate all additional rents that the buyer would otherwise receive when contracting took place later.

11.2.3 Sequential Mechanism Design

In this section, we consider the same model as in the previous section except that we now have more than one potential buyer. This extends the analysis of the unit auction from Section 3.2 to the case that the buyers' private information about their valuations arrives sequentially.

There is now a set $I = \{1, \ldots, N\}$ of potential buyers where buyer i's utility when she gets the good and pays a transfer t_i is $\theta_i - t_i$, and her utility when she does not get the good is $0 - t_i$. The seller's utility when she gets transfers t_i from buyers $i = 1, \ldots, N$ is $\sum_{i \in I} t_i$.

Initially, each buyer i privately observes the signal τ_i which is distributed with cumulative distribution $G_i(\tau_i)$ and density function $g_i(\tau_i) > 0$ on the support $[\underline{\tau}, \bar{\tau}]$. Only after the buyer has accepted the selling mechanism proposed by the seller, buyer i observes her actual valuation θ_i. We assume that conditional on τ_i, the valuation θ_i is distributed with cumulative distribution $F_i(\theta_i|\tau_i)$ and density function $f_i(\theta_i|\tau_i) > 0$ on the support $[\underline{\theta}, \bar{\theta}]$ where $0 \leq \underline{\theta} < \bar{\theta}$. As in the single buyer case, we assume that the conditional distributions of buyer i's valuation are ordered in the sense of first-order stochastic dominance, that is, $\partial F_i(\theta_i|\tau_i)/\partial \tau_i < 0$. Moreover, we assume that all distributions are stochastically independent across buyers i. Note that we assume that the supports of τ_i and θ_i are independent of i. This is for convenience only.

We denote by $\tau = (\tau_1, \ldots, \tau_N)$ and $\theta = (\theta_1, \ldots, \theta_N)$ the vectors of types and define $\mathcal{T} = [\underline{\tau}, \bar{\tau}]^N$, $\Theta = [\underline{\theta}, \bar{\theta}]^N$. Moreover, we write $F(\theta|\tau)$ (resp. $G(\tau)$) for the distribution of θ conditional on τ (resp. of τ).

The seller's objective is to design a selling mechanism that maximizes her expected revenue. As in Section 3.2 and in the single buyer case, the revelation principle applies, and we can restrict attention to direct and incentive-compatible mechanisms. A direct mechanism consists of functions q and t_i (for $i \in I$) where

$$q : \mathcal{T} \times \Theta \to \Delta$$

and

$$t_i : \mathcal{T} \times \Theta \to \mathbb{R}.$$

Here, the allocation rule $q(\tau,\theta) = (q_1(\tau,\theta), \ldots, q_N(\tau,\theta))$ describes the probability with which each buyer i gets the good, and $1 - \sum_{i\in I} q_i(\tau,\theta)$ is the probability with which the good is not sold. The set Δ is defined in the same way as in Section 3.2.2. As in Section 3.2.2, we define the expected probability that buyer i gets the good by

$$Q_i(\tau_i,\theta_i) = \int_{\mathcal{T}_{-i}} \int_{\Theta_{-i}} q_i(\tau_i,\theta_i,\tau_{-i},\theta_{-i})\ dF_{-i}(\theta_{-i}|\tau_{-i}) dG_{-i}(\tau_{-i})$$

and buyer i's expected payment by

$$T_i(\tau_i,\theta_i) = \int_{\mathcal{T}_{-i}} \int_{\Theta_{-i}} t_i(\tau_i,\theta_i,\tau_{-i},\theta_{-i})\ dF_{-i}(\theta_{-i}|\tau_{-i}) dG_{-i}(\tau_{-i}),$$

where $G_{-i}(\tau_{-i}) = \prod_{j\in I\setminus\{i\}} G_j(\tau_j)$ is the marginal distribution of ex ante types τ_{-i} excluding type τ_i. The definition of incentive compatibility carries over from Definition 11.2, where we only have to replace q by Q_i and t by T_i and add the buyer index i to (τ,θ) and u. Analogously, Propositions 11.3–11.6 carry over with the appropriate notational modifications. In particular, the multiple buyer version of Proposition 11.5 pins down the expected transfers T_i as a function of the expected winning probabilities Q_i up to the expected transfer $T_i(\underline{\tau},\underline{\theta})$ for the lowest type. Therefore, the same remark as after Proposition 3.2 applies: The seller's choice variables are the allocation rule q and the payment $T_i(\underline{\tau},\underline{\theta})$ by the lowest type.

Analogously to the single buyer case, the expected payments the seller collects from buyer i under an incentive-compatible mechanism can be computed as

$$\int_{\underline{\tau}}^{\bar{\tau}} \int_{\underline{\theta}}^{\bar{\theta}} \psi_i(\tau_i,\theta_i) Q_i(\tau_i,\theta_i)\ dF_i(\theta_i|\tau_i) dG_i(\tau_i) - U_i(\underline{\tau}),$$

where

$$\psi_i(\tau_i,\theta_i) = \theta_i + \frac{1 - G_i(\tau_i)}{g_i(\tau_i)} \frac{\partial F_i(\theta_i|\tau_i)/\partial\tau_i}{f_i(\theta_i|\tau_i)}$$

is buyer i's virtual valuation, and $U_i(\underline{\tau})$ is buyer i's expected utility as seen from the ex ante stage when her type is $\underline{\tau}$. Therefore, summing over all buyers and expanding Q_i delivers the seller's total expected revenue:

$$\sum_{i\in I}\left[\int_{\mathcal{T}}\int_{\Theta}\psi_i(\tau_i,\theta_i)q_i(\tau,\theta)\ dF(\theta|\tau)dG(\tau)\right]-\sum_{i\in I}U_i(\underline{\tau}).$$

If we find an allocation rule $q = (q_1, \ldots, q_N)$ together with values $U_i(\underline{\tau})$ that maximize this expression and, in addition, transfer $T = (T_1, \ldots, T_N)$ that make the allocation rule incentive-compatible and individually rational, then we have found an optimal mechanism for the seller. Analogously to Proposition 11.7, individual rationality is equivalent to granting the lowest buyer type a positive utility, and hence maximizing the seller's objective while guaranteeing individual rationality requires $U_i(\underline{\tau}) = 0$. Moreover, as a weighted sum of the virtual valuations ψ_i, the seller's objective is maximized when full weight is placed on the maximal virtual valuation, provided that it is positive. If none of the virtual valuations is positive, the seller optimally retains the good. Thus, the allocation rule which specifies for all $i \in I$ that

$$q_i(\tau,\theta)=\begin{cases}1 & \text{if } \psi_i(\tau_i,\theta_i)>0, \text{ and } \psi_i(\tau_i,\theta_i)>\psi_j(\tau_j,\theta_j) \text{ for all } j\neq i,\\ 0 & \text{otherwise}\end{cases}$$

maximizes the seller's objective. According to Propositions 11.3–11.6, there is an expected transfer scheme T that makes the allocation rule q incentive-compatible if the induced probability of winning for buyer i, $Q_i(\tau_i, \theta_i)$, is increasing in both arguments. Similarly to the single buyer case, we impose a regularity assumption which ensures that this is the case.

Assumption 11.2 *$\psi_i(\tau_i, \theta_i)$ is increasing in τ_i and θ_i.*

It can be easily verified that Assumption 11.2 implies that the expected winning probability $Q_i(\tau_i, \theta_i)$ is increasing in both arguments. Thus, we can conclude the following:

Proposition 11.11 *Suppose Assumption 11.2 holds. Then an incentive-compatible and individually rational direct mechanism $\{(q_i, t_i)\}_i$ is optimal if and only if for all $i \in I$ and all $(\tau, \theta) \in \mathcal{T} \times \Theta$:*

(i) $$q_i(\tau,\theta)=\begin{cases}1 & \textit{if } \psi_i(\tau_i,\theta_i)>0, \textit{ and } \psi_i(\tau_i,\theta_i)>\psi_j(\tau_j,\theta_j) \textit{ for all } j\neq i,\\ 0 & \textit{otherwise}\end{cases}$$

(ii) $T_i(\tau_i, \theta_i) = T_{i0}(\tau_i) + \theta_i Q_i(\tau_i, \theta_i) - \int_{\underline{\theta}}^{\theta_i} Q_i(\tau_i, x)\, dx,$

where, analogously to Proposition 11.5, we have

$$T_{i0}(\tau_i) = T_i(\underline{\tau}, \underline{\theta}) - \underline{\theta} Q_i(\underline{\tau}, \underline{\theta}) + \int_{\underline{\tau}}^{\tau_i} \int_{\underline{\theta}}^{\bar{\theta}} Q_i(\hat{\tau}, \hat{\theta}) \frac{\partial F_i(\hat{\theta}|\hat{\tau})}{\partial \tau_i}\, d\hat{\theta} d\hat{\tau}$$
$$+ \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\theta}}^{\hat{\theta}} [Q_i(\tau_i, x) f_i(\hat{\theta}|\tau) - Q_i(\underline{\tau}, x) f_i(\hat{\theta}|\underline{\tau})]\, dx d\hat{\theta},$$

and $T_i(\underline{\tau}, \underline{\theta})$ is pinned down by the condition that the expected utility of the lowest type $\underline{\tau}$ is zero.

As with Proposition 3.4, 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.

In the special case that for all $i \in I$, the expression

$$\frac{\partial F_i(\theta_i|\tau_i)/\partial \tau_i}{f_i(\theta_i|\tau_i)}$$

is equal to some function $\phi_i(\tau_i)$ that does not depend on θ_i, the optimal mechanism can be implemented by an indirect mechanism that is similar to the menu of option contracts presented in the single buyer case. The virtual valuation is then of the form

$$\psi_i(\tau_i, \theta_i) = \theta_i + \frac{1 - G_i(\tau_i)}{g_i(\tau_i)} \phi_i(\tau_i).$$

The indirect mechanism is a so-called handicap auction that consists of two rounds: In the first round, each buyer can acquire, for a fee $t_{i0}(\tau_i)$, a premium $p_i(\tau_i)$ from a set of premia $\{p_i(\tau_i)|\tau_i \in \mathcal{T}\}$. In the second round, buyers bid in a second price auction without reserve price with the twist that the winning buyer has to pay not only the second highest bid but also the premium she acquired in the first round. In the second round, the buyer's effective valuation is thus equal to her true valuation minus the acquired premium, and therefore it is a weakly dominant strategy for each buyer to bid her valuation minus the premium. Hence, by setting

$$p_i(\tau_i) = -\frac{1 - G_i(\tau_i)}{g_i(\tau_i)} \phi_i(\tau_i),$$

the optimal allocation rule is implemented, provided that buyer i with type τ_i finds it optimal to buy the premium $p_i(\tau_i)$ in the first round. For this to be the case, we only need to set the ex ante fee $t_{i0}(\tau_i)$ so that the resulting expected payments for buyer i with type τ_i are the same as the expected payments under the optimal, direct selling mechanism.

11.3 DYNAMIC ALLOCATIONS

In the previous section, we extended the static mechanism design approach toward considering dynamic asymmetric information and saw that an optimal mechanism makes use of all the new information that is received by the agent over time. In this section, we extend the static approach by considering dynamic allocations rather than dynamic asymmetric information. In this case, the buyer's behavior in early periods reveals information about her type to which the mechanism may respond in later periods. However, the key insight in this section will be that, with dynamic allocations, the optimal mechanism will not react to new information that is revealed by the buyer's behavior. Instead, the optimal mechanism will be a repetition of the optimal static mechanism.

To show this, consider a seller who, in each period $\tau = 1, \ldots, T$, seeks to sell one indivisible good to a buyer with some valuation $\theta \geq 0$. Since we want to focus on the impact of dynamic allocations on mechanism design, we intentionally fix the private information as static. Consequently, we assume that the buyer's valuation does not change over the periods.

As in Chapter 2, let F denote the seller's subjective cumulative distribution function of θ. In line with Chapter 2, we consider the case of a risk-neutral agent with an additively separable utility over time. In this dynamic environment the economic allocation describes, for each period $\tau = 1, \ldots, T$, whether the buyer obtains the good from the seller and what transfer the buyer pays the seller. To make this more precise, let $\delta \in [0, 1)$ represent the common discount factor of the buyer and seller.[1] Moreover, let q_τ represent the probability that the agent obtains the good in period τ. Then given a schedule of transfers $t_1, \ldots, t_T$, the agent's utility is

$$\sum_{\tau=1}^{T} \delta^{\tau-1}(\theta q_\tau - t_\tau),$$

while the seller's utility is

$$\sum_{\tau=1}^{T} \delta^{\tau-1} t_\tau.$$

The following definition extends naturally the notion of a direct mechanism to our dynamic setting.

Definition 11.4 *A (dynamic) "direct mechanism" consists of functions* $q = (q_1, \ldots, q_T)$ *and* $t = (t_1, \ldots, t_T)$*, where*

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

and

$$t : [\underline{\theta}, \bar{\theta}] \to \mathbb{R}^T.$$

As in the static case, the interpretation is that in a direct mechanism the buyer is asked to report θ. The mechanism then, once and for all, commits the seller to transferring the good to the buyer with probability $q_\tau(\theta)$ in period τ if the buyer reports that her type is θ, and the buyer has to pay the seller $t_\tau(\theta)$ in period τ if she reports that her type is θ.

Defining $u(\theta) = \sum_{\tau=1}^{T} \delta^{\tau-1}[\theta q_\tau(\theta) - t_\tau(\theta)]$, we similarly extend the notion of incentive compatibility and individual rationality.

Definition 11.5 *A (dynamic) direct mechanism is "incentive-compatible" if truth telling is optimal for every type—that is, if*

$$u(\theta) \geq \sum_{\tau=1}^{T} \delta^{\tau-1}[\theta q_\tau(\theta') - t_\tau(\theta')] \qquad \textit{for all } \theta, \theta' \in [\underline{\theta}, \bar{\theta}].$$

Definition 11.6 *A (dynamic) direct mechanism is "individually rational" if the buyer, conditional on her type, is voluntarily willing to participate—that is, if*

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

As shown in Section 2.2, the optimal selling mechanism, when the seller and the buyer only interact for one period $(T = 1)$, is an indirect mechanism that allows the buyer to buy the good at a posted price p^*, where p^* maximizes $p(1 - F(p))$. We recall from Proposition 2.5 that the corresponding incentive-compatible direct mechanism $(\bar{q}^s, \bar{t}^s)$ is

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

and

$$\bar{t}^s(\theta) = \begin{cases} p^* & \text{if } \theta \geq p^*, \\ 0 & \text{if } \theta < p^*. \end{cases}$$

We are now able to demonstrate the main result of this section that the optimal dynamic mechanism is simply a repetition of this posted price mechanism for each period.

Proposition 11.12 *The (dynamic) direct mechanism* (q^*, t^*) *with* $q^*_\tau(\theta) = \bar{q}^s(\theta)$ *and* $t^*_\tau(\theta) = \bar{t}^s(\theta)$ *for all* $\tau = 1, \dots, T$ *is an optimal selling mechanism.*

Proof

First note that, because $(\bar{q}^s, \bar{t}^s)$ is incentive-compatible and individually rational, also the dynamic mechanism (q^*, t^*) is incentive-compatible and individually rational. The dynamic mechanism (q^*, t^*) yields the principal the payoff

$$\int_{\underline{\theta}}^{\bar{\theta}} \sum_{\tau=1}^{T} \delta^{\tau-1} t^*_\tau(\theta) f(\theta)\, d\theta = p^*(1 - F(p^*))\beta,$$

where $\beta \equiv \sum_{\tau=1}^{T} \delta^{\tau-1} = (1 - \delta^T)/(1 - \delta)$.

Now suppose to the contrary that (q^*, t^*) is not optimal. Under this alternative hypothesis, there exists a dynamic mechanism $(\hat{q}, \hat{t})$ that is incentive-compatible and individually rational, which yields the seller strictly more, that is,

$$\int_{\underline{\theta}}^{\bar{\theta}} \sum_{\tau=1}^{T} \delta^{\tau-1} \hat{t}_\tau(\theta) f(\theta)\, d\theta > p^*(1 - F(p^*))\beta.$$

Given $(\hat{q}, \hat{t})$, define the static mechanism $(\hat{q}^s, \hat{t}^s)$ as follows:

$$\hat{q}^s(\theta) = \sum_{\tau=1}^{T} \delta^{\tau-1} \hat{q}_\tau(\theta)/\beta \quad \text{and} \quad \hat{t}^s(\theta) = \sum_{\tau=1}^{T} \delta^{\tau-1} \hat{t}_\tau(\theta)/\beta \qquad \text{for all } \theta.$$

Because $(\hat{q}, \hat{t})$ is incentive-compatible and individually rational by assumption, it follows that the static mechanism $(\hat{q}^s, \hat{t}^s)$ is incentive-compatible and individually rational. Moreover, the seller's payoff under $(\hat{q}^s, \hat{t}^s)$ is

$$\int_{\underline{\theta}}^{\bar{\theta}} \hat{t}^s(\theta) f(\theta)\, d\theta = \int_{\underline{\theta}}^{\bar{\theta}} \left[\sum_{\tau=1}^{T} \delta^{\tau-1} \hat{t}_\tau(\theta)/\beta \right] f(\theta)\, d\theta,$$

which under our alternative hypothesis strictly exceeds $p^*(1 - F(p^*))$ and therefore contradicts the optimality of $(\bar{q}^s, \bar{t}^s)$. □

Because the optimal dynamic mechanism is a repetition of the static posted price mechanism, the mechanism does not display any real dynamics. In particular, the optimal mechanism does not respond to the agent's buying behavior over the periods.

At first sight, this result may be surprising, because it seems that the mechanism does not optimally exploit the information that is revealed through the buyer's behavior. If, for instance, the buyer refuses to buy at the posted price p^*, then this reveals that the buyer's value lies below the seller's price and she will also not buy in the future. Because the posted price strictly exceeds the seller's cost, it is then in the interest of the seller to reduce the posted price at least a little to induce the buyer to buy in the next period. Likewise, by buying at the posted price p^*, the buyer reveals to the seller that her valuation exceeds p^*, which makes it profitable for the seller to raise the price a little bit.

An important insight of the proposition is that the reasoning in the previous paragraph is misleading: Any mechanism that responds to the buyer's buying behavior is suboptimal. The reason for this is that a dynamic pricing schedule also alters the buying behavior of the buyer, because the buyer will take into account how her current buying behavior affects future prices. This leads to a buying behavior that eliminates any potential benefits from a responsive price schedule.

This discussion also reveals that the optimal mechanism is time inconsistent in the weak sense that, if the buyer's valuation lies below the price p^*, both the seller and the buyer gain from renegotiating the mechanism. Hence, even though it is optimal for the seller to offer the unresponsive mechanism *before* the execution of the mechanism, this is no longer the case *during* the execution of the mechanism. Consequently, the feasibility of this mechanism depends crucially on the full commitment power assumed by the traditional theory of mechanism design: As was explicitly pointed out on page 4 in Chapter 1, the rules, once announced, are set in stone. Especially in dynamic environments, the assumption of full commitment may, however, be rather strong to be appropriate for studying actual applications.

We conclude this section by remarking that the previous arguments apply unchanged to the case when the seller, instead of a single buyer, faces multiple buyers whose valuations for the object do not change over time. Then the optimal mechanism is a repetition of the optimal static second price auction with an optimally chosen reserve price.

11.4 REMARKS ON THE LITERATURE

In extending static mechanism design toward a dynamic framework, we focused on revenue maximization and considered the two most straightforward dynamic extensions: dynamic private information and dynamic allocations.

Our exposition of the sequential screening model in Subsection 11.2.1 closely follows Courty and Li (2000). These authors identified dynamic frameworks in which the usual techniques of static mechanism design extend. Subsections 11.2.2 and 11.2.3 are based on Esö and Szentes (2007a, b). These authors were the first to decompose the buyer's information in initial and additional information—a technique known as inverse transform sampling in the generation of pseudo-random numbers—and apply this to dynamic mechanism design.

Baron and Besanko (1984) were the first to show that a model with dynamic allocations but static information does not lead to meaningful dynamics of the mechanism. Our exposition in Section 11.2 is a modified version of the argument in Laffont and Tirole (1993), who refer to this result as "false dynamics" and point out the commitment issue. Relaxing the full commitment assumption, Laffont and Tirole (1988) study a framework with dynamic allocations without intertemporal commitment, while Laffont and Tirole (1990) consider a framework with dynamic allocations under intertemporal renegotiation. These models raise issues concerning proper versions of the revelation principle as discussed in Bester and Strausz (2001, 2007).

Baron and Besanko (1984), Battaglini (2005), Pavan et al. (2014), and Boleslavsky and Said (2013) study mixed models with both dynamic information and dynamic allocations and many periods. These papers adopt the same local approach as followed in this chapter. In particular, they exploit regularity conditions which guarantee that global first period incentive compatibility constraints can be ignored in the seller's problem. Battaglini and Lamba (2014) argue that such regularity conditions are in general hard to satisfy. In this case, the local approach cannot be used and some non-local incentive constraints will be binding at the optimum and little is known about the general structure of optimal dynamic mechanisms.

In this chapter we focused on revenue maximization rather than efficiency. Bergemann and Välimäki (2010) show how to extend the pivot mechanism as introduced in Chapter 3 to a dynamic setting. Athey and Segal (2013) extend the budget balanced mechanism proposed by d'Aspremont and Gerard-Varet (1979) to a dynamic setting. These results are derived for dynamic models with both dynamic private information and dynamic allocations.

In addition to "dynamic private information" and "dynamic allocations," a further strand of the literature introduces dynamics in the mechanism design problem by considering "dynamic populations." The basic idea underlying this extension is that the seller faces a population of buyers that changes over time and the seller cannot offer an overall mechanism to all possible agents in the first period. When dealing with a dynamic population of buyers, the seller faces the natural tradeoff between selling to the current buyer today or waiting for a future buyer with a possibly higher valuation. The models studying this problem combine ideas of mechanism design and operation research. Because the characterization of optimal mechanisms requires techniques that go beyond those developed in this book, we refer readers to the surveys by Bergemann and Said (2011) and Gershkov and Moldovanu (2012) for further details.

11.5 PROBLEMS

(a) Consider the sequential screening model of Section 11.2.1 with $\tau \in [0,1], \theta \in [0,1]$, and

$$G(\tau) = \tau, \quad \text{and} \quad F(\theta|\tau) = \theta + \tau\left(\left(\theta - \frac{1}{2}\right)^2 - \frac{1}{4}\right).$$

  (i) Show that ψ is increasing in τ and θ and, hence, satisfies Assumption 11.1.
  (ii) Show that the optimal selling mechanism is ex post efficient if the seller incurs no opportunity cost when selling the good.
  (iii) Suppose the seller incurs an opportunity cost $c \in (0,1)$ when selling the good. Show that the optimal selling mechanism (q^*, t^*) is not ex post efficient and that it satisfies

$$q^*(\tau,\theta) = \begin{cases} 1 & \text{if } \theta > \dfrac{c\tau + \sqrt{c - (1-c)c\tau^2}}{1+\tau}, \\ 0 & \text{otherwise.} \end{cases}$$

(b) Consider the sequential screening model of Section 11.2.1. A direct mechanism (q,t) is called *deterministic* if $q(\tau,\theta) \in \{0,1\}$ for all $(\tau,\theta) \in [\underline{\tau},\bar{\tau}] \times [\underline{\theta},\bar{\theta}]$. Show that a direct deterministic direct mechanism (q,t) is incentive-compatible if and only if the schedule q is increasing in both τ and θ, and the transfer schedule t satisfies the expression in Proposition 11.4(ii).[2]

(c) Consider the sequential screening model of Section 11.2.1 with the only difference that the buyer's ex ante private information τ is discrete: $\tau \in \{\tau_L, \tau_H\}$. Let each ex ante type be equally likely, and for $\theta \in [0,1]$ let $F(\theta|\tau_H) = \theta$ and $F(\theta|\tau_L) = \sqrt{\theta}$. A direct mechanism (q,t) can now be represented as a quadruple $(q_L, q_H, t_L, t_H) : [0,1] \to [0,1]^2 \times \mathbb{R}^2$.
  (i) Define the random variable $\gamma_i = F(\theta|\tau_i)$ for $i \in \{L,H\}$. Verify that γ_i is stochastically independent of τ and uniformly distributed over $[0,1]$.
  (ii) Let $(\tilde{q}^*, \tilde{t}^*)$ be a direct mechanism that is optimal among all mechanisms that are incentive-compatible with observable γ. Show

$$\tilde{q}_H^*(\gamma) = 1 \qquad \forall \gamma \in [0,1], \qquad \tilde{q}_L^*(\gamma) = \begin{cases} 0 & \text{if } \gamma < 1/2, \\ 1 & \text{else.} \end{cases}$$

  (iii) Show that the optimal mechanism with observable γ yields ex ante type τ_H an expected utility of $1/12$.

(iv) For the remainder of this problem, assume that the buyer observes γ privately. Suppose that the schedule transfer pair $(\tilde{q}_L^*, \tilde{t}_L)$ induces ex ante type τ_L to report γ truthfully after having reported τ_L. Show that there then exists a $\bar{t}_L \in \mathbb{R}$ such that

$$\tilde{t}_L(\gamma) = \begin{cases} \bar{t}_L & \text{if } \gamma < 1/2, \\ \bar{t}_L + 1/4 & \text{else.} \end{cases}$$

(v) Show that if $(\tilde{q}_L^*, \tilde{t}_L)$ induces ex ante type τ_L to report γ truthfully, then, for any $\gamma \in (1/4, 1/2)$, it induces ex ante type τ_H not to report γ truthfully after having reported τ_L, but instead to report that γ exceeds $1/2$.

(vi) Show that if a schedule transfer pair $(\tilde{q}_L^*, \tilde{t}_L)$ induces ex ante type τ_L to report γ truthfully, it yields ex ante type τ_H expected utility that exceeds τ_L's expected utility by $11/96 > 1/12$.

(vii) Explain that parts (iii) and (vi) imply that the seller must concede a higher information rent to ex ante type τ_H for implementing $(\tilde{q}_L^*, \tilde{q}_H^*)$ when the buyer observes γ privately compared to when γ is publicly observable. Relate this result to Proposition 11.10.[3]

(d) Consider a sequential screening setup between a seller and a buyer where the ex ante and ex post types are both discrete: Before the buyer meets the seller, she receives a binary signal $\tau \in \{1, 3\}$. The seller then offers a mechanism. After accepting the mechanism, the buyer receives a second private signal $\sigma \in \{1, 3\}$. The buyer's valuation is $\theta = \theta_{\tau\sigma} = \tau + \sigma$. Let all draws be equally likely. The seller's opportunity cost of selling the good is $c = 1/2$. A direct mechanism (q, t) now specifies selling probabilities $q_{\tau,\sigma} \in [0, 1]$ and transfers $t_{\tau,\sigma} \in \mathbb{R}$ for all $(\tau, \sigma) \in \{1, 3\}^2$.

(i) Show that any direct incentive-compatible mechanism induces lying off the equilibrium path in the sense that it is optimal for ex ante type $\tau = 3$ to always report $\sigma = 3$ after having reported $\tau = 1$ and that it is optimal for ex ante type $\tau = 1$ to always report $\sigma = 1$ after having reported $\tau = 3$.

(ii) Show that an optimal mechanism (q, t) satisfies $q_{11} = 1$ and $q_{13} = q_{31} = q_{33} = 1$.

(iii) Suppose the buyer does not obtain the ex post signal σ so that ex ante type $\tau = 1$ values the good at (expected) value $\theta_1 = 3$, and ex ante type $\tau = 3$ values the good at (expected) value $\theta_3 = 5$. Show that a (static) direct mechanism $(q_\tau, t_\tau)_{\tau=1,3}$ with $q_1 = q_3 = 1$ is optimal.

(iv) Show that the buyer is worse off when she receives the ex post signal σ privately rather than not at all.[4]

NOTES

Preface

1. Note that the verb "to implement" is used in this book, and, when it is used, it is not meant to refer to implementation in the sense of the implementation literature. Hopefully, the intended meaning of the verb will become clear below.
2. If you find errors of any kind in this book, please send me an email at tborgers@umich.edu. A list of corrections for this book will be maintained online at http://www-personal.umich.edu/~tborgers/.

Chapter 2

1. From now on, when we refer to "utility," we shall mean "von Neumann–Morgenstern utility."
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.
3. Throughout the book we shall say that a function f is "increasing" if it is weakly monotonically increasing, that is, if $x > x'$ implies $f(x) \geq f(x')$.
4. Where "almost all θ" means "for a set of θ with Lebesgue measure $\bar{\theta} - \underline{\theta}$."
5. More precisely: $\hat{q}(\theta) \neq 0$ for a set of θ with strictly positive Lebesgue measure.
6. More precisely: $q(\theta^*) \in (0, 1)$ for a set of θ^* with strictly positive Lebesgue measure.
7. One can show that under Assumption 2.1, which we introduce below, deterministic mechanisms are optimal [Proposition 8 in Maskin and Riley (1989)].

Chapter 3

1. We ignore questions of existence and uniqueness of the conditional expected values referred to in this paragraph.
2. 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.
3. 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 3.4 for an example where this type is not the lowest type in types' numerical order.
4. An exception is Section 3.4.4, where we consider profit-maximizing trading mechanisms.
5. We use the expression "social surplus" to distinguish this expression from social welfare as defined in equation (3.60). The first best decision rule maximizes both, social welfare and social surplus.

Chapter 4

1. This use of the phrase "dominant" is slightly sloppy. We shall comment on terminology at the end of this section.
2. This expression is not commonly used. I made it up for this chapter.
3. The word "type" is used in a loose sense here, as the value is not necessarily within the type range.
4. The example is due to Jan-Henrik Steg. I would like to thank him for allowing me to include it in this book.
5. Dominant strategy mechanisms for public goods economies with more than two agents are further discussed, for example, in Serizawa (1999).

Chapter 5

1. The literature defines this adjective differently from us, but for the purposes of this chapter it is convenient to use the word as defined here.
2. In the following, we use the term "order" synonymously with the word "binary relation."
3. We know this if we have tackled problem (c) in Section 2.6.

Chapter 6

1. In this definition we use the same notion of "equivalent direct mechanism" as we used earlier, for example in Proposition 3.6. To make the result as transparent as possible, we repeat the definition of "equivalent direct mechanisms" in the result.
2. Franklin (1980), p. 56.
3. Because the set of types is finite, prior distributions can be regarded as vectors of real numbers. The notion of "closeness" of prior distributions that we refer to here is the Euclidean distance of the corresponding vectors.
4. Using a different notion of genericity than Heifetz and Neeman, Chen and Xiong (2011) have obtained the opposite result of Heifetz and Neeman's.

Chapter 8

1. Reny writes that his proof is inspired by the proof of Arrow's theorem in Geanakoplos (2005).

Chapter 9

1. 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.
2. Jehiel and Moldovanu (2001) allow more general convex signal spaces with nonempty interior, but for ease of exposition we restrict attention to the case that signals are contained in $[0, 1]^K$.
3. A formal definition of ex post Bayesian equilibrium will be provided in Chapter 10; see Definition 10.13.

Chapter 10

1. Of course, one might also wish a mechanism to be robust with respect to the possibility that agents make errors. This idea is developed in Eliaz (2002).
2. For the details see Mertens and Zamir (1985) and Brandenburger and Dekel (1993).
3. We assume that the sets T_i have a measurable structure so that probability measures on them can be defined.
4. This has first been shown by Mertens and Zamir (1985). A simplified proof is due to Brandenburger and Dekel (1993).
5. The concept of finite types is from Mertens and Zamir (1985) and was emphasized recently in Dekel et al. (2006).
6. I and the readers of this book are grateful to Takuro Yamashita for pointing out that an earlier version of this proof was incorrect.
7. Theorem 1* in Hylland (1980). We use here the version of Hylland's theorem that is Theorem 1 in Dutta et al. (2007) with the correction in Dutta et al. (2008).
8. If $p_i = 1$ for some $i \in I$, that is, when dictatorship is deterministic, not random, then obviously no mechanism and Bayesian equilibrium can be ranked above dictatorship by a mechanism designer pursuing interim Pareto welfare.
9. The difference between the condition in Proposition 10.6 and the linear independence condition is discussed in Crémer and McLean (1988).

Chapter 11

1. If the discount differs, then the buyer and the seller have a different rate of substitution between money in the present and the future. As a result, the intertemporal gains of trade are infinite.
2. This exercise illustrates an observation in Krähmer and Strausz (2011).
3. This exercise illustrates an observation in Krähmer and Strausz (2014).
4. This exercise illustrates an observation in Krähmer and Strausz (2008).

BIBLIOGRAPHY

William J. Adams and Janet L. Yellen. Commodity bundling and the burden of monopoly. *Quarterly Journal of Economics*, **90**: 475–498, 1976.

Kenneth Arrow. *Social Choice and Individual Values*, second edition. John Wiley & Sons, New York, 1963.

Susan Athey and Ilya Segal. An efficient dynamic mechanism. *Econometrica*, **81**: 2463–2485, 2013.

Salvador Barbera. An introduction to strategy-proof social choice functions. *Social Choice and Welfare*, **18**: 619–653, 2001.

David Baron and David Besanko. Regulation and information in a continuing relationship. *Information Economics and Policy*, **1**: 267–302, 1984.

Marco Battaglini. Long-term contracting with markovian consumers. *American Economic Review*, **95**: 637–685, 2005.

Marco Battaglini and Rohit Lamba. Optimal dynamic contracting. Princeton University, 2014.

Dirk Bergemann and Stephen Morris. Robust mechanism design. Unpublished, November 2001.

Dirk Bergemann and Stephen Morris. Robust mechanism design. *Econometrica*, **73**: 1771–1813, 2005.

Dirk Bergemann and Stephen Morris. An introduction to robust mechanism design. *Foundations and Trends in Microeconomics*, **8**: 169–230, 2013.

Dirk Bergemann and Maher Said. Dynamic auctions: A survey. In *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Hoboken, NJ, 2011.

Dirk Bergemann and Juuso Välimäki. The dynamic pivot mechanism. *Econometrica*, **78**: 771–789, 2010.

Dirk Bergemann, Stephen Morris, and Satoru Takahashi. Efficient auctions and interdependent types. *American Economic Review (Papers and Proceedings)*, **102** (3): 319–324, 2012.

Helmut Bester and Roland Strausz. Contracting with imperfect commitment and the revelation principle: The single agent case. *Econometrica*, **69**: 1077–1089, 2001.

Helmut Bester and Roland Strausz. Contracting with imperfect commitment and noisy communication. *Journal of Economic Theory*, **136**: 236–259, 2007.

Felix Bierbrauer, Axel Ockenfels, Andreas Pollak, and Désirée Rückert. Robust mechanism design and intentions. Unpublished, January 2014.

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.

Raphael Boleslavsky and Maher Said. Progressive screening: Long-term contracting with a privately known stochastic process. *Review of Economic Studies*, **80**: 1–34, 2013.

Patrick Bolton and Mathias Dewatripont. *Contract Theory*. MIT Press, Cambridge, 2005.

Tilman Börgers. (No) foundations of dominant-strategy mechanisms: A comment on Chung and Ely (2007). Unpublished, November 2013.

Tilman Börgers and Peter Norman. A note on budget balance under interim participation constraints: The case of independent types. *Economic Theory*, **39**: 477–489, 2009.

Tilman Börgers and Peter Postl. Efficient compromising. *Journal of Economic Theory*, **155**: 2057–2076, 2009.

Tilman Börgers and Doug Smith. Robustly ranking mechanisms. *American Economic Review (Papers and Proceedings)*, **102**: 325–329, 2012.

Tilman Börgers and Doug Smith. Robust mechanism design and dominant strategy voting rules. *Theoretical Economics*, **9**: 339–360, 2014.

Steven Brams and Peter Fishburn. *Approval Voting*, second edition. Springer, New York, 2007.

Adam Brandenburger and Eddie Dekel. Hierarchies of beliefs and common knowledge. *Journal of Economic Theory*, **59**: 189–198, 1993.

Yi-Chun Chen and Siyang Xiong. The genericity of beliefs-determine-preferences models revisited. *Journal of Economic Theory*, **146**: 751–761, 2011.

Kim Sau Chung and Jeffrey C. Ely. Foundations of dominant strategy mechanisms. *Review of Economic Studies*, **74**: 447–476, 2007.

Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, **8**: 19–33, 1971.

Pascal Courty and Hao Li. Sequential screening. *Review of Economic Studies*, **67**: 697–717, 2000.

Jacques Crémer and Richard McLean. Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica*, **56**: 1247–1257, 1988.

Claude d'Aspremont and Louis-Andre Gerard-Varet. Incentives and incomplete information. *Journal of Public Economics*, **11**: 25–45, 1979.

Eddie Dekel, Drew Fudenberg, and Stephen Morris. Topologies on types. *Theoretical Economics*, **1**: 275–309, 2006.

Bhaskar Dutta, Hans Peters, and Arunava Sen. Strategy-proof cardinal decision schemes. *Social Choice and Welfare*, **28**: 163–179, 2007.

Bhaskar Dutta, Hans Peters, and Arunava Sen. Strategy-proof cardinal decision schemes (erratum). *Social Choice and Welfare*, **30**: 701–702, 2008.

Kfir Eliaz. Fault tolerant implementation. *Review of Economic Studies*, **69**: 589–610, 2002.

Peter Esö and Balasz Szentes. Optimal information disclosure in auctions and the handicap auction. *Review of Economic Studies*, **74**: 705–731, 2007a.

Peter Esö and Balasz Szentes. The price of advice. *RAND Journal of Economics*, **38**: 863–880, 2007b.

Vitor Farinha Luz. Surplus extraction with rich type spaces. *Journal of Economic Theory*, **148**: 2749–2762, 2013.

Joel Franklin. *Methods of Mathematical Economics*. Springer, New York, 1980.

Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, Cambridge, 1993.

John Geanakoplos. Three brief proofs of arrow's impossibility theorem. *Economic Theory*, **26**: 211–215, 2005.

Alex Gershkov and Benny Moldovanu. Dynamic allocation and pricing: A mechanism design approach. *International Journal of Industrial Organisation*, **30**: 283–286, 2012.

Alex Gershkov, Jacob K. Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi. On the equivalence of Bayesian and dominant strategy implementation. *Econometrica*, **81**: 197–230, 2013a.

Alex Gershkov, Benny Moldovanu, and Xianwen Shi. Optimal mechanism design without money. Discussion paper, April 2013b.

Alan Gibbard. Manipulation of voting schemes. *Econometrica,* **41**: 587–601, 1973.

Alan Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica,* **45**: 665–681, 1977.

Alan Gibbard. Straightforwardness of game forms with lotteries as outcomes. *Econometrica,* **46**: 595–614, 1978.

Jerry Green and Jean-Jacques Laffont. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica,* **45**: 727–738, 1977.

Theodore Groves. Incentives in teams. *Econometrica,* **41**: 617–631, 1973.

Werner Güth and Martin Hellwig. The private supply of a public good. *Journal of Economics, Supplement* 5, 121–159, 1986.

Kathleen Hagerty and William Rogerson. Robust trading mechanisms. *Journal of Economic Theory,* **42**: 94–107, 1987.

John C. Harsanyi. Games with incomplete information played by Bayesian players. *Management Science,* **14**: 159–182, 320–334, 486–502, 1967–1968.

Aviad Heifetz and Zvika Neeman. On the generic (im)possibility of full surplus extraction in mechanism design. *Econometrica,* **74**: 213–233, 2006.

Jobst Heitzig and Forrest Simmons. Some chance for consensus: Voting methods for which consensus is an equilibrium. *Social Choice and Welfare,* **38**: 43–57, 2012.

Benjamin Hermalin. *Lecture Notes for Economics.* 2005. URL http://faculty.haas.berkeley.edu/hermalin/.

Bengt Holmström. Groves' scheme on restricted domains. *Econometrica,* **47**: 1137–1144, 1979.

Aanund Hylland. Strategy proofness of voting procedures with lotteries as outcomes and infinite sets of strateges. Unpublished 1980.

Matthew Jackson and Hugo Sonnenschein. Overcoming incentive constraints by linking decisions. *Econometrica,* **75**: 241–275, 2007.

Philippe Jehiel and Benny Moldovanu. Efficient design with interdependent valuations. *Econometrica,* **69**: 1237–1259, 2001.

Philippe Jehiel and Benny Moldovanu. Allocative and informational externalities in auctions and related mechanisms. In *Proceedings of the 9th World Congress.* Econometric Society, London, 2005.

Philippe Jehiel, Moritz Meyer ter Vehn, Benny Moldovanu, and Bill Zame. The limits of ex post implementation. *Econometrica,* **74**: 585–611, 2006.

Grigory Kosenok and Sergei Severinov. Individually rational, budget-balanced mechanisms and allocation of surplus. *Journal of Economic Theory,* **140**: 126–161, 2008.

Daniel Krähmer and Roland Strausz. Optimal procurement contracts with pre-project planning. *Review of Economic Studies,* **78**: 1015–1041, 2011.

Daniel Krähmer and Roland Strausz. Ex post information rents in sequential screening. Unpublished 2014.

Daniel Krähmer and Roland Strausz. Ex post private information and monopolistic screening. *The B.E. Journal of Theoretical Economics, Topics,* **8**(1), Article 25, 2008.

Vijay Krishna. *Auction Theory.* Academic Press, San Diego, 2002.

Vijay Krishna and Eliot Maenner. Convex potentials with an application to mechanism design. *Econometrica,* **69**: 1113–1119, 2001.

Jean-Jacques Laffont and David Martimort. Mechanism design with collusion and correlation. *Econometrica*, **68**: 309–342, 2000.

Jean-Jacques Laffont and Jean Tirole. The dynamics of incentive contracts. *Econometrica*, **56**: 1153–1175, 1988.

Jean-Jacques Laffont and Jean Tirole. Adverse selection and renegotiation in procurement. *Review of Economic Studies*, **57**: 579–625, 1990.

Jean-Jacques Laffont and Jean Tirole. *A Theory of Incentives in Regulation and Procurement*. MIT Press, Cambridge, 1993.

David G. Luenberger. *Optimization By Vector Space Methods*. John Wiley & Sons, New York, 1969.

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.

Eric Maskin and John Riley. Optimal multi-unit auctions. In Frank Hahn, editor, *The Economics of Missing Markets, Information, and Games*, pages 312–335. Clarendon Press, Oxford, 1989.

Jean-François Mertens and Shmuel Zamir. Formulation of Bayesian analysis for games with incomplete information. *International Journal of Game Theory*, **14**: 1–29, 1985.

Paul Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, Cambridge, 2004.

Paul Milgrom and Nancy Stokey. Information, trade and common knowledge. *Journal of Economic Theory*, **26**: 17–27, 1982.

Antonio Miralles. Cardinal Bayesian allocation mechanisms without transfers. *Journal of Economic Theory*, **147**: 179–206, 2012.

Dilip Mookherjee and Stefan Reichelstein. Dominant strategy implementation of Bayesian incentive compatible allocation rules. *Journal of Economic Theory*, **56**: 378–399, 1992.

Herve Moulin. On strategy-proofness and single peakedness. *Public Choice*, **35**: 437–455, 1980.

Michael Mussa and Sherwin Rosen. Monopoly and product quality. *Journal of Economic Theory*, **18**: 301–317, 1978.

Roger Myerson. Optimal auction design. *Mathematics of Operations Research*, **6**: 58–73, 1981.

Roger B. Myerson and Mark A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, **29**: 256–281, 1983.

Zvika Neeman. The relevance of private information in mechanism design. *Journal of Economic Theory*, 117: 55–77, 2004.

Efe A. Ok. *Real Analysis with Economic Applications*. Princeton University Press, Princeton, NJ, 2007.

Alessandro Pavan, Ilya Segal, and Juuso Toikka. Dynamic mechanism design: A Myersonian approach. *Econometrica*, **82**: 601–653, 2014.

Philip Reny. Arrow's theorem and the Gibbard-Satterthwaite theorem: A unified approach. *Economics Letters*, **70**: 99–105, 2001.

Jacques Robert. Continuity in auction design. *Journal of Economic Theory*, **55**: 169–179, 1991.

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.

Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, **16**: 191–200, 1987.

Tyrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

Alvin Roth. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory*, **36**: 537–569, 2008.

Halsey Royden and Patrick Fitzpatrick. *Real Analysis*, fourth edition. Prentice Hall, Upper Saddle River, NJ, 2010.

Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, New York, 1976.

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.

Shigehiro Serizawa. Strategy-proof and symmetric social choice functions for public good economies. *Econometrica*, **67**: 121–145, 1999.

Doug Smith. A prior free efficiency comparison of mechanisms for the public goods problem. Unpublished, 2010.

William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, **16**: 8–37, 1961.

Takuro Yamashita. A necessary condition for implementation in undominated strategies, with applications to robustly optimal trading mechanisms. Unpublished, 2012.

INDEX