Studies in Public Economics

Volume 1

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM · NEW YORK · OXFORD

Incentives in Public Decision-Making

JERRY R. GREEN
Harvard University

JEAN-JACQUES LAFFONT
Université des Sciences Sociales de Toulouse and Ecole Polytechnique

1979

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM · NEW YORK · OXFORD

Publishers

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

Sole distributors for the U.S.A. and Canada

ELSEVIER NORTH-HOLLAND, INC.
52 VANDERBILT AVENUE
NEW YORK, N.Y. 10017

Library of Congress Cataloging in Publication Data

Green, Jerry R.
Incentives in public decision-making.
(Studies in public economics; v. 1)
Bibliography: p. 283
Includes index.
1. Social choice. I. Laffont, Jean-Jacques, joint author. II. Title. III. Series.
HB99.G74 350 78–19132
ISBN 0 444 85144 5

PRINTED IN THE NETHERLANDS

INTRODUCTION TO THE SERIES

This series will present research on a broad range of different aspects of the economics of the public sector. In addition to studies of taxation, we expect to publish volumes on such diverse subjects as public expenditure planning, social insurance, fiscal policy, and the economics of public and regulated enterprises. The editors will also encourage the publication of high quality monographs in particular areas of the public and nonprofit sector including transportation, law enforcement and education.

Recent research in public economics has used a wide variety of theoretical and empirical approaches. We hope that the studies in this series will reflect this spectrum of methods.

In publishing this international series of studies, we express our conviction that public economics is an applied subject with methods and conclusions that transcend the institutions of individual countries. Although the editors welcome empirical studies of the experience in a particular country, the analysis or results should be of interest to scholars in other countries as well.

Two neighbours may agree to drain a meadow, which they possess in common; because 'tis easy for them to know each others mind; and each must perceive, that the immediate consequence of his failing in his part, is the abandoning the whole project. But 'this very difficult, and indeed impossible, that a thousand persons shou' d agree in any such action; it being difficult for them to concert so complicated a design, and still more difficult for them to execute it; while each seeks a pretext to free himself of the trouble and expence, and wou 'd lay the whole burden on others.

David Hume

A Treatise of Human Nature
(Oxford University Press, p. 538)

PREFACE

In a society of ever-increasing complexity, rules and methods for public decision making tend to become institutionalized. The settings in which collective actions are taken often are precisely those in which the forces of competition may fail to rectify errors of inappropriate or faulty decision processes. It is therefore incumbent upon the designer of the system to choose methods that can be expected to perform well within their range of application. We consider a central decision-maker whose action is selected from among a given set of alternatives. He is benevolent and, ideally, if the full description of the economic system were known to him, he would select according to some rule from among the Pareto optima. However, this information is not at his command; more explicitly, it is widely scattered. Various members of society have different pieces to puzzle, and it is not necessarily in their interests to reveal them. Thus the dual features of conflicting individual goals and diverse information combine to create the structure of the problems we treat in this book.

Our goal has been to explore the possibility of overcoming these obstacles to efficient decision-making. The book is organized into four parts. We begin by setting out the basic negative results regarding the possibility of constructing general methods that lead to Pareto optima and are immune to individual gamesmanship. The interconnections between the recent strands of this branch of social choice theory and its intellectual forerunners in public finance, moral philosophy and economics of centrally planned society is set out in chapter one. A more formal treatment is given in chapter two.

Part II is devoted to an exploration of the basic technique for inducing revelation of individual preferences for collective action. Chapter three describes this idea, and shows how it can be applied to both public and private allocation problems. Chapter four treats this in a rigorous way, extending and generalizing methods set out in chapter three. The basic result is that, under some highly restrictive, but not uninteresting, economic conditions, a class of methods for eliciting individuals' preferences and selecting, simultaneously, Pareto optima does exist. The method is characterized in this chapter. Therefore, if one maintains these two properties as strict requirements, the search for a suitable public decision making rule must be limited to this class. Chapter five goes on to show that the results of chapter four cannot be generalized in any of several, potentially desirable, ways. We show that, in each case, extending these methods entails the loss of one of its desirable characteristics. We are therefore faced with the problem of trading-off these properties against other features which might be more important in particular applications. The remainder of the book can be viewed as an exploration of these trade-offs.

In chapter six a consideration of great importance for many public decision making contexts, that has not been mentioned thus far, is introduced. In a voluntary society, one cannot presume that individuals will elect to participate in a social decision making procedure which might be to their own disadvantage. Nonparticipation, or abstention, must be viewed as a realistic possibility. Accepting the criterion that participation be individually advantageous allows us to narrow the class of satisfactory decision making procedures of chapter four quite substantially. Fortunately however, some of the more interesting and useful procedures having the desirable properties mentioned above can satisfy this requirement as well.

In chapter seven we explore the way in which some of the negative results of chapter five can be overcome by weakening the form of the incentives for individuals to correctly reveal their privately held information. Three alternative incentive compatibility requirements are treated: individuals maximize expected utility given their beliefs, individuals follow a maximin criterion, individuals optimize given a knowledge of the strategies played by all other agents.

Although the results of part II have left us in a somewhat uneasy state one important feature of public decision-making processes has been neglected. This is the fact that, very often, large numbers of economic agents are involved. In part III we investigate the potential for overcoming problems of incentive compatibility through the use of large-numbers, approximative, methods. Each of the primary difficulties described in chapter five, which cannot be handled in a deterministic, small-numbers setting is treated in this part. Chapter eight gives an overview of the main ideas we will utilize and surveys our "large-number" methods in this type of economic analysis. Chapter nine develops the problem of the imbalance between receipts and payments of monetary transfers entailed by the use of the incentive compatible decision making rules. Chapter ten explores

the possibility that the problems due to misrepresentation of preferences by coalitions, rather than individuals, can be overcome in large systems when the methods of chapter four are used. Chapter eleven deals with income effects. Here it is shown that, unlike the favorable results attained in chapters nine and ten, income effects represent a more robust type of difficulty. Although some improvements over naive decision making methods can be obtained with many economic agents, the cost imposed by the imperfections of information cannot be completely overcome. Chapter twelve introduces the idea that in large systems, it may only be necessary to elicit the preferences from a sample drawn randomly from a population, in order to make an efficient decision with high probability. This consideration is particularly relevant when the process of elicitation of preferences itself is costly. The properties of the optimal sample size and its dependence on the prior beliefs of the planner are explored for several forms of this cost function. Chapter thirteen addresses a new type of imperfection of information. Unlike the rest of the book, in which information is perfect at the personal level and failures persist from the point of view of the planner only, this chapter makes explicit the individual's ignorance about characteristics of the social decision to be taken. For economists trained in the neo-classical tradition of consumer sovereignty, these considerations may seem rather odd. But they are an important, practical, serious aspect of virtually all governmental decision-making. Therefore, even though our results in this regard are limited to certain special cases, we felt that the pervasiveness of the problem made them worthy of our attention.

All of the results of parts II and III have been cast in the mold of a single play game. That is, individuals are assumed to select strategies which jointly, determine the outcome. In fact, these strategies may be complex, and the game may actually be played in extensive form over a considerable period of time. Viewed in this light, the essential assumption is really that individuals play the game in which they are embedded with perfect rationality and foresight. Because many social decision-making processes are dynamic in character, considering at each stage the revision or modification of the status-quo condition by small steps, the assumption of individual rationality may be modified to its polar opposite of individual myopia. Part IV is devoted to a study of incentive compatibility under this type of behavioral hypothesis. In the literature on planning procedures, the assumption of myopia is commonly employed. In chapter fourteen the basic structure of this type of model is set up. Chapter fifteen concerns the treatment of mechanisms in which individuals are assumed to play maximin strategies at each stage of the game. This method can be viewed as a successful way of avoiding the problem of income effects which was resistant to the large-numbers approach of part III. One may object, however to the restrictive character of the maximin assumption. Therefore, we have attempted in our final chapter to integrate dominant strategy procedures in this dynamic context. Our methods have met with only limited success. Substantial advantages of computational complexity can be achieved, but the existence of the income effects in general appear resistant still. However, if the public decision can be parameterized in one dimension, then these methods achieve optimal results in the limit, with or without income effects.

In preparing this book we attempted to steer a course compatible with both technical rigor and generality on the one hand, and accessibility and potential applicability on the other. Unfortunately, our results are of such a character that we can not, unambiguously, come out in "favor" of the utilization of a certain type of public decision making procedure. Nevertheless, we hope that we have been able to provide some guidance as to the qualities of alternative procedures, and the likelihood of finding one suitable for any given circumstance. Research on this topic is far from complete and we look forward to exciting new results for some time to come.

Cambridge, Paris
July 1977

J.R. Green
J.-J. Laffont

ACKNOWLEDGEMENT

Although neither of us is a specialist in public finance or in social choice theory, we found ourselves attracted to the topic of this volume. Our friends and colleagues are primarily responsible. It was their vigorous interest in the many facets of the incentive problem that spurred our original efforts and encouraged us to undertake completing the manuscript.

First of all our thanks are due to Kenneth Arrow. In the spring of 1974 he focused our attention on the relationship between the incentive problems of resource allocation teams and some recent results of social choice theory. Our debt in this respect is great, but it is still small compared with the sum already accumulated as colleague, teacher, and friend.

Our work began that summer during a seminar at the Institute for Mathematical Studies in the Social Sciences at Stanford University. Frank Hahn, Robert Aumann, Eytan Sheshinski and Elon Kohlberg contributed greatly to our momentum, as well our understanding, in these crucial early stages.

By the end of that academic year the broader structure of the problem was well defined in our minds and some of the pieces were taking shape. A preliminary summary of our work, which was more of a research outline at this stage, was presented to an International Conference on Public Economics at Kibbutz Kiryat Anavim, Israel, in June 1975. Daniel McFadden and James Mirrlees provided some useful feedback that markedly shaped some of our late work. Shortly after the conference we decided to write this book; Martin Feldstein and Eytan Sheshinski, the editors of this series, were instrumental in this choice and have provided the necessary flexibility which made the writing easier.

The next two years saw a good deal of travel. Our wives, Miki and Colette, withstood our absences in good humor and were excellent hostesses on the alternate occasions.

In the final stages of the writing we benefited greatly from the comments of Vernon Smith, John Roberts, Howard Raiffa, Roger Guesnerie and Paul Champsaur, among many others, and from the collaboration of Eric Maskin on some results which are reported in chapters 4 and 7.

We are also greatful to the participants at the two seminars we organized concerning the problems addressed in this book: the NSF-NBER Decentralization seminar, held at Harvard in April 1977, and the European Summer Workshop of the Econometric Society, held in Paris in June 1977.

Financial assistance was generously provided by the National Science Foundation, the NATO Systems Science Program, and by the Ecole Polytechnique through the invaluable support of Claude Henry.

The manuscript was typed excellently and cheerfully by Cynthia Gaulin, Mary Bosco, Martine Vidoni and Alves de Souza. Takatoshi Ito provided fine assistance through many hours of proof-reading and also prepared the index.

CONTENTS

NOTATION*

N	number of economic agents
a	a social state
A	set of alternative social states
$\boldsymbol{P}_i$	preferences of i, an ordering over A
$\boldsymbol{P}$	$(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N)$
$\Sigma(A)$	set of all allowable private preferences
$S(A)$	set of all allowable social preferences
$\mathscr{F}$	social welfare function
ℓ	social choice function or outcome function
U_i	utility of i
x_i	consumption of i of private goods
d	social decision
v_i	willingness to pay (for $d = 1$ against $d = 0$)
$\bar{x}_i$	initial allocation of private good
S_i	strategy set for i
S	$S_1 \times \ldots \times S_n$
f	outcome function
(S, f)	a mechanism
t_i	transfer to i
h_i	arbitrary real valued functions (to modify the transfer)
w_i	strategy for i, a professed willingness-to-pay
$\mathscr{K}$	space of social decisions
V_i	allowable willingness to pay functions for i, defined over $\mathscr{K}$
V	$V_1 \times \ldots \times V_N$
F	set-valued outcome function
D	set-valued decision function
K^*	the maximizer of $\sum_{i=1}^{N} v_i(K)$

* Only the principle notation which is used repeatedly is given here. Other usage corresponds to generally accepted mathematical conventions, or is defined and used entirely within one section.

ξ_i	the set of optimal strategies for i
$\mathscr{D}_i$	set of all dominant strategies for i
$\bar{K}$	"status quo" project
ϑ	a parameter of the willingness to pay function
Θ	parameter space
λ_i	welfare weight of i in utilitarian objective
η_i	characteristic of i on which welfare weight should be based
S_i^a	admissible strategies for i under a "no-bankruptcy" constraint
C	a coalition
$\mathscr{C}$	set of all possible coalitions
μ	subjective beliefs about strategies of the other agents
x	sum of other agents willingness to pay (for $d = 1$) $\sum_{j \neq i} w_j$
P_i	subjective distribution of x
p_i	density of P
ν	subjective, ex ante, belief about own true willingness to pay, v_i
π	an arbitrary measure on $\mathscr{K}$
π_j	induced measure on $\mathscr{K}$ given μ and $w_j(\cdot)$
m_i	a message sent by i
F	distribution of willingnesses to pay
f	the density of F
n	sample size
m	mean of F
$N(0, r^2)$	subjective beliefs about own value of v_i
c	cost of ascertaining correct information
$c^*(n)$	cutoff point above which agents do not acquire information
$\sigma^2(c_i)$	variance of information around true v_i if it is obtained

PART I

Chapter 1

INCENTIVES IN ECONOMIC SYSTEMS

1.1. Three roots of this study

The concern with incentives as a problem in economic theory and particularly the incentives to correctly reveal privately held information for public use, grows out of a confluence of three streams of thought. The market socialism[1] literature of the 1930s and 1940s has developed into a sophisticated set of theories of economic planning. As the questions they approached became increasingly complex, incorporating public expenditures as well as private investment decisions, serious difficulties arose in regarding the interests of the agents as fully compatible with the planner's objectives. Economic theory then tried to incorporate incentives directly into the planning procedures – and actual socialist economies paralleled this with the institution of private rewards and penalties related to the attainment of target output and productivity levels. Both in theory and in practice this line of thought was pursued, leading to increasing degrees of complexity in process design and individual response rules. Quite naturally, it was asked whether procedures could be found that were immune to private strategic manipulation under any circumstances.[2]

[1] The theory of market socialism begins with Lange [1936] and Lerner [1944] who tried to duplicate the workings of a perfectly competitive market through a decentralized planning process in which incomes could be controlled. This method is made rigorous by Arrow and Hurwicz [1960]. Public goods are introduced into planning processes by Drèze and de la Vallée Poussin [1971], Malinvaud [1972a] and Heal [1973]. Bergson [1967] presents a survey of the market socialism literature.

[2] Since the 1964 Economic Reform in the Soviet Union (see Lieberman [1965]), the question of private incentives, particularly for managers of production units has been the subject of much concern. This material is surveyed in Berliner [1976] (see Weitzman [1976] for an analytical formulation of the new incentive system in the Soviet Union).

The second strand arises out of quite a different set of considerations. ›eginning with the seminal work of Arrow [1951] in social choice theory, conomists and political scientists have felt the need to seek collective ecision rules that are at least somewhat satisfactory, perhaps under mited circumstances.[3] For the most part this has taken the form of estricting the structure of the social states themselves. But social choice ieory concerns itself with rational aggregation of known preferences. ven if such rules can sometimes be found, it is not clear that agents would eveal their true tastes, and hence such a decision process may be finding ptimal social choices for a distorted set of preferences, resulting in allo- ations that are far from the true optima.

Finally, there is an impetus to study incentives from the point of view f market oriented processes in the presence of various kinds of imper- ctions. When externalities or common property resources exist, it is ell-known that perfectly competitive markets, as they are usually con- eived, fail to achieve Pareto efficiency. Two types of remedies are proposed. there is sufficient time, and if transactions costs can be neglected, it is iggested that free bargaining[4] will lead to a Pareto optimal final outcome. nother possibility is to create artificial property rights for the exter- alities[5] and achieve efficiency by treating these, along with the ordinary ommodities, on a much expanded set of competitive markets. Either of iese schemes presumes that individuals will continue to act according to ieir true preferences. But the bilateral and multilateral negotiations quired by the former, and the thinness of some markets generally arising the latter, are both well-known causes of the failure of correct incentives r agents to follow exactly such a course of behavior.

The common set of open problems left in the wake of these approaches quires a general study of the role of incentives in various forms of eco- omic organizations. Several years ago this task was begun in earnest by urwicz [1959, 1972] and others. Initially the results were mixed, and even omewhat conflicting in their implications. In some circumstances it ppeared that the lack of accurate information necessary for the planning ocess engendered by inappropriate private incentives could be com-

3 Single-peaked preferences provide the possibility for successfully aggregating individuals' stes, as shown by Black [1948] in one dimension and Tullock [1959] in two dimensions. n [1970] gives a detailed treatment of other sufficient conditions for restricted domains der which Arrow's axioms are consistent.

4 See Coase [1960] for a defense of this position.

5 These ideas were discussed by Coase [1960] and Meade [1952], and were formalized in row [1969].

pletely overcome without loss;[6] in others it appeared to be insurmount- able.[7] Our objective in this book is to make some progress in the study of incentives in economic organizations by using a unifying framework and hopefully to provide useful results and interpretations. We will study how privately held information can be elicited for public use, and the effects of the elicitation process on economic efficiency. In this way we can ascertain the extent to which the problems of imperfect public information can be mitigated. We will be referring to the case of unknown individual preferences for public goods, and the incentives for revealing them. But this should be viewed primarily as a vehicle through which the more general problem of obtaining private information useful for social decisions can be made concrete for expository purposes.

1.2. Incentives and public goods

Since Wicksell, economists have been aware of the difficulties that public goods create in allocation mechanisms from an incentive point of view. Hurwicz [1972] has recently pointed out that an analogous difficulty exists for exchange economies with private goods in the sense that no individually rational mechanism exists for which truthful behavior is a Pareto optimal Nash equilibrium.[8] However, when the size of the population is large, the incentive to "misbehave" becomes small, as shown by Roberts and Postlewaite [1976]. On the contrary, with public goods, incentives to lie and distort one's preferences seem to increase with the size of the population.[9] Thus public goods present a more serious problem than purely private goods economies.

At the end of the 19th century a lively debate over public finance took place among European economists.[10] The "benefit" approach, holding that agents should be taxed according to the benefits they derive from public goods, and the "ability to pay" approach, believing that agents should be taxed according to their wealth, were opposed. Most authors

6 These positive results are used in Groves and Loeb [1975], Groves [1973] and Smets [1972] in various contexts.

7 See Hurwicz [1960], Gibbard [1973] and Satterthwaite [1975].

8 The requirement that true revelation of preferences be a Nash equilibrium for all possible configurations of the economy is equivalent to the property that the truth be a dominant strategy.

9 This has been made precise by Roberts [1976].

10 The reader interested in the history of this controversy is referred to Musgrave [1959] and Musgrave and Peacock [1967].

such as J.B. Say and J.S. Mill considered that taxation had to be more or less independent of public spending and therefore favored the ability to pay approach. From 1880 on, a number of economists – in particular Mazzola, Pantaleoni, and de Viti de Marco in Italy, and Sax in Austria – began to use the "modern" concepts of marginal utility and subjective value in the study of public services and of their financing. They revived the benefit approach which is implicit in the writings of many authors of the 18th century, such as Bentham, Locke and Rousseau.

A seminal step was taken in 1896 when, Knut Wicksell published "Ein neues Prinzip der gerechten Besteuerung", a book of three essays in public finance (Wicksell [1896]). In his discussion of Mazzola's contribution to the above debate, he pointed out what became known later as the free rider problem, which had been ignored in the benefit literature. Mazzola had formulated a law of fiscal economics summarized by Wicksell [1896, p. 81] as follows:

> The imposition of taxes is just and applies without arbitrariness and error when each taxpayer succeeds in distributing his resources in such a manner that his utility is maximized.

For Wicksell this requirement is meaningless [1896, p. 81]:

> If the individual is to spend his money for private and public uses so that this satisfaction is maximized, he will obviously pay nothing whatsoever for public purposes (at least if we disregard fees and similar charges). Whether he pays much or little will affect the scope of public service so slightly, that for all practical purposes, he himself will not notice it at all. Of course if everyone were to do the same, the State would soon cease to function.

Wicksell recognized the triviality of his remark but emphasized its fundamental character. He concludes that an efficient allocation of resources cannot be realized in economies with public goods through a decentralized behavior [1896, p. 82]:

> The equality between the marginal utility of public goods and their price cannot therefore be established by the single individual, but must be secured by consultation between him and all other individuals or their delegates. How is such consultation to be arranged so that the goal [efficiency?] may be realized. This is precisely the question which ought to be decided.

The problem was clearly stated, and this was probably Wicksell's main contribution to the theory of public finance. He suggested also a solution, the principle of (approximative) unanimity and voluntary consent. Each item in the public budget must be voted simultaneously with the determination of its financing and must be accepted only if unanimity (or quasi-unanimity) is obtained. This suggestion has been formalized by D. Foley

[1967] with his concept of politico-economic equilibrium. Let us note that Wicksell insisted on the existence of an appropriate initial distribution of resources in order for his principle to be valid. If we could ignore the information and communication costs, as well as threats and strategic behaviors, this process would lead to a Pareto optimum. However, which one of the Pareto optima will be reached depends upon the sequential realization of the decision-making process. Indeed, this is the main reason justifying strategic behavior by the participants as they try to manipulate the path of the procedure.

In "die Gerechtigkeit der Besteuerung", Lindahl [1919] has presented what is often considered as the final version of the benefit approach. More precisely, Lindahl has given, in an economy with public goods, a (partial equilibrium) characterization of the Pareto optimum for which each agent's contribution to the financing of a given public good equals his marginal utility for this good multiplied by its quantity, or in Lindahl's terms, for which the relative share of a given agent in the cost of a public good equals his marginal utility for that good.

The partial equilibrium character of the analysis can be easily legitimated if each agent has a constant marginal utility of income as shown by Samuelson [1969]. The only serious weakness of Lindahl's presentation is to ignore that even if his mechanism starts with an optimal distribution of income it can reach an inappropriate distribution, that is, a Pareto optimum which is not the best in the sense of a given social welfare function. It was only through Samuelson's [1954, 1955] general equilibrium approach that the multiplicity of Pareto optima in economies with public goods was well understood. Given initial resources, one can associate to each Pareto optimum a vector of pseudo-prices and lump sum transfers such that the given Pareto optimum can be reached as a competitive equilibrium with these pseudo-prices and transfers. This is called a pseudo-equilibrium. The qualification "pseudo" is due to the fact that a different price for the public good is quoted to each agent. Because the prices are personalized in this way, thinness of markets is likely to preclude competitive behavior. In this framework it is appropriate to call the pseudo-equilibrium for which lump sum transfers are zero, a Lindahl equilibrium. It is the state of the economy which would eventually be reached if, starting with the given initial resources of the economy, the artificial markets for public goods with personalized prices could be organized in a competitive way.

Lindahl's contribution provides the first real characterization of Pareto optimality in economies with public goods. But Wicksell's fundamental criticism has not been accounted for in this analysis, since it assumes that preferences are known. To sum up, the goal is more clearly defined, but it still seems unattainable.

In his 1939 paper, Musgrave [1939] presents the debate summarized above, emphasizing the free rider problem. Samuelson is also very clear in his mathematical version of his pure theory of public finance [1954] about the impossibility to decentralize a pseudo-equilibrium.

It remained necessary to attack this problem directly.

1.3. Some early attempts to solve the free rider problem

H.R. Bowen [1943] seems to have been the first economist to make an important contribution on the free rider problem stated by Wicksell. Moreover his paper contains an independent discovery of Lindahl's results and an example showing how majority voting can lead to Pareto optimality, a first step toward Black's result [1948], that majority voting provides optimal social choices according to a transitive rule when individuals have single peaked preferences.

First Bowen establishes the conditions for Pareto optimality and rejects the possibility that the market alone can achieve them. Instead he suggests a voting mechanism. He realizes that it is hopeless to expect that agents will reveal their preferences for a given project through voting if they know that their share of the cost is linked to their vote. Therefore, he assumes an a priori fixed imputation of costs (equal shares if the initial distribution of income is fair) from which individuals derive their net willingnesses to pay. An agent then votes for his most preferred level of the public good, as determined by his net willingness to pay function. The decision is then taken at the median of these expressed levels for the public good. This scheme is seen to be non-manipulable in the sense that it is in each agent's interest to vote according to his true preferences. Misrepresenting his preferences, he either leaves the median level of the public good unchanged or changes the median, but in a direction he would not wish. Truthful behavior is obtained, but Pareto optimality is reached only if the mean and the median of the distribution of net willingnesses to pay are equal. If one attempts to reach Pareto optimality by letting agents express the intensity of their preferences in their vote, agents realize the influence of their vote on the decision mechanism and modify their votes in an attempt to obtain their preferred outcome. Suppose for example that a public project is undertaken if the sum of individual announced evaluations

exceeds the cost of the project which has been equally shared. Then, an agent for whom the utility of the project exceeds the per capita cost will answer his largest allowable evaluation in order to make the project accepted. On the contrary, an agent with a negative willingness to pay will announce the lowest possible evaluation in order to kill the project.

More recently, E. Thompson [1967] has suggested an insurance system which realizes an efficient allocation if the government knows the probabilities (assumed to be fixed) that agents impute to the success or the failure of the vote for a given public project. It is possible to eliminate this strong informational assumption by a process suggested by Savage [1971] which elicits these probabilities (see Kurz [1974]). However, if the agents connect the two steps of Kurz's procedure, the revelation of probabilities and the revelation of willingnesses to pay, they can successfully misrepresent their preferences. Newbery [1974] has given a mechanism with random payments which integrates both steps of Kurz's procedure. However, it remains that truthful revelation requires the hypothesis according to which an agent cannot influence the outcome of the collective decision process.

In this review of efforts to solve the free rider problem we should say a word about Tiebout's [1956] suggestion for local public goods. Tiebout observed that with local public goods there exists an additional dimension which can ideally help to solve the revelation problem. Agents have indeed an additional choice variable, the neighborhood where they choose to live. According to Tiebout, the behavior of local governments combined with the mobility of agents leads to homogenous communities within which problems of divergent tastes disappear. Many difficulties exist however in this proposition and the interested reader is referred to the growing literature on local public goods.[11]

1.4. Dynamic planning procedures

The approach described up to now is a global one, that is to say it is an attempt to define mechanisms for which the strategy spaces of the agents are their own preference relations and which have good incentive properties leading to a satisfactory allocation of resources. A different approach consists in designing multi-stage or dynamic procedures in which questions

[11] See Westhoff [1976] and Stiglitz [1974].

are asked in order to reveal local information about preferences with the hope of converging to a Pareto optimal allocation.

Bowen [1943] was the first to suggest such a procedure, by providing a dynamic version of his voting mechanism which, with the assumption of symmetry of tastes, has the properties that no agent has any incentive to lie and it converges to a Pareto optimal allocation if agents say the truth, at least in systems with a single public good.

Drèze and de la Vallée Poussin [1971] and Malinvaud [1972a] have both constructed, independently, a planning procedure with public goods which has the following properties. Revelation of the true marginal rates of substitution at each instant is a maximin strategy; and in addition, at equilibrium, the mechanism achieves a truthful revelation of marginal rates of substitution as a Nash equilibrium. Champsaur, Drèze and Henry [1977] have embedded this planning procedure for public goods in a competitive economy with private goods, obtaining essentially the same properties. Recent work by Roberts [1976a] and others has pursued the study of incentives properties along the trajectory of this procedure. In part IV we will present a detailed analysis of dynamic planning mechanisms and their incentive properties.

1.5. Empirical evidence on the free-rider problem

Before engaging ourselves in a lengthy theoretical study of the free rider problem it is surely worth asking if there is empirical evidence concerning this well-accepted hypothesis. Surprisingly enough there is no empirical study in which a strong tendency towards free rider behavior is in evidence. However, it must be admitted that the scope and extent of such work is still very limited.

P. Bohm [1972] was the first to report the results of a well-conducted experiment designed for testing the existence of free riding. A sample of the Swedish population, after receiving a fixed fee for participation in the experiment, was confronted with the following: Subgroups of the sample were asked to report their willingnesses to pay for a given television program which was to be shown on a closed circuit network. The questions were put in different ways so as to provide incentives to understate, overstate or correctly state the willingness to pay. The main result of this experiment was that no significant differences in responses under alternative incentives were obtained.

Bohm announced some requirements for such experimental designs

which were essentially satisfied by his own experiment: The outcome of the experiment should actually be implemented, it should involve a large number of agents; the public good itself and its cost should be well defined; and exchange of ideas between the subjects should be possible before requiring them to answer. All these conditions are aimed at reproducing a realistic situation.

The major weakness of the experiment may be that because of the high participation fee they received, members of the population felt a moral obligation to accept losing a small part of their fee in the course of revealing their true preferences. Further, even though the sample was random, only some of those invited actually participated; and they were likely to be a group of concerned citizens. Finally, at the time of this experiment, there were no mechanisms such as those described in this book, for which truthful revelation can be expected in general, to provide a benchmark or control against which to measure the extent of free riding.

L. Johansen [1977] has recently questioned the practical significance of the problem of correct revelation of preferences "at the level of governments, states or countries". His main argument is related to the existence of elected intermediaries in public decision processes, who, because of public debates or because of the need for a simple behavior pattern observable by their constituents, could not afford misrepresentation. This is an interesting point, however, far from convincing on theoretical grounds. For example, the possibility of free riding of the electorate in attempting to influence its representatives' impression of their true preferences is overlooked. The need for more empirical evidence recognized by Johansen is, however, hardly subject to quarrel.

Scherr and Babb [1975] report an experiment in which they compare the results obtained through voluntary payments (which should lead to understatements, through a pricing method (which should induce global overstatement), and through one of the mechanisms to be discussed later in the book, the Clarke mechanism (which should induce exact revelation). They find, as expected, that the answers with the pricing method are larger than under the Clarke mechanism, but that voluntary contributions are larger than both. It appeared through discussions with the agents that the complexity of the mechanisms was the main reason for these awkward results.

Right now, the available evidence is not conclusive.[12] What has been

[12] A few other experiments are available. J. Sweeney [1973] provided a group of undergraduates an opportunity to pedal an exercise bicycle under six alternatives designed to

shown is that in experiments where the good provided is of very limited value to the individual and the cost is very small, diverse motives may counteract the free riding forces in some situations. Interdependent utility functions, demonstration effects or the desire to acquire a social status when anonymity is not guaranteed, and most importantly moral concerns, may indeed be enough to overcome the free riding effect. How far these motives can be pushed is an important sociological question. It is certainly likely, however, that when large amounts are at stake and individuals are anonymous a fair amount of free riding is to be expected.

provide various incentives, and concluded that persons perceiving they are in a small group will contribute more toward a shared goal than those perceiving they are in a large group and that participants in a large group will make greater contributions toward a collective goal if provision of a private good is tied in with it. He obtained significant personal contributions even when agents had an imperceptible impact on the total and where they could not be excluded if the good was provided to the group. However, it is not clear that the undergraduates perceived bicycle pedaling as a real cost to them, as it should have been.

The availability of the mechanisms studied in this book makes new experiments very interesting. Smith [1976] is currently running such experiments in an iterative context with an explicit rule for stopping which insures convergence.

Chapter 2

SOCIAL CHOICE THEORY

2.1. Introduction

The purpose of this chapter is to briefly recall some fundamental results of social choice theory which will enable us to put our work in the broad perspective of the socio-politico-economic theory which has been developed in the last twenty-five years. No comprehensive treatment of this theory is offered. The reader is referred to Arrow [1951] and Sen [1970] for an introduction.

At the origin of this theory we find Arrow's Impossibility Theorem generalizing Condorcet's paradox, the first theorem of political science. This states that there does not exist a non-dictatorial social welfare function satisfying a number of "reasonable" conditions one would be willing to impose on a mechanism designed to aggregate individual preferences. In section 2.2 we review Arrow's theorem and discuss the interpretation of the conditions imposed by Arrow on the functioning of desirable social choice mechanisms.

Then, in section 2.3 we present and prove the most important result of social choice theory from the point of view of this book, a recent theorem by Gibbard [1973] and Satterthwaite [1975], which shows that there does not exist a social choice mechanism which is non-dictatorial and which can never be advantageously manipulated by some agent. Intuitively, this says that all the non-dictatorial social choice mechanisms will be subject to misrepresentation of preferences, if these are privately held pieces of information. This negative result will be interpreted from the point of view of the free rider problem.

The axiomatic structure of these two impossibility theorems shows the directions in which one might hope for positive results by relaxing some of the conditions they impose. A considerable literature has been devoted on the one hand to the strengthening of Arrow's result by weakening some conditions,[1] and on the other hand to obtain "possibility theorems" by weakening them even further. In particular, early positive results have been given to Arrow's aggregation problem by Black [1948], restricting the space of preferences.[2] Recently, Maskin [1976a] has given a complete characterization of the spaces of preferences which admit "satisfactory" social welfare functions (see section 2.4).

In a sense, part of the objectives of this book can be viewed as a similar effort with respect to the Gibbard–Satterthwaite theorem instead of Arrow's theorem. To what extent can strategic manipulation be ruled out or its adverse effects reduced if more structure is imposed on the social decision process and on the individuals involved.

2.2. Arrow's impossibility theorem

Consider a *finite* number N of agents, $i = 1, \ldots, N$, and a set A of alternatives which consists of at least three elements. An alternative or social state is a complete description of the environment and of the activities of the agents. For example, in an exchange economy an alternative is a feasible set of consumption bundles for all the agents. When public goods are present the quantities of public goods and the financing of their production must be specified.

Each agent i has an *ordering*[3] $\boldsymbol{P}_i$ over the set of alternatives A. Let $\Sigma(A)$ denote the class of *allowed* individual orderings on A, and let $S(A)$ be the class of allowed social orderings on A. A generic element of $[\Sigma(A)]^N$, $(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N)$, is a *preference profile*, which is denoted $\boldsymbol{P}$.

Definition 2.1. A *Social Welfare Function* (*SWF*) (or constitution) $\mathscr{F}$ is a function from $[\Sigma(A)]^N$ into $S(A)$, which assigns a social ordering to any allowed preference profile.

Imbedded in the definition of a social welfare function is the requirement of collective rationality: the fact that expressed social choices can be

[1] Arrow's theorem can be strengthened as discussed in Sen [1970, chapter 4].
[2] Positive results are obtained by Sen [1970, chapter 10*].
[3] An ordering is a complete, transitive, asymmetric binary relation. For simplicity of presentation, we exclude indifference but, when not specified, the results of this chapter can be generalized to allow for indifference.

derived from an ordering, instead of a less structured binary relation.[4]

In his pathbreaking study Arrow [1951] has shown that if a number of restrictions are imposed on the social welfare function, the only possible *SWF* is a dictatorial *SWF*. These conditions are stated below.

Definition 2.2. A *SWF* satisfies the *Pareto principle* (*P*) if, when an alternative a_1 is preferred to an alternative a_2 by every individual, then the social ordering also ranks a_1 above a_2. In more formal terms: If $\boldsymbol{P} = (\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N)$ is such that $a_1 \boldsymbol{P}_i a_2$ for all i, then $a_1 \mathscr{F}(\boldsymbol{P}) a_2$.

Definition 2.3. A *SWF* satisfies the *Independence of Irrelevant Alternatives* (*IIA*), if the ranking of two alternatives by the social ordering depends only on the ranking of these alternatives by the agents. That is, for an arbitrary pair of alternatives a_1, a_2 and a pair of preferences profiles $\boldsymbol{P} = (\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N)$ and $\boldsymbol{P}' = (\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_N)$ such that, $a_1 \boldsymbol{P}_i a_2$ if and only if $a_1 \boldsymbol{P}'_i a_2$, then the social welfare function satisfies $a_1 \mathscr{F}(\boldsymbol{P}) a_2$ if and only if $a_1 \mathscr{F}(\boldsymbol{P}') a_2$.

The assumption of independence of irrelevant alternatives is, in essence, a formal way of removing any possibility for the social welfare function to take "intensities" of preference into account. If $a_1 \mathscr{F}(\boldsymbol{P}) a_2$ and a_1 and a_2 are adjacent elements in the orderings of all agents, i, for whom $a_1 \boldsymbol{P}_i a_2$, then the social preference is not allowed to change in favor of a_2 in another situation where a_2 is still preferred to a_1 by all these agents, but there are many other alternatives between them in all their orderings. Point voting and related mechanisms[5] are consequently excluded by virtue of this assumption.

Definition 2.4. A *SWF* satisfies the assumption of universal domain (*UD*) if $\Sigma(A)$ is the class of all orderings on A.

With the requirement of universal domain, no restriction on preferences of the individuals is allowed. The social welfare function must be valid for

[4] Brown [1973] has dealt with the case of an acyclic binary relation, generalizing the results of Sen [1970, chapter 6]. Mas-Colell and Sonnenschein [1972] have used quasi-transitivity instead of full transitivity, but have introduced the additional hypothesis that the social decision function not respond perversely to changes in the individuals' orderings.
[5] The best known of these is the "Borda rule" (see Sen [1970, chapter 7] according to which the alternatives are given points corresponding to their order in the set, and the one with the highest score is the social choice.

any group of agents. If the set of alternatives has a natural linear structure, this rules out conditions such as convexity of individuals' preferences.[6]

Definition 2.5. A *SWF* is *dictatorial* if there exists an agent whose preferences determine the social ordering. That is, $i \in \{1, \dots, N\}$ is a dictator if for all $(\boldsymbol{P}_1, \dots, \boldsymbol{P}_N) \in [\Sigma(A)]^N$, $\mathscr{F}(\boldsymbol{P}_1, \dots, \boldsymbol{P}_N) = \boldsymbol{P}_i$.

Theorem 2.1. (Arrow [1951]). Any *SWF* which satisfies *P*, *IIA*, *UD* is dictatorial.

2.3. Gibbard-Satterthwaite Theorem[7]

The purpose of Arrow's inquiry was the study of the aggregation of given individual preference relations according to a self-consistent scheme. From the point of view of decision making a social ordering is an intermediary step used to define, when it exists, an optimal social state. That is to say, if a single, best, element could be found in the set of social states, we really would not need an entire social ordering. This leads naturally to the concept of a social choice function.

Definition 2.6. A *Social Choice Function* (*SCF*) is a function ℓ from $[\Sigma(A)]^N$ into A, which assigns to any allowed preference profile $\boldsymbol{P}$ an alternative a.

The true preference ordering of each agent is known only to himself. Here, we are envisioning a context in which agents announce their preferences, which serve as the arguments of the social choice function. The objective is now to design a *SCF*, which defines a game for each true preference profile, leading to optimal outcomes.

Definition 2.7. An *SCF* ℓ is *manipulable* at $(\boldsymbol{P}_1, \dots, \boldsymbol{P}_N) \in [\Sigma(A)]^N$ if there exists $\boldsymbol{P}'_i \in \Sigma(A)$ such that

$$\ell(\boldsymbol{P}_1, \dots, \boldsymbol{P}'_i, \dots, \boldsymbol{P}_N)\ \boldsymbol{P}_i\ \ell(\boldsymbol{P}_1, \dots, \boldsymbol{P}_i, \dots, \boldsymbol{P}_N).$$

[6] Convexity of individuals' preferences would not upset the negative conclusion of Arrow's theorem. If, however, the set of alternatives capable of defeating any fixed alternative were convex, then there would exist some alternative that could not be defeated by any other. This has been shown by Sonnenschein [1971] and has been further explored by Zeckhauser [1973].

[7] In this section we draw heavily on the presentation of the Gibbard–Satterthwaite theorem given in Schmeidler–Sonnenschein [1974].

In the above definition, if $(\boldsymbol{P}_1, \dots, \boldsymbol{P}_N)$ are seen as the true preference manipulability means that agent i by announcing $\boldsymbol{P}'_i$ instead of his tru ordering $\boldsymbol{P}_i$ can secure a better outcome for himself.

Definition 2.8. A *SCF* ℓ is *strongly individually incentive compatibl* (*SIIC*)[8] if there is no preference profile at which it is manipulable.

This means that the true ordering is a dominant strategy for any preferenc profile. More formally, for any $\boldsymbol{P} = (\boldsymbol{P}_1, \dots, \boldsymbol{P}_N) \in [\Sigma(A)]^N$ and any $\boldsymbol{P}'_i \in \Sigma(A$ $\ell(\boldsymbol{P}_1, \dots, \boldsymbol{P}_i, \dots, \boldsymbol{P}_N)\ \boldsymbol{P}_i\ \ell(\boldsymbol{P}_1, \dots, \boldsymbol{P}'_i, \dots, \boldsymbol{P}_N)$.

The definition of universal domain for *SCF* extends trivially fro definition 2.4.

Definition 2.9. If ℓ is a *SCF* and ℓ has range $A' \subseteq A$, then ℓ will b called an *SCF with range A'*.

Definition 2.10. A *SCF* with range A' is *dictatorial* if there exists an agen whose favorite announced alternative in A' is always the social choice That is, $i \in \{1, \dots, N\}$ is an *SCF* dictator if for any $\boldsymbol{P} = (\boldsymbol{P}_1, \dots, \boldsymbol{P}_N) \in [\Sigma(A)]$ and any $a \in A'$ such that $a \neq \ell(\boldsymbol{P})$ then $\ell(\boldsymbol{P})\ \boldsymbol{P}_i a$.

Theorem 2.2. (Gibbard [1973] and Satterthwaite [1975]).[9] If A' has a least three alternatives, a *SCF* with range A' satisfying *SIIC* and *UD* is dictatorial.

The proof is now presented in a sequence of lemmas from Schmeidler-Sonnenschein [1974].

Lemma 2.1. If there exists $i \in \{1, \dots, N\}$, such that $\ell(\boldsymbol{P}_1, \dots, \boldsymbol{P}_i, \dots, \boldsymbol{P}_N) = a_1$, $\ell(\boldsymbol{P}_1, \dots, \boldsymbol{P}'_i, \dots, \boldsymbol{P}_N) = a_2$ with $a_1 \neq a_2$, and if $a_2 \boldsymbol{P}_i a_1$, $a_1 \boldsymbol{P}'_i a_2$, then ℓ is manipulable at either $(\boldsymbol{P}_1, \dots, \boldsymbol{P}_i, \dots, \boldsymbol{P}_N)$ or $(\boldsymbol{P}_1, \dots, \boldsymbol{P}'_i, \dots, \boldsymbol{P}_N)$.

Proof. Obvious from definition 7.

Lemma 2.2. If ℓ is *SIIC* with range $A' \subseteq A$, and if $B \subseteq A'$ and $(\boldsymbol{P}_1, \dots, \boldsymbol{P}_N$ is a preference profile such that for all i and all pairs a_1, a_2 with $a_1 \in B$ $a_2 \in A$ and $a_2 \notin B$ (such that) $a_1 \boldsymbol{P}_i a_2$, then $\ell(\boldsymbol{P}_1, \boldsymbol{P}_2, \dots, \boldsymbol{P}_N) \in B$.

[8] Instead of *SIIC* the words cheat-proof, straightforward or strategy-proof are sometimes used in the literature.

[9] Pattanaik [1975] has obtained a similar result.

Proof. Suppose it is not true and let $f(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N) = a_2 \notin B$. Since $B \subset A'$, $\exists$ $(\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_N) \in [\Sigma(A)]^N$ such that $f(\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_N) = a_1 \in B$. Consider now the set of preference profiles $\{a_{i3}\}_{i=0}^{i=N}$ defined by $a_{i3} = f(\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_i, \boldsymbol{P}_{i+1}, \ldots, \boldsymbol{P}_N)$ and let j be the last integer such that $a_{i3} \in B$, then

$$f(\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_j, \boldsymbol{P}_{j+1}, \ldots, \boldsymbol{P}_N) = a_1{}^* \in B,$$
$$f(\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_{j-1}, \boldsymbol{P}_j, \boldsymbol{P}_{j+1}, \ldots, \boldsymbol{P}_N) = a_2{}^* \notin B,$$

and $a_1{}^* \boldsymbol{P}_j a_2{}^*$.

Therefore $f(\cdot)$ is manipulable at $(\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_{j-1}, \boldsymbol{P}_j, \ldots, \boldsymbol{P}_N)$. Q.E.D.

Proof of Theorem 2.2. The idea of this proof is to construct from the *SCF* f, a *SWF* $\mathscr{F}$ which fulfills all the conditions of Arrow's theorem to conclude that $\mathscr{F}$ is dictatorial and consequently also that f is dictatorial.

We now proceed to a construction of a *SWF* $\mathscr{F}$ which fulfills the conditions of Arrow's theorem.

Given any preference profile $(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N)$ and any pair (a_1, a_2), let us construct a modified preference profile $(\tilde{\boldsymbol{P}}_1, \ldots, \tilde{\boldsymbol{P}}_N)$ as follows; $\tilde{\boldsymbol{P}}_i$ is derived from $\boldsymbol{P}_i$ by moving a_1 and a_2 to the top of i's list and preserving the $\boldsymbol{P}_i$ ordering within $\{a_1, a_2\}$ and within $\{a \mid a \in A,\ a \neq a_1,\ a \neq a_2\}$. Let $\phi_{a_1a_2}(\boldsymbol{P}_i)$ denote $\tilde{\boldsymbol{P}}_i$.

For the pair (a_1, a_2) we then write

$$a_1 \boldsymbol{P} a_2 \quad \text{if} \quad a_1 = f(\tilde{\boldsymbol{P}}_1, \ldots, \tilde{\boldsymbol{P}}_N).$$

This process repeated for each pair of alternatives yields a binary relation $\boldsymbol{P}$, and repeated for each allowable preference profile an entire potential *SWF* $\mathscr{F}: (\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N) \to \boldsymbol{P}$, is constructed.

$\mathscr{F}$ satisfies (*UD*) by assumption.

$\mathscr{F}$ satisfies (*P*) on A' by lemma 2.2.

$\mathscr{F}$ satisfies (*IIA*); if not, a proof analogous to the proof of lemma 2.2 would show that f is manipulable.

To complete the argument it remains to be shown that $\boldsymbol{P}$ is an ordering for any $(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N)$. This means that $\boldsymbol{P}$ is *complete*, *asymmetric* and *transitive*. Completeness and asymmetry are obvious. Transitivity is the hard part.

Suppose $\boldsymbol{P}$ is not transitive; then there exists $(\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N)$, a_1, a_2, a_3, with:

$$\begin{aligned} a_1 &= f[\phi_{a_1a_2}(\boldsymbol{P}_1), \ldots, \phi_{a_1a_2}(\boldsymbol{P}_N)], && \text{if and only if } a_1 \boldsymbol{P} a_2, \\ a_2 &= f[\phi_{a_2a_3}(\boldsymbol{P}_1), \ldots, \phi_{a_2a_3}(\boldsymbol{P}_N)], && \text{if and only if } a_2 \boldsymbol{P} a_3, \\ a_3 &= f[\phi_{a_1a_3}(\boldsymbol{P}_1), \ldots, \phi_{a_1a_3}(\boldsymbol{P}_N)], && \text{if and only if } a_3 \boldsymbol{P} a_1. \end{aligned}$$

Let $(\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_N)$ be defined by taking a_1, a_2 and a_3 to the top of i's list and

preserving the $\boldsymbol{P}_i$ ordering within the sets $\{a_1, a_2, a_3\}$, $\{a \mid a \in A,\ a \neq a$ $a \neq a_2,\ a \neq a_3\}$. Then, suppose that

$$a_1 = f[\boldsymbol{P}'_1, \ldots, \boldsymbol{P}'_N].$$

Let $(\boldsymbol{P}''_1, \ldots, \boldsymbol{P}''_N)$ be defined by moving a_2 to the third position in each c the rankings $\boldsymbol{P}'_i$. Since $a_1 \boldsymbol{P}_i a_3$ if and only if $a_1 \boldsymbol{P}''_i a_3$, $a_3 \boldsymbol{P} a_1$ and *IIA* implie $f(\boldsymbol{P}''_1, \ldots, \boldsymbol{P}''_N) = a_3$.

Let $a_{4i} = f(\boldsymbol{P}''_1, \ldots, \boldsymbol{P}''_{(i-1)}, \boldsymbol{P}''_i, \boldsymbol{P}'_{(i+1)}, \ldots, \boldsymbol{P}'_N)$, $i = 1, \ldots, N$, and let j b the first integer such that $a_{4j} \neq a_1$. If $a_{4j} = a_3$, then f is manipulable b lemma 2.1. If $a_{4j} = a_2$, then f is manipulable by j at $(\boldsymbol{P}''_1, \ldots, \boldsymbol{P}''_j, \boldsymbol{P}'_{(j+1)}, \ldots$ $\boldsymbol{P}'_N)$. This contradicts the non-manipulability of f and therefore establishe the transitivity of $\boldsymbol{P}$.

From Arrow's theorem we know that $\mathscr{F}$ is a dictatorial *SWF*. Th dictator for $\mathscr{F}$ is also a dictator for f. Q.E.D.

2.4. Possibilities for positive results

Theorem 2.2 shows that there is no social choice procedure which car provide efficient choices and truthful revelation of preferences unde general circumstances. The conditions of this theorem themselves sugges the possibility of generalization and of positive results.

The first direction for generalization would allow for more complex strategy spaces and social decision rules than simply having agents announce preference orderings. This approach has been taken by Gibbard and Satterthwaite, leading to an extension of the negative results demonstrated above to arbitrary strategy spaces. (Since "truth telling" may not have a precise meaning in these contexts, it is replaced by the requirement that some dominant strategy exists corresponding to each preference relation.) The importance of this generalization is that arbitrary strategy spaces allow, in principle, for sequential social choice procedures, complex point-voting rules and systems of representative governments. No such ideas, no matter how complicated, can preserve incentives in the form of dominant strategies without losing either optimality or non-dictatorship.

In the search for positive results, other avenues must be pursued. A logical choice is to look for restrictions on the domain of preferences. These allow solutions to Arrow's problem to be found and the close connection between these results hints at successful application here. For example, single peakedness admits the median voter rule as a solution to

Arrow's problem, and with the set of alternatives as the strategy space, this social choice function is also non-manipulable. Recently, the precise relationship between these problems[10] has been given by Maskin [1976b] and by Kalai and Muller [1976]. If the set of social states is fixed, the class of preference orderings for which Arrow's axioms are consistent coincides with the class for which non-manipulable social choice functions exist.[11] Moreover, this class of individual orderings is independent of the size of the population whose tastes are to be aggregated. Unfortunately the characterization of this class is not very transparent, nor can it be stated in constructive terms that can readily be checked by methods short of complete enumeration. The real power of these results is in the unification they provide for these two theories.

The approach taken by Maskin and Kalai and Muller is special in two respects. First, it is highly symmetric across persons in that the same domain restriction is applied to each agent. Second, it is only the domain of preferences that is restricted; the frequency distribution of preferences within the allowable domain is arbitrary. Viewing the distribution of tastes present in the population as a probability measure over the set of orderings, it is as if only the supports of these measures were constrained, and the social choice function were required to operate on the domain of all measures with support contained in the indicated set.

Both of these features suggest further generalization in the search for positive results. Symmetry, anonymity or other egalitarian principles are generally regarded as desirable features of social choice processes. This was certainly the motivation behind the identical domain restrictions used in these results. However, the social choice processes that involve the government or a central planner as an active participant may require a different strategy space for this agent than for all the others. This point needs some elaboration for it is intimately linked with the description of the set of alternatives, or social states.

Up to now, the set of social states has been regarded as an abstract collection. But more structure is often built into the problems for which social choice functions are used. In particular in resource allocation

[10] The results as characterized below hold only if indifference between social states is not allowed. They become somewhat more complex if it is permitted, but their general character is unchanged.

[11] More precisely there exists an N-person nondictatorial SWF satisfying (P), (IIA) and a monotonicity assumption on restricted domains $\Sigma_1(A) \times \ldots \times \Sigma_N(A)$ if and only if there exists a "selfconsistent" collection of SCF which are $SIIC$ (see Maskin [1976a]).

situations, the possibility of free disposal introduces a natural way of appending an extra agent onto the set of actual participants. This agent, the government or planner as he may be called, serves as a source or sink for unused resources. Keeping to the spirit of our original social choice problem means that the government's tastes must be neutral as between alternative private resource allocation patterns – unlike the agents themselves whose actual preferences are the heart of the matter. Incorporating the government in the planning process means that its preferences are really specified in advance as part of the system itself and are not free to vary as those of the agents. Typically the government is assumed either to have monotonic preferences in the surplus goods it collects, or to be indifferent to all social states. Formally speaking, we are outside the context of the Maskin–Kalai–Muller analysis and new incentive compatible social choice processes may be found with superior properties. In this book we will be exploring precisely such a class of processes, in which the central planner's asymmetry plays a key role.

Another type of generalization that may be useful for obtaining positive results is the postulate that statistical information about the distribution of tastes rather than merely knowledge of the domain of preferences, is available. This may be valuable, especially if approximation results are the objective. Grandmont [1976] has taken this approach by placing restrictions on the qualitative nature of the distribution of preferences in the population. He has shown that if it is possible to parameterize preferences in such a way that the induced distribution of tastes in the parameter space is symmetric in all directions, then majority rule is a consistent social welfare function – that is, it is non-dictatorial and satisfies the criteria of Pareto optimality and independence of irrelevant alternatives.

Random social choice procedures in place of deterministic ones may be another fruitful direction to pursue. In this framework a social choice function maps announced preferences into lotteries over the alternative social states. This is suggested by schemes such as choosing an agent at random and using his announced preferences as the social ordering. Unfortunately, as long as Pareto optimality is maintained as a strict requirement and no domain restrictions are imposed, it can be shown that there are no random social choice functions other than this random dictator scheme. Indeed, Gibbard [1977] has shown that even if Pareto optimality is dropped completely, the requirement that truthful revelation of preferences be a dominant strategy for all agents limits the set of random social choice functions to probability mixtures between two schemes. The first is the random dictator and the second is a two stage process in which a random pair of alternatives is selected and a voting rule using the announced preferences then determines the social choice between these two. On the basis of these results it would appear that randomness alone cannot improve matters very much. In this book we will show, however, how randomness can lead to much greater efficiency and stronger incentive properties than can be achieved by either domain restrictions or weakening the dominant strategy requirement alone.

Before discussing the latter possibility, which markedly increases the complexity of the strategic considerations we turn to another potential weakening of the Gibbard–Satterthwaite postulates that is quite natural for economic theory. Very often the space of alternatives has a natural topological structure. For example if they are allocations of commodities, they can be viewed as points in a Euclidean space, and agent's preferences can be assumed to be continuous in this topology. In this way we can speak of approximate Pareto optimality even though cardinal preference relations have not been specified. We can then seek mechanisms that achieve approximate Pareto optimality with certainty, or stochastically with high probability. Such an approach is taken in part III of this book.

In a related manner we can expand the set of social states under consideration beyond those that are actually feasible, and look for mechanisms that select approximately feasible points as their dominant strategy equilibria.

Finally, we must discuss the possibility of obtaining positive results on the preference revelation problem by weakening the requirement that a dominant strategy exists for each agent. There are many possibilities in this direction. All of them require the basic specification of two additional characteristics of the agents – their attitudes toward risk and their expectations concerning the strategies to be played by the other individuals. If their beliefs are independent of the strategies actually played, the optimal actions will constitute an equilibrium point of a game with imperfect information. This has been explored by d'Aspremont and Gérard-Varet [1977] and by Arrow [1977]. In the case of extreme risk averseness, we are led to the maximin concept, as long as beliefs about the others are diffuse. This approach has been taken by Dubins [1974]. In chapter 7 we study these procedures, and others, which do not rely on the existence of dominant strategies.

At the other extreme in terms of hypotheses concerning knowledge of others' strategy choices we have the Nash equilibrium concept in which each agent's action is optimal given the actions of everyone else. Unlike the concept of equilibrium in expectations, Nash equilibria require a

process of successive responses until an equilibrium can be reached.[12] For this reason they may be much more costly in terms of their requirements for communication – costs which are hard to model precisely but are very real nonetheless. The Nash concept may still offer the best hope for positive results and may realistically describe behavior when actual strategic choices are observable and sufficient time exists to revise strategies in response to these observations. In the general social choice context, Schmeidler and Hurwicz [1976] have given a mechanism that has several desirable properties when the Nash equilibrium concept is employed. In a more restrictive resource allocation setting, but still a rather general one, Groves and Ledyard [1975] have given a mechanism all of whose Nash equilibria are Pareto optimal.

[12] We believe that this interpretation of Nash equilibria is the only one possible in our context of imperfect information. It would be meaningless to presume that agents are able to compute the best strategies of the others as it is generally assumed in game theory.

PART II

Chapter 3

THE BASIC MODEL

3.1. Decisions about fixed size projects

In the light of part I, a general solution to the incentive problem seems beyond reach. In the present chapter we concentrate on a highly restrictive case in which a positive result is attainable. We envision the social state as being describable in two parts. There are those aspects that are of a purely private nature, and there are those public aspects which are perceived in common by all agents. The decision we consider first in this chapter concerns changing the public characteristics of a given social state, the status quo, to a well-defined proposed alternative, but not disturbing the private aspects of the system.

We suppose that all individuals can evaluate these two social states precisely. That is, they fully comprehend both the status quo and the alternative in all relevant details. This assumption is critical to our entire study (except in chapter 13 where we consider relaxing it to some extent) and yet it may be a very poor one in practice. But without something of this sort there is little basis on which rational social decision processes can be built. The assumption that the resulting social state is a perfectly predictable outcome of the changes contemplated may also be quite problematical. Consider the proposal of reducing the speed limit on highways. To evaluate the effect of this measure requires the individual to forecast how his fellow-motorists will react to the new law, and then to be able to predict the resulting traffic pattern, in space and time. Matters may be even worse when the proposed change cannot be viewed as a small one and when multiple equilibria raise their ugly heads. In the speed limit example the price of gasoline may be affected, and the general equilibrium consequences of this may not be well-defined at all. Taking an expected utility viewpoint, the participants in our system will be required to have subjective probability distributions over the workings of the economic process – most of which is unobservable – in order that they can derive expectations about the ultimate outcome of the alternative policy.

For the present we take the bold step of believing that this process of evaluation and comparison can be carried out costlessly and precisely. In many instances this may not be wrong. Small groups engaged in simple choices are the best candidates for such an analysis: a group of office workers deciding on whether to buy a coffee pot to be shared equally, a committee considering the election of its chairman from among two candidates, two business partners deciding who will take his vacation first, assuming that one must remain on the job at all times, or a town deciding whether to build a new public swimming pool are some examples.

The idea behind our analysis is that the decision-making mechanism is to be designed in advance of its actual implementation. One often thinks that the mechanism, once established is to be used in many particular separate instances. In each application it must produce a definite outcome on the basis of the announced preferences of the participants in such a way that an optimum is attained relative to the true preferences.

Requiring merely Pareto optimality in the two-alternative case will hardly limit the outcomes at all. Unless there is unanimity, either decision will qualify as an optimum. Simple decision processes such as majority rule will have the appropriate incentive properties and will always result in an outcome selected within the optimal set. More generally, however, when there are three or more alternatives, the results of part I indicate that either incentives or efficiency would rapidly break down.

To avoid these difficulties in multi-option cases we take the seemingly paradoxical step of further increasing the set of possible social decisions – but in a very special way. We assume that among the private aspects of the social state there exists a transferable resource. In this way we reduce the set of public aspects of those decisions that are associated with Pareto optimal social states. Many public options, even if favored by some agents with the initial private resource distribution, can be dominated by other decisions in which a system of compensatory transfers of this private good has been arranged. Although the transferability of this resource apparently increases the complexity of the social decision, it will be seen that it permits us to arrange for the appropriate private incentives necessary to select optimal outcomes, at least when some further restrictions on preferences are employed.

3.2. Costless projects and financing decisions

When the proposed alternative would involve the expenditure of real resources, the transferable private resource plays a second role. We will be assuming that any such costs can be denominated in units of this resource. Because this resource is initially held by individuals in the system, their shares in the financing of the alternative would certainly influence their preferences regarding its adoption. (Since in this chapter we are concerned with only a single alternative to the status quo, we will not consider choices among different financing plans for the same public proposal[1]). Thus, a specification of the cost shares must be presented as part of the description of the alternative itself.

Let us consider, first, an alternative that is costless to implement. The aspects of the social state relevant to any agent are the public decision taken and his level of consumption of the transferable resource.

We denote by:

x_i the consumption of the transferable resource by agent i,
d the decision taken (or the project selected),
$\{0, 1\}$ the set of possible projects,

where:

$d = 0$ means "reject alternative project",
$d = 1$ means "accept alternative project".

A utility index for individual i can be written as

$$U_i(d, x_i), \tag{3.1}$$

which is just equivalent to specifying two utility indices for the private good under the two possible social decisions. These indices can be written:

$$U_i^0(x_i),$$

and (3.2)

$$U_i^1(x_i),$$

respectively. Both are assumed to be strictly monotone increasing in x_i.

Thus far these are only ordinal representations of the underlying preferences. A particular cardinalization will be of interest. Let us consider the transformation:

[1] See section 4.6.

$$\tau : \mathbf{R} \to \mathbf{R}, \tag{3.3}$$

given by

$$\tau = U_i^{0^{-1}} \tag{3.4}$$

It is well-defined, by the monotonicity of U_i^0.

The utility function $\tau(U_i(d, x_i))$ clearly represents the same preferences as $U_i(d, x_i)$. Writing

$$u_i^0(x_i) = \tau(U_i(0, x_i)) = \tau(U_i^0(x_i)), \tag{3.5}$$

and

$$u_i^1(x_i) = \tau(U_i(1, x_i)) = \tau(U_i^1(x_i)), \tag{3.6}$$

we see, by definition, that $u_i^0(x_i) = x_i$. The quantity $u_i^1(x_i) - x_i$ has a ready interpretation. It is the amount of the private good having the property that the individual with an initial consumption level of x_i would be indifferent between this amount of additional consumption or keeping the level at x_i or having the project adopted instead. Of course this is not necessarily the same quantity as that which would make him indifferent between having the project rejected and a consumption of x_i or having it accepted, but being forced to give up that amount.

Noting this distinction, we write

$$v_i(x_i) = u_i^1(x_i) - x_i, \tag{3.7}$$

and refer to this quantity as the willingness to pay for the project at income level x_i.

This construction can be extended directly to the case of costly projects. Let c denote the individual's cost share if the project is accepted. The amount needed to compensate him for the loss of the project, when his income with it would be x_i is

$$u_i^1(x_i) - x_i - c. \tag{3.8}$$

This is because c is refunded to him if the project is cancelled. Thus, in the case of costly projects, we can refer to his willingness to pay as

$$v_i(x_i) = u_i^1(x_i) - x_i - c, \tag{3.9}$$

observing as before that this differs from the amount he would be willing to give up, in addition to the cost share, in the event the project is adopted and his initial income is x_i.

Comparing (3.7) and (3.9) we see that, by definition, the willingness to

pay for a costly project is exactly c less than if it were costless. We therefore follow the convention of including individuals' cost shares in the specification of the project. Overall, the entire proposal can be regarded as a single costless project throughout the analysis, with the valuation function given by (3.9). We will follow the convention of reasoning with costless projects throughout most of this book. It will be shown in section 4.6, that there is no real alternative to this approach; the financing decision cannot be effected simultaneously with the elicitation of preferences. Special problems involving costly projects in some other circumstances will be discussed in chapters 7, 15 and 16.

3.3. Willingness to pay and the separability assumption

Despite the strong assumptions made thus far, we will be unable to give a satisfactory solution to the preference revelation problem without a further restriction on utility functions. This is the condition that $v_i(x_i)$ can be written in the additively separable form

$$v_i(x_i) = v_i + x_i, \tag{3.10}$$

where v_i is a constant.

This is just the statement that the project is not subject to any income effects. The two measures of compensation – for having the project or for being denied it – mentioned in the last section, will coincide at v_i. Without ambiguity therefore we are entitled to call v the *willingness-to-pay* for the project.

It is important to keep in mind that v_i may be negative, for it includes the cost c assigned to the individual. If the underlying preference relation had the property that the project is surely desirable in the absence of costs then $-c$ could be considered a lower bound on the willingness-to-pay. However, since we believe this method to be applicable to decisions more general than that concerning "building public projects" in the traditional sense, no a priori bounds on v_i can be given. Reconsider, for a moment, the examples of the speed limit decrease, the committee chairmanship election and the vacation sequencing problem discussed in section 3.1. In each of these cases, some individuals may have large negative evaluations for one of the alternatives. Only in the speed limit case are there any obvious real resource costs: those of posting new signs and perhaps tightening enforcement policy; and these seem unrelated to the individual's welfare gains or losses that would result from such a decision.

Moreover, we have been generally stating the public decision problem in an unnecessarily asymmetric fashion. The "proposed alternative" may actually be to abandon a project or an ongoing program. The cost shares may be rebate shares instead. Our previous argument would then indicate that, although there may be some cases in which an upper bound can be placed on v_i, we prefer to think of v_i as an unconstrained parameter for each individual.

There is one final remark to be made concerning preferences. Our interpretation of v_i as a willingness-to-pay or alternatively as a necessary compensation level hinges on not bounding potential consumption of the transferable private resource from below. An individual with less than v_i units of this resource at his disposal initially, and who is indifferent between having the project accepted or receiving v_i units of it, will still be unable to pay this much, unless negative consumption can be given a meaningful interpretation. It is therefore incorrect to say that his willingness-to-pay is v_i. This problem goes back to our original specification that preferences be defined over $\{0, 1\} \times \mathbf{R}$ and not, say, over $\{0, 1\} \times \mathbf{R}_+$. For the present we will maintain the symmetry of the problem by allowing negative consumption of the transferable resource, in principle. In section 5.3 we will consider the implications of bounding it below.

3.4. Social states, allowable transfers and the characterization of optimality

The goal of the decision-making procedures that we envision is to choose a social state in an optimal manner, in the absence of information on individual's preferences. It is therefore necessary to clarify the set of social states. In any situation, the decision concerning the project must be uniquely determined. Consumptions of the transferable private good will generally differ among the individuals. If there are N individuals,

$$a = (d, x_1, \ldots, x_N), \tag{3.11}$$

is a social state, where x_i is the consumption of individual i, $i = 1, \ldots, N$.

Typically in economic theory, one considers situations in which individuals have an initial endowment of the transferable resources. Let the non-negative number, $\bar{x}_i$, be the initial endowment of individual i. Then, since the resource is freely transferable, the requirement of feasibility would be

$$\sum_i x_i \leqq \sum_i \bar{x}_i. \tag{3.12}$$

In the context of public decision-making, however, we envision a somewhat more flexible situation. There is another economic agent, hidden

behind the scenes, whom we can identify with the public agency or government charged with undertaking the decision-making process itself. This agent is endowed with the transferable resource and it is a valuable commodity for him as well. Nevertheless he may draw on his stock of his resource as required by the procedure if transfers to the other agents are necessary. Since we have neglected, temporarily at least, any constraints on the consumptions of individuals,we consider the set of all social states to be

$$A = \{a \mid a = (d, x_1, \ldots, x_N)\} = \{0, 1\} \times \mathbf{R}^N, \tag{3.13}$$

without further qualification.

Implicitly we understand this to mean that the net transfer from the agents to the decision-maker is

$$\sum_i (\bar{x}_i - x_i). \tag{3.14}$$

Individuals' preferences can be defined directly on A, by considering the appropriate projection of A onto $\{0, 1\} \times \mathbf{R}$. Without fear of confusion it is possible to write

$$u_i(a) = u_i(d, x_i). \tag{3.15}$$

Our definition of Pareto optimality respects the transfer to the outside agency:

Definition 3.1. A social state $a \in A$ is said to be *Pareto optimal* if for any $a' \in A$ involving transfers to the outside agency at least equal to those of a,

$$u_i(a') \geqq u_i(a), \qquad i = 1, \ldots, N,$$

implies

$$u_i(a') = u_i(a), \qquad i = 1, \ldots, N.$$

Equivalently we could give the outside agency an utility function, $u_0(a)$, such that, if $a = (d, x_1, \ldots, x_N)$, then $u_0(a) = -\sum_i x_i$. Pareto optimality would then be defined as usual for the $N + 1$ agents, $i = 0, \ldots, N$.

The special nature of the utility functions and social states we have described allows us the following obvious, but useful, result:

If $\sum_i v_i > 0$ then the set of Pareto optimal states is $\{a \mid a = (d, x_1, \ldots, x_N), d = 1\}$.

If $\sum_i v_i < 0$ then the Pareto optima are $\{a \mid a = (d, x_1, \ldots, x_N), d = 0\}$. That is to say, the set of Pareto optima can be completely characterized by the public decision taken. We can thus refer to a decision, d, as being a Pareto optimal decision, if

$$d = 1 \quad \text{whenever} \quad \textstyle\sum_i v_i > 0$$

and (3.16)

$$d = 0 \quad \text{whenever} \quad \textstyle\sum_i v_i < 0.$$

In the case where $\sum_i v_i = 0$, the set of Pareto optima coincides with the entire set A.

3.5. Mechanisms as non-cooperative games

As in part I, the essential difficulty to be overcome is that the willingnesses-to-pay v_i, $i = 1, \ldots, N$, are unknown at the time the decision-making process is devised.

Since the goal is the attainment of Pareto optimality it is really not necessary to discover the v_i, per se. However, as we shall see, optimal decision making processes can sometimes be viewed as doing exactly that. Our objective is to design procedures for communication and decision making that will result in a Pareto optimal outcome in every instance. The ground rules under which we work are that the agents are assumed always to act in their own interest, and that they understand the workings of the procedure in which they are embedded. They are presumed to be ignorant of the true utility functions of the other individuals who are participating in the procedure.

More formally, we make the following definition:

Definition 3.2. A *mechanism* is a list of N sets $S_1, \ldots, S_N$, and a function, f, $f: S_1 \times \ldots \times S_N \to A$.

Denoting $S = S_1 \times \ldots \times S_N$, we may refer to the mechanism (S, f). The sets, S_i, are called strategy spaces, and are for the present just abstract sets.

A mechanism is not a game because the utilities have not yet been specified. Gibbard [1973] has used the term "game form". In the language of chapter 2, it is a social choice function,[2] but with perhaps more general strategy spaces than the set of allowable preferences.

Assigning an utility function to an agent we can discuss how he might

[2] As defined a mechanism cannot be dictatorial. The ith agent would always want more of the private good than he is given at any $a \in A$.

play the resulting game. In general, given expectations about the strategies to be played by the other agents, each of his strategies induces a probability distribution over the outcomes of A. The utility function we discussed above describes only an ordinal ranking of A. The individual's attitude towards risk will determine which strategy is optimal for particular expectations. Since many preferences for lotteries over the outcomes are consistent with the same ordering for sure prospects, the individual's choice of a strategy will generally not be the same for different risk preferences. If, however, there existed a strategy such that no superior outcome could be attained by any other strategy in any event, then his expectations and attitude towards risk would not be a determinant of the optimal strategy. Otherwise, more explicit assumptions regarding the role of uncertain beliefs in individual decision making would have to be made.

3.6. The dominant strategy property

Consider a typical individual i; he is uncertain about the vector $(s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_N)$ which represents the strategies selected by the other agents. Let us denote this vector by s_{-i}, and write $s = (s_1, \ldots, s_N)$ as (s_{-i}, s_i).

Definition 3.3. A strategy $s_i \in S_i$ is said to be *dominant* for the mechanism (S, f) if $u_i(f(s_{-i}, s_i)) \geqq u_i(f(s_{-i}, s_i'))$ for all $s_{-i} \in S_1 \times \ldots \times S_{i-1} \times S_{i+1} \times \ldots \times S_N$ and for all $s_i' \in S_i$.

The great advantage of mechanisms such that every agent will always have a dominant strategy is that they remove the game theoretic aspect of the decision-making procedure, at least as far as individual behavior is concerned. No one can conceivably gain by playing any other strategy. Moreover, no threats by any individual to deviate from his optimal strategy can be viable; that is to say, they will not cause anyone to change his behavior, and therefore cannot be to the advantage of the threatener. However, cooperative play may be to the mutual advantage of a group of individuals. Such behavior requires communication among the agents and, while it cannot be ruled out on a priori grounds, it can be treated to some extent within the present framework.[3] For the most part, this book will treat the individual as the sole decision-making unit. There is strong

[3] See chapter 10.

reason to believe that if dominant strategies exist, individuals will use them.

Although the requirement that dominant strategies exist is a strong one, there are some interesting processes that have it. In the present context, with only two possible values for the social decision, majority rule, without any monetary transfers being made, is a mechanism with dominant strategies for every agent. Voting against one's more preferred alternative is never superior to voting in favor of it, as it can only worsen the result in borderline cases. Unfortunately the results of this voting procedure may be far from optimal; the definition of optimality we have given depends heavily on the intensity of individuals' preferences, and majority rule eliminates such considerations.

Observe that revealing one's true preferences is the dominant strategy in this example. Whenever an agent's strategy space coincides with the space of potential preference patterns, we may be interested in inducing truth-telling behavior per se. As a first step we consider in this chapter only mechanisms designed to produce truth-telling dominant strategies and Pareto optimal outcomes. Of course, truth telling, for its own sake, is not a genuine objective in the design of mechanisms. We consider more general situations in the next chapter.

A second example is due to J. Marschak[4] – called a seller's price auction. A seller of an object fixes a price in advance which is unknown to the potential buyer. If the buyer bids any amount above the seller's price, then he acquires the object and pays the price previously specified by the seller. If he bids less, the seller retains the item and no transfers are made.

The essence of the mechanism studied in most of this book is really contained in this example. The buyer will always bid the true value of the item to him. Exaggerations of this can only lead him to be forced to buy at some unfavorable prices, and reductions may cause him to lose some opportunities for desirable transactions.

On the seller's side, however, it is not a dominant strategy to set his true reservation price. By exaggerating it slightly he may lose a few valuable customers, but he will increase his revenues in many other cases. If the seller is at all uncertain about the buyer's bids, he should set a price above the true reservation price. He has no dominant strategy.

[4] J. Marschak described this idea in an unpublished lecture about 20 years ago.

3.7. Satisfactoriness

Since the goal of our analysis is to construct mechanisms whose outcomes are optimal, we should not care about the dominant strategy property, per se. It only provides a convenient way of isolating a particular[5] outcome, the dominant strategy equilibrium, on which we can fix our attention, and at the same time, allows us to neglect an explicit treatment of individuals' expectations. If the dominant strategy equilibrium were also an optimum, then, in a sense, we would be in the best of all possible worlds.

Definition 3.4. A mechanism (S, f) with the properties that for every agent, i, and every willingness-to-pay, v_i, there exists a dominant strategy $s_i^* \in S_i$ and for every vector $(v_1, \ldots, v_N)$, $f(s_1^*, \ldots, s_N^*)$ is a Pareto optimum is called *satisfactory*.

Neither of the examples of the last section were satisfactory; however a bidding system related in spirit to the seller's price auction can be constructed in a satisfactory way. This was first pointed out by Vickrey [1961]. We refer to this mechanism as the "second-price auction".

We consider an item to be allocated between two buyers[6], according to the following rule. Each buyer writes down, independently, the amount he supposedly would be willing to pay for the item. Of course these need not be truthful. The one whose bid is higher gets it. But the price he pays is not what he wrote, it is that written by his opponent.

It is easy to see that each individual is placed in exactly the same situation as the buyer in the seller's price auction. Here, however, there is a crucial difference in the way in which the private good is allocated. In the seller's price auction the seller receives the resources given up by the buyer; but here, the loser of the auction receives nothing. In particular, he cannot influence his level of consumption of this resource except through making a winning bid – all losing bids are rewarded equally, independently of the situation of the winner. It is necessary that the revenue collected accrue to an outside agent or the auctioneer, and not be redistributed to the participants themselves.

[5] In the case in which there are multiple dominant strategies, and hence multiple dominant strategy equilibria, the outcomes generated must all be within the same indifference class for agents whose dominant strategy is not unique. These cases are studied in section 4.5; see theorem 4.8, in particular.

[6] One of these individuals can be thought of as the "seller" of the "seller's price auction", whose willingness-to-pay is just his reservation price.

3.8. The idea behind satisfactory mechanisms

The mechanism for allocating a private resource between two buyers suggested above is reminiscent of a much more familiar economic prescription: Everyone who causes an externality should be taxed according to the full social cost that he imposes on others. In the present context, the structure of the externality is not a datum, but rather the mechanism itself defines the way in which strategies cause the social state selected to vary, and hence the way in which the welfare of others is affected. When one buyer changes his bid so that he now receives the item instead of his opponent, the amount to be paid is the cost he imposes on the rest of the system. This is just the value that the item would have had for his opponent had he received it.

There is, however, one crucial difference between the externality-internalizing prescription and construction of satisfactory mechanisms. In the former, the utility functions of all agents are implicitly assumed to be known – social action is required only to create the appropriate incentives for internalization. But the present context is one where the planner's information is highly imperfect; indeed he is assumed to have none at all. The burden placed on satisfactory mechanisms is that they must overcome this informational imperfection without destroying the internalization feature.

Satisfactory mechanisms accomplish this by separating the two phases of the public decision-making process. Truthful preferences are elicited by making the individual's payoff a step-function of his strategy, the size of the step being equal to his strategy at the point of discontinuity. The optimality of the outcome is insured by making the point of discontinuity vary with the statement of the other individual so that it always occurs where the optimal social decision should switch, assuming truthful revelation by both parties.

The burden of incomplete information does, however, have its costs. With perfect information, nonconvexity problems aside, the planner need only satisfy the first-order conditions in order to attain a first-best situation. By employing a satisfactory mechanism, however, the agent to whom the item goes is taxed and this amount of the transferable resource must leave the system. If the planner is a government or another central agency which has a use for these revenues, the mechanism might not be thought of as inefficient. But in small group planning procedures, such as the vacation sequencing problem or the coffee pot problem, implementation of the analogous satisfactory mechanism would entail giving any of the resource

generated to an outside agency, or wasting it in some other way, to no advantage of the participants.[7]

The term "satisfactory" can therefore be justifiably challenged in such circumstances. Although Pareto optimal under our definition, this satisfactory mechanism at least is not optimal from the point of view of the individuals who would undertake to employ it. The lost revenues are the price of the imperfection of information. Later in this book we will study whether, and how, the cost described above can be overcome. In order to approach this question we first try to see if perhaps there are other ways of constructing satisfactory mechanisms. Though there are, in fact, many such mechanisms, there is no way to avoid the problem of lost revenues completely.

3.9. Groves mechanisms and the pivotal mechanism

As a prelude to the more general analysis to follow, we continue with our study of the two alternative case, using the model of the second price auction as an example. Let the "status quo" decision be that individual 1 obtains the good and let the alternative be that it goes to individual 2. Writing the rules of the second-price auction in mathematical form[8] we have:

$$\begin{aligned} &S_1 = S_2 = \mathbf{R}, \\ &f(s_1, s_2) = (d(s_1, s_2), \bar{x}_1 + t_1(s_1, s_2), \bar{x}_2 + t_2(s_1, s_2)), \end{aligned} \tag{3.17}$$

where:

$$\begin{aligned} d(s_1, s_2) &= 1, && \text{if and only if } s_2 \geqq s_1, \\ t_1(s_1, s_2) &= -s_2, && \text{if } s_1 > s_2, \\ &= 0, && \text{if } s_2 \geqq s_1, \\ t_2(s_1, s_2) &= -s_1, && \text{if } s_2 \geqq s_1, \\ &= 0, && \text{if } s_1 > s_2. \end{aligned} \tag{3.18}$$

The payment made by the agent i is $-t_i$, so that his consumption level can be written as the initial endowment $\bar{x}_i$ plus t_i.

[7] The impossibility of closing the system while maintaining satisfactoriness is proven in section 5.3, theorem 5.4.

[8] To complete a specification of the second-price auction we must associate a social state to all patterns of bids, and in particular to cases in which there is a tie in the bids. We adopt the convention of giving the good to agent 2 (the "buyer") in this event, without loss of generality.

It is easy to generalize this system by adding, for each agent, a function depending only on the other's strategy, but not on his own. This has the effect of shifting his utility as a function of his own strategy by a constant. The location of the optimum remains unchanged, and as this is true for every value of the strategy of the other, the existence of a dominant strategy has remained undisturbed. Incorporating the constants $\bar{x}_i$ into this function, we have a class of mechanisms defined by:

$$\begin{aligned} d(s_1, s_2) &= 1, && \text{if and only if } s_2 \geqq s_1, \\ x_1(s_1, s_2) &= -s_2 + h_1'(s_2), && \text{if } s_1 > s_2, \\ &= h_1'(s_2), && \text{if } s_2 \geqq s_1, \\ x_2(s_1, s_2) &= -s_1 + h_2'(s_1), && \text{if } s_2 \geqq s_1, \\ &= h_2'(s_1), && \text{if } s_1 > s_2, \end{aligned} \tag{3.19}$$

where h_1' and h_2' are arbitrary real-valued functions.

This can be brought into an equivalent, but more readily interpretable, form by regarding the variables w_1 and w_2 as the strategies and using the transformations

$$\begin{aligned} w_1 &= -s_1, \\ w_2 &= s_2, \end{aligned} \tag{3.20}$$

and

$$\begin{aligned} h_1(w_2) &= h_1'(s_2) - s_2, \\ h_2(w_1) &= h_2'(s_1). \end{aligned} \tag{3.21}$$

Thus we obtain the mechanism:

$$\begin{aligned} d(w_1, w_2) &= 1, && \text{if and only if } w_1 + w_2 \geqq 0, \\ x_1(w_1, w_2) &= w_2 + h_1(w_2), && \text{if } w_1 + w_2 \geqq 0, \\ &= h_1(w_2), && \text{if } w_1 + w_2 < 0, \\ x_2(w_1, w_2) &= w_1 + h_2(w_1), && \text{if } w_1 + w_2 \geqq 0, \\ &= h_2(w_1), && \text{if } w_1 + w_2 < 0. \end{aligned} \tag{3.22}$$

In this way we can regard w_1 and w_2 as the professed willingnesses to pay for the project ($d = 1$). This transformation is therefore the natural one, since individual 1's willingness to pay for the loss of the good should be the negative of his bid.

This form of the mechanism suggests a natural generalization to the N-person case. For each agent, i, the jump discontinuity in his allocation of the private good must occur where $w_i = -\sum_{j \neq i} w_j$. On either side of this discontinuity, any constant payoff will maintain the dominance of the

truthful strategy. In this way we derive a class of mechanisms which we call Groves mechanisms.

Definition 3.5. A mechanism (S, f) is called a *Groves mechanism* if

$$\begin{aligned} S_i &= \mathbf{R}, && i = 1, \dots, N, \\ d(w) &= 1, && \text{if } \textstyle\sum_i w_i \geqq 0, \\ &= 0, && \textstyle\sum_i w_i < 0, \\ x_i(w) &= \textstyle\sum_{j \neq i} w_j + h_i(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_N), && \text{if } \textstyle\sum_i w_i \geqq 0, \\ &= h_i(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_N), && \textstyle\sum_i w_i < 0, \end{aligned}$$

where $h_i(\cdot)$ is an arbitrary deterministic real-valued function, $i = 1, \dots, N$.

Let us introduce the notation

$$w_{-i} = (w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_N). \tag{3.23}$$

The following theorem due to Groves and Loeb [1975] is a formal statement of the desirable properties of mechanisms in this class.

Theorem 3.1. All mechanisms in the class of Groves mechanisms are satisfactory. Moreover, the truth is the dominant strategy for all agents.

Proof. Consider the utility obtained by an agent whose true willingness to pay is v_i and who plays the strategy, w_i. Let (S, f) be a Groves mechanism. Suppose that

$$d(w_1, \dots, w_N) = d(w_1, \dots, w_{i-1}, v_i, w_{i+1}, \dots, w_N). \tag{3.24}$$

Then according to the definition of Groves mechanisms,

$$x_i(w_1, \dots, w_N) = x_i(w_1, \dots, w_{i-1}, v_i, w_{i+1}, \dots, w_N), \tag{3.25}$$

and thus the individual achieves the same utility with either the response v_i or w_i.

Suppose that

$$\begin{aligned} d(w_1, \dots, w_N) &= 1, \\ d(w_1, \dots, w_{i-1}, v_i, w_{i+1}, \dots, w_N) &= 0. \end{aligned} \tag{3.26}$$

Then,

$$\begin{aligned} x_i(w_1, \dots, w_N) &= \textstyle\sum_{j \neq i} w_j + h_i(w_{-i}), \\ x_i(w_1, \dots, w_{i-1}, v_i, w_{i+1}, \dots, w_N) &= h_i(w_{-i}). \end{aligned} \tag{3.27}$$

The difference in utilities is therefore

$$v_i + \sum_{j \neq i} w_j. \tag{3.28}$$

But by (3.26), and the definition of $d(\cdot)$ for Groves mechanisms, (3.28) is necessarily negative. The case in which

$$\begin{aligned} &d(w_1, \ldots, w_N) = 0, \\ &d(w_1, \ldots, w_{i-1}, v_i, w_{i+1}, \ldots, w_N) = 1, \end{aligned} \tag{3.29}$$

is analogous.

Thus, using any strategy $w_i \neq v_i$ can only be harmful, compared to $w_i = v_i$; Groves mechanisms induce truthful dominant strategies. The optimality of the decisions taken follows directly from the definition of the decision function, d. Hence these mechanisms are satisfactory. Q.E.D.

One member of this class was discovered by Clarke [1971]. It corresponds to the particular choice of the arbitrary functions h_i given by

$$h_i(w_{-i}) = \min\left(-\sum_{j \neq i} w_j, 0\right) + \bar{x}_i, \qquad i = 1, \ldots, N. \tag{3.30}$$

Using this mechanism, the transfer to individual i is

$$\begin{aligned} t_i(w_{-i}) &= \sum_{j \neq i} w_j, \quad &&\text{if } |w_i| \geqq \Big|\sum_{j \neq i} w_j\Big| \\ & &&\text{and } w_i > 0 > \sum_{j \neq i} w_j, \\ &= -\sum_{j \neq i} w_j, \quad &&\text{if } |w_i| \geqq \Big|\sum_{j \neq i} w_j\Big| \\ & &&\text{and } w_i < 0 < \sum_{j \neq i} w_j, \\ &= 0 &&\text{otherwise.} \end{aligned} \tag{3.31}$$

Note that t_i is never positive. It represents the tax levied on individual i if his strategy, w_i, causes the sign of $\sum_i w_i$ to differ from the sign of $\sum_{j \neq i} w_j$. In such cases the individual in question is forcing his will on the other participants, as far as acceptance or rejection of the project is concerned. Without him they would have chosen differently.[9] According to the basic principle of resource allocation in the presence of externalities, such an individual must pay enough to compensate the others for the harm he causes. Of course in this case they do not actually receive this compensation, the planner does. We call an individual who is in this position *pivotal* and the mechanism defined by (3.31) is called the *pivotal* or *Clarke* mechanism.

The second-price auction with N bidders is a particular case of the pivotal mechanism. Here, the winner of the auction is always pivotal because he gets the commodity instead of the one whose bid came in second. All the individuals except the winner, taken collectively, experience a total utility loss equal to that of the second place bidder, since the others are completely unaffected. This explains why the second-price auction is the natural private goods counterpart of the mechanism as defined by Clarke.

[9] This statement is strictly true for a costless project, or one such that v_i can be specified for each agent independently of the size of the system in which he is embedded. A costly project, with costs shared equally, would be associated with different values of v_i for those remaining when one of the agents has been deleted.

3.10. A first characterization theorem

We have seen in the last section that there exists a class of mechanisms inducing dominant strategies and achieving Pareto optimal outcomes. One member of this class, it has been argued, seems intuitively very reasonable, but it has the drawback of generating a surplus which must be disposed of, if it is to keep its desirable properties. It is natural to ask, therefore, if there might be other satisfactory mechanisms, perhaps more attractive from this and other points of view.

We will now show that this is not true for the present situation with the real line for the strategy space and two possible public decisions. An extension of this characterization theorem to many decisions and arbitrary strategy spaces will be given in chapter 4.

Let us first observe that Groves mechanisms are characterized by the following four properties:

(1) $d(w) = 1$ if and only if $\sum_i w_i \geqq 0$,
(2) $x_i(w)$ is constant with respect to w_i on $\{w \mid d(w) = 1\}$ for all i,
(3) $x_i(w)$ is constant with respect to w_i on $\{w \mid d(w) = 0\}$ for all i,
(4) if $d(w) = 1$, $d(w') = 0$, and $w_j = w'_j$ for all $j \neq i$, then $x_i(w) - x_i(w') = \sum_{j \neq i} w_j$.

We will argue further that any mechanism that does not satisfy (1)–(4) is either not satisfactory or does not induce truthful strategies.

Theorem 3.2. All satisfactory mechanisms inducing truthful strategies satisfy (1)–(4) above and thus the set of all such mechanisms coincides with the set of Groves mechanisms.

Proof. Assume that (1) holds and that (2) fails. Then there exists an individual i, and statements $w_i, w_i', w_j, j \neq i$ such that

$$\sum_{j\neq i} w_j + w_i \geqq 0, \qquad (3.32)$$
$$\sum_{j\neq i} w_j + w_i' \geqq 0,$$
$$x_i(w_{-i}, w_i) > x_i(w_{-i}, w_i').$$

Letting $v_i = w_i'$ we see that w_i' could not be a dominant strategy for this agent. Similarly, if (3) fails, then some agent may not have a truthful dominant strategy in an environment where the optimal action is to reject the project. Suppose (4) fails but that (1), (2) and (3) hold. We will then be in one of the two following cases:

Case I. There exists i, w_{-i}, w_i, w_i' such that

$$x_i(w_{-i}, w_i) - x_i(w_{-i}, w_i') = \sum_{j\neq i} w_j + \varepsilon,$$
$$\sum_{j\neq i} w_j + w_i \geqq 0,$$

and

$$\sum_{j\neq i} w_j + w_i' < 0, \qquad (3.33)$$

for $\varepsilon > 0$.

Case II. There exists i, w_{-i}, w_i, w_i' such that

$$x_i(w_{-i}, w_i) - x_i(w_{-i}, w_i') = \sum_{j\neq i} w_j - \varepsilon, \qquad (3.34)$$
$$\sum_{j\neq i} w_j + w_i \geqq 0,$$

and

$$\sum_{j\neq i} w_j + w_i' < 0,$$

for $\varepsilon > 0$.

Let us consider case I first. Since (2) holds, $x_i(w_{-i}, w_i)$ is invariant with respect to changes in w_i, say to $\tilde{w}_i$, as long as $\sum_{j\neq i} w_j + \tilde{w}_i$ remains positive.

Letting

$$\tilde{w}_i = -\sum_{j\neq i} w_j + \frac{\varepsilon}{2}, \qquad (3.35)$$

we have:

$$x_i(w_{-i}, w_i) = x_i(w_{-i}, \tilde{w}_i), \qquad (3.36)$$
$$x_i(w_{-i}, w_i) + \tilde{w}_i = x_i(w_{-i}, w_i') + \frac{\varepsilon}{2}, \qquad (3.37)$$
$$x_i(w_{-i}, w_i) + \tilde{w}_i > x_i(w_{-i}, w_i'). \qquad (3.38)$$

Letting $v_i = w_i$ we see that (3.38) implies that the truth would not be told. Similarly, in case II we can use

$$\tilde{w}_i = -\sum_{j\neq i} w_j + \frac{\varepsilon}{2} \qquad (3.39)$$

to get the same conclusion by choosing $v_i = \tilde{w}_i$.

Finally if (1) does not hold, then either the mechanism does not induce the truth, or it will make a non-optimal decision in some cases. Q.E.D.

Chapter 4

SATISFACTORY MECHANISMS

4.1. Description of social states and allowable utility functions

We have seen that when economic environments are restricted by the assumptions of separability and transferability, and when an outside agency is created to collect any surplus revenues, the negative conclusions of part I may be somewhat reversed. In this chapter we extend these results to a generalized model in which the social decision process is to select one project from among a larger set of alternatives. The basic principles will remain the same, and the generalization will serve to highlight the main ideas. We will characterize the class of dominant strategy mechanisms that always select Pareto optima, and then discuss several applications of these procedures – including some that would be desirable, but which are in fact not feasible for various reasons.

As above, there is a unique transferable private good, which may be thought of as money. The set of possible public decisions is denoted $\mathscr{K}$. We assume that $\mathscr{K}$ can be described as a compact subset of a topological space. This fits many possible contexts. Examples include:

(i) a fixed size unique public project; then $\mathscr{K} = \{0, 1\}$ where 1 represents the realization of the project and 0 the non-realization,
(ii) a variable size unique public project; then the allowable size should belong to a bounded closed subset of $\mathbf{R}$; for example, $\mathscr{K} = 0 \cup [K_1, K_2]$ if a minimal size is required technologically or $\mathscr{K} = \{0, K_1, \ldots, K_L\}$, if indivisibilities exist,
(iii) a set of L variable size public projects; then, for example,
$\mathscr{K} = \prod_{l=1}^{L} [0, K_l]$,
(iv) a compact set of functions (e.g., tax laws); then, for example,

$\mathscr{K} \subseteq \mathscr{C}[0, 1]$, set of continuous real functions on the interval [0, 1].

Let K be a typical element of $\mathscr{K}$.

The set of social states is $\mathscr{K} \times \mathbf{R}^N$, where N denotes the number of individuals in the population. Thus a typical social state is denoted $(K, x_1, \ldots, x_N) \in \mathscr{K} \times \mathbf{R}^N$. Individual i's preferences concerning alternative social states depend only on K and x_i. The additive separability assumption of chapter 3 can easily be expressed through the following definition in the multi-project case.

Definition 4.1. Preferences are *separable* if they are representable by functions u_i: $\mathscr{K} \times \mathbf{R} \to \mathbf{R}$, for each $i = 1, \ldots, N$, of the form

$$u_i(K, x_i) = v_i(K) + x_i$$

We assume that the individuals have initial endowments $\bar{x}_i$ and that their consumptions, x_i, represent the sum of the endowment and a transfer t_i made as part of the resource allocation process. Thus an ordinally equivalent representation of preferences is

$$u_i(K, t_i) = v_i(K) + t_i \tag{4.1}$$

where the $v_i(\cdot)$ are the same functions of K alone as those described above. For the most part, this chapter will be concerned with situations in which the level of endowment plays no role. Therefore, preferences will be expressed by (4.1).

As in the case of chapter 3, the separability assumption implies the absence of income effects concerning the willingness-to-pay for alternative public decisions. This is a necessary condition[1] for the existence of mechanisms with the desirable characteristics we seek. In cases in which $\mathscr{K}$ is a multi-dimensional space describing the levels of several types of public goods, the willingness-to-pay is jointly determined by their entire specification. We do not ascertain willingnesses-to-pay for each of the components separately, nor are these even well-defined. If we would have additive separability of the v function itself, then the separate willingnesses-to-pay for each component of the project space could be elicited by using the mechanism repeatedly for each part of the decision.

One way in which the public decision mechanisms we describe are more general than those sometimes encountered in planning processes is that no restriction is placed on the concavity of $v_i(\cdot)$, nor need $\mathscr{K}$ have any

[1] See section 5.1 for an elaboration of this assertion.

linear structure at all. This is due to the global, once-and-for-all nature of the method by which a Pareto optimum is attained. Requiring that the entire function $v_i(\cdot)$ be communicated, however, has its costs in terms of information transmission and processing.[2]

4.2. Mechanisms

We now discuss how processes can be established for discovering and successfully aggregating these preferences.

Definition 4.2. A *mechanism*, $M = (S, f)$, is a set of strategy spaces S_i, $i = 1, \ldots, N$, and a function $f(\cdot) = [d(\cdot), t_1(\cdot), \ldots, t_N(\cdot)]$ from $\prod_i S_i = S$ into $\mathscr{K} \times \mathbf{R}^N$ such that for each $s \in S$.

(1) the accepted project is $d(s)$,

(2) the transfer to agent i is $t_i(s)$ for $i = 1, \ldots, N$.

The components of $f(\cdot)$ are referred to as the decision function and the transfer functions, respectively.

With this rather general specification, many social decision procedures are allowed. For example, the individuals may first be asked to reveal their preferences for some public goods and then, depending on the outcome attained, they may be asked a variety of different questions concerning the other public goods. Another possibility is that they are divided into groups and in the first stage they elect a representative from each group on the basis of the preferences of his constituency. Then, in a later stage of the decision process, the representatives choose strategies which jointly determine the social outcome. As long as the strategy space is specified with sufficient complexity to account for the action of the individual, conditional on every sequence of events that he might experience, the mechanism will be well-defined.

There are two important restrictions embodied in this description of the mechanism. The first is that a unique non-random outcome be defined for each N-tuple of strategies. This rules out situations in which the strategy spaces, S_i, are randomized mixtures from some more primitive set of basic strategies, for then the outcome would generally vary with the results of

[2] In part IV we study a dynamic process which, though less demanding in its information structure at each instant, requires an infinite number of iterations to compute an optimum precisely. Nevertheless the total quantity of information elicited during the dynamic process will be seen to be strictly less than that discovered in the global processes presently considered.

these individualized randomizations. It will be seen in what follows that this restriction is not a critical one, because we will define mechanisms that have dominant strategies for each individual. Randomization by individuals will not be of value.

For a variety of reasons, however, another type of randomized social choice procedure may be useful. A subset of the individuals may be selected randomly from the population and then a mechanism – applied to a much lower dimensional joint-strategy space – is used to select the social state. Such mechanisms are implicitly ruled out in the present context because we assume that $f(\cdot)$ is properly a function. We return to them in part III, chapters 12 and 13.

Due to the potential complexity of the strategies and of the decision and transfer functions allowable in a general mechanism as described, it may be valuable to consider some simpler, more restrictive forms of social choice devices. Since Pareto optimality in systems with separable preferences requires that K be chosen such that $\sum_i v_i(K)$ be maximized over $\mathscr{K}$, and places no restrictions at all on the transfers, we might therefore want to find mechanisms whose strategies could be interpreted as "professed preferences" of the agents.

Definition 4.3. A *revelation mechanism*, $\boldsymbol{RM} = (V, f)$, is a mechanism such that

$$\begin{aligned} V_i &= \{w_i \mid w_i\colon \mathscr{K} \to \mathbf{R}\}, \\ V &= V_1 \times \ldots \times V_N, \\ S_i &= V_i, \qquad i = 1, \ldots, N. \end{aligned}$$

In a revelation mechanism, each agent is asked the question: "What is your valuation function?", and the social decision is determined directly by the answers. Of course, it is only relevant to use a revelation mechanism when preferences are known to be separable, provided that Pareto optimality is the single goal. The professed preferences, $w_i(\cdot)$, may of course differ from the true ones, $v_i(\cdot)$, which are known only to the agent himself.

When a revelation mechanism is used, it is natural to think of the decision function which chooses the apparently optimal project. This is not required of a general revelation mechanism. The government may understand that individuals will be systematically distorting their answers and will attempt to correct this before computing the optimum.[3]

[3] In chapter 7 we will study a somewhat unusual mechanism having the properties that individuals distort their answers but the government, acting strictly according to the announced preferences, nevertheless attains an optimal outcome.

If the mechanism does accept the professed preferences as valid, we are led to the following definition.

Definition 4.4 A *direct revelation mechanism*, $\boldsymbol{DRM} = (V, f)$, is a revelation mechanism for which

$$d(w_1(\cdot), \ldots, w_N(\cdot)) \text{ maximizes } \sum_i w_i(\cdot) \text{ over } \mathscr{K}.$$

There may be several maximizers of $\sum_i w_i(\cdot)$ over $\mathscr{K}$ in any particular situation. Any selection from this set will define a ***DRM***.

More generally, we can allow for set valued mechanisms, or mechanisms in which non-unique outcomes result.

Definition 4.5 An *extended direct revelation mechanism*, $\boldsymbol{EDRM} = (V, \boldsymbol{F})$, is such that

$$V = V_1 \times \ldots \times V_N,$$

that is, the joint strategy space is the product of the spaces of possible valuation functions, and

$$\boldsymbol{F}\colon V \to \mathscr{K} \times \mathbf{R}^N \qquad F(w) = (D(w), T_1(w), \ldots, T_N(w))$$

is a correspondence such that for every $w \in V$, $D(w)$ is a subset of the maximizers of $\sum_i w_i(\cdot)$.

Before ***DRM***s or ***EDRM***s can be applied, it is necessary to insure that maxima will always exist. Otherwise, it would be impossible to define the decision function. The simplest way to do this is to restrict V_i to the set of upper semi-continuous functions over $\mathscr{K}$. The sum of such functions will always have a maximum. This means that even though it may be known by the planner that the true valuation functions are upper semi-continuous, it is still necessary to make this restriction to avoid potential distortions that do not have this property, and which therefore might disrupt the decision-making mechanism.

This is a rather mild restriction on V_i since it means only that the preferences generating this utility have closed upper contour sets.[4] Under this condition, each single agent could in fact select his most preferred point from any compact subset of $\mathscr{K}$.

It is important to note that no further restrictions such as the convexity of the underlying preference relation are assumed. If these hypotheses

[4] The property is precisely the one used to guarantee the non-emptiness of demand correspondences in Walrasian equilibrium theory.

were made, there is a potential for the existence of more mechanisms with desirable properties, much as domain restrictions in the social choice context may permit social welfare functions to perform better. We explore this possibility in section 4.4, with negative results.

In order to set up the correct incentives for the participants in the system, the transfer functions must be constructed in an appropriate way. The following class of mechanisms was proposed by Groves [1973]:

Definition 4.6 A *Groves mechanism*, $\boldsymbol{GM} = (V, f)$, is a direct revelation mechanism in which the transfer functions satisfy

$$t_i(w(\cdot)) = \sum_{j \neq i} w_j(K^*(w(\cdot))) + h_i(w_{-i}(\cdot))$$

where $K^*(w(\cdot))$ is a maximizer of $\sum_i w_i(\cdot)$ and the functions $h_i(\cdot)$ are arbitrary deterministic real-valued functions depending on all the announced valuation functions except that of the individual himself.

The important feature of Groves mechanisms is that the only way in which an individual's statement affects the transfer he receives (or pays) is through the shift it causes in the apparently optimal social decision. More precisely, when an individual changes his strategy, the change in the transfer he receives will be exactly the cost that this shift induces for the other members of society through the variation in the decision taken. The parallel with the second-price auction and the other public decision mechanisms discussed in chapter 3 is clear; this is the straightforward generalization to an arbitrary space of alternative decisions.

The pivotal mechanism in this context was first proposed by Clarke [1971]. It is defined by:

Definition 4.7 The *Clarke mechanism* (or the *pivotal mechanism*) is the Groves mechanism for which

$$h_i(w_{-i}(\cdot)) = -\sum_{j \neq i} w_j(K_i^*(w_{-i}(\cdot)))$$

where $K_i^*(w_{-i}(\cdot))$ is a maximizer of $\sum_{j \neq i} w_j(\cdot)$ over $\mathscr{K}$.

Just as in chapter 3, we ask how the stated welfare of others would have changed if their collectively optimal project[5] were adopted instead, and we tax individuals accordingly.

We will also have use for the following concept which specializes ***EDRM***s to the Groves case:

[5] See footnote 9, chapter 3, for the interpretation in the case of costly projects.

Definition 4.8. An *extended Groves mechanism*, $\boldsymbol{EGM} = (V, \boldsymbol{F})$, is an extended direct revelation mechanism such that

$$D(w(\cdot)) \text{ is a subset of the maximizers of } \sum_i w_i(\cdot)$$

and

$$t_i(w(\cdot)) = \sum_{j \neq i} w_j(K^*(w(\cdot))) + h_i(w_{-i}(\cdot))$$

where

$K^*(w(\cdot)) \in D(w(\cdot))$ and $h_i(\cdot) \in H_i(\cdot)$ and $H_i(\cdot)$ is an arbitrary deterministic correspondence from $\prod_{j \neq i} V_j$ into $\mathbf{R}$.

4.3. Optimality and characterization theorems: definitions and objectives

Before discussing the properties of alternative mechanisms, it is necessary to have a theory describing how the strategies to be played by the individuals are selected. For given preferences, mechanisms are N-person, non-zero sum games. Moreover, there is no reason to presume that the agents know each other's utilities as a function of the joint strategy played. Indeed it is the heart of the matter that they do not; for if they did, and each perceived that the others did, then there would be a strong reason to presume that some form of bargaining, outside the mechanism, could lead to Pareto optimality. The incompleteness of information and the potential for disguising it from colleagues is pervasive and is the raison d'être for the mechanism itself. We are in the world of N-person, non-zero sum game theory with incomplete information.

In complete generality there would be little predictive content in any solution concept for such games, with or without complete information. Let us suppose, therefore, that when faced with any mechanism, the individual uses his subjectively uncertain beliefs about others' preferences and strategies to be played conditional on holding particular preferences to derive a subjective joint probability distribution for the joint strategy of all the others. No attempt will be made to relate these beliefs to any particular game theoretic concept of equilibrium in the case of a general mechanism. However, for mechanisms of particular interest such a procedure will form a crucial vehicle for our analysis.

Let $\xi_i(v_i, \boldsymbol{M})$ be the set of strategies that are optimal in mechanism $\boldsymbol{M}$

as calculated by individual i using his own beliefs, when his preferences are $v_i = v_i(\cdot)$.[6]

Due to the lack of specificity of the ξ_i correspondences, we should be particularly pleased if they were independent of the individual's beliefs.

Definition 4.9. A strategy $s_i \in S_i$ is said to be a *dominant strategy for agent i* if

$$v_i(d(s_i, s_{-i})) + t_i(s_i, s_{-i}) \geqq v_i(d(s_i', s_{-i})) + t_i(s_i', s_{-i})$$

for all $s_i' \in S_i$ and all $s_{-i} \in \prod_{j \neq i} S_j$.

The set of all dominant strategies for i is denoted by $\mathscr{D}_i(v_i, \boldsymbol{M})$. It depends in general on the individual's tastes and on the mechanism being used. Further, it might vary with the individuals themselves, both because the strategy spaces for different agents could be different and because the mechanism might treat them asymmetrically. Throughout this book we will deal primarily with mechanisms that are symmetric in the sense that the strategy spaces are the same for all agents and the decision and transfer functions are symmetric functions of the strategies played. In this case, $\mathscr{D}_i(v_i, \boldsymbol{M})$ could be written as $\mathscr{D}(v_i, \boldsymbol{M})$.

The generality of the description of mechanisms used thus far admits sequential decision making procedures and other potentially useful devices such as representative elections within subgroups of agents to form a legislative body). We do not want to rule out such possibilities ex ante. However, the force of the present section is to demonstrate that they are unnecessary and that the class of Groves mechanisms defined above exhausts all of those that can guarantee Pareto optimality. To do so we retain the more general description.

Any particular mechanism may yield dominant strategies for some agents in some situations but not others. If dominant strategies always exist, then the problems of non-cooperative game theory under any informational specification would never arise.

Definition 4.10. A mechanism $\boldsymbol{M}$ is said to be *decisive* if

$$\mathscr{D}_i(v_i, \boldsymbol{M}) \neq \phi$$

for all i and all $v_i: \mathscr{K} \to \mathbf{R}$.

Decisive mechanisms do not, however, rule out the possibility of cooperative behavior that is mutually advantageous for the agents involved.

[6] From this point onward we write v_i for $v_i(\cdot)$ and w_i for $w_i(\cdot)$ when it is understood that these are defined to be real-valued preference functions over $\mathscr{K}$.

Dominant strategies are best against any particular, fixed, strategies played by the others. But if by choosing another strategy the choices of others can be influenced, then it may be wise to do so. Whether cooperative behavior is to be regarded as a real problem depends to a large extent on considerations that are outside of the description of the mechanism itself: Will the strategy choices be revealed after the play is completed? Will the transfers made to each player be revealed to the others? Are there any methods for enforcement or checking up on agreements made concerning the joint strategy choice of a coalition? Will the mechanism be repeated later, perhaps concerning a different public project, as yet with unknown characteristics? Is tacit collusion a possibility? These issues, and the cooperative behavior that results will be treated in chapter 10. For the present, we will concentrate on the purely non-cooperative case which may be justified by assuming that there is no information available about others' strategies ex post and that no repetitions of the mechanism are envisioned by the agents in the future.[7]

Under these conditions we can be sure that the individual will play one of his dominant strategies. His tastes are unknown, but we can narrow down the strategies that might conceivably be played to those that would be dominant for some value of his true preferences.

Let

$$S_i' = \bigcup_{v_i \in V_i} \mathscr{D}_i(v_i, \boldsymbol{M}) \tag{4.2}$$

and

$$S' = \prod_i S_i'. \tag{4.3}$$

Strategy spaces for decisive mechanisms can be restricted to $\{S_i'\}$ without loss of generality.

Whether or not the mechanism is decisive, we are interested in the suitability of the outcomes it produces. For the type of preferences we consider, it is simple to characterize the Pareto optimal points. They are all those values of K such that $\sum_i v_i(K)$ is maximized.

Definition 4.11. A mechanism $\boldsymbol{M}$ is said to be *successful* if whenever $s_i \in \xi_i(v_i; \boldsymbol{M})$, for all i, $d(s)$ maximizes $\sum_i v_i(K)$ over $\mathscr{K}$.

[7] With repetions of the same mechanism for preferences that are not unrelated in different iterations, the possibility of a myopically non-optimal play which enables the individual to learn indirectly about the strategic choices of the other members of his coalition, could not be ruled out.

Successfulness captures the idea of achieving Pareto optimality, independently of the initial beliefs. However, the particular Pareto optimal point selected may vary with these beliefs. Further, the interactive problems of a game theoretic nature, such as threats and promises which may be fulfilled by subsequent transfers of the private good, have not been avoided. By requiring both decisiveness and successfulness we insure the attainment of an optimum while guaranteeing that no individual can affect anyone else's response by deviating from his optimal strategy.

Definition 4.12. A mechanism that is decisive and successful is called *satisfactory*.

The design and analysis of satisfactory mechanisms is the central focus of this part of the book.

Since some of the mechanisms to be treated below are revelation mechanisms, we can introduce a truth-telling concept that will strengthen the notion of decisiveness.

Definition 4.13. A revelation mechanism, ***RM***, is *strongly individually incentive compatible* (*SIIC*) if the truth is a dominant strategy for each individual; that is if

$$v_i \in \mathscr{D}_i(v_i; \boldsymbol{RM}) \quad \text{for all } i \text{ and all } v_i.$$

4.4. Optimality and characterization theorems: Results

The basic interest in the mechanisms considered in this book can be traced to the fact that Groves mechanisms are *SIIC* and successful revelation mechanisms. This accounts for the nice properties of second-price auctions as noted in chapter 3.

Theorem 4.1. (Groves and Loeb [1975]). A Groves mechanism is a *SIIC* direct revelation mechanism.

Proof. For any $w_{-i} \in V_{-i}$ and any $w_i \in V_i$,

$$\begin{aligned} &u_i(w_{-i}, v_i; \boldsymbol{GM}) - u_i(w_{-i}, w_i; \boldsymbol{GM}) \\ &\quad = v_i(K^*(w_{-i}, v_i)) + \sum_{j \neq i} w_j(K^*(w_{-i}, v_i)) \\ &\qquad + h_i(w_{-i}) - v_i(K^*(w_{-i}, w_i)) \\ &\qquad - \sum_{j \neq i} w_j(K^*(w_{-i}, w_i)) - h_i(w_{-i}) \end{aligned} \tag{4.4}$$

$$\begin{aligned} &= \max_{K \in \mathscr{K}} [v_i(K) + \sum_{j \neq i} w_j(K)] \\ &\quad - [v_i(K^*(w_{-i}, w_i)) + \sum_{j \neq i} w_j(K^*(w_{-i}, w_i))] \\ &\geqq 0 \end{aligned}$$

Q.E.D.

It is evident from the proof above, and the definition of Groves mechanisms, that the actual set of dominant strategies for an individual whose true preference is v_i includes all the functions which differ from this by a constant throughout $\mathscr{K}$.

$$w_i(K) = v_i(K) + \alpha. \tag{4.5}$$

This is because w_i influences u_i only through the choice of the project. Shifting functions by a constant does not affect the location of their maxima; hence α is really irrelevant.

More precisely, for any true preference relation on the underlying set $\mathscr{K} \times \mathbf{R}$, the individual's true preferences are equally well represented by all such functions. Our choice of a particular representation v_i is therefore arbitrary, and all members of this class are equivalent to it. If the strategies were the preference relations themselves, rather than their representations, there would be a unique dominant strategy, and it would be the truth.

Since it is easier to work with valuation functions than preference relations, we shall actually be concerned with the equivalence classes of valuation functions that represent the same underlying ordering on $\mathscr{K} \times \mathbf{R}$. To describe an equivalence class of announced valuations, it is simplest to choose a single project, $\bar{K}$, which we will call the "null project", the "no-action" project or the "status quo", according to the context, and require that

$$v_i(\bar{K}) = w_i(\bar{K}) = 0 \quad \text{for all } i. \tag{4.6}$$

Thus, in responding to a revelation mechanism, the announced valuation of the project $\bar{K}$ may be fixed at zero; all other elements of $\mathscr{K}$ are ranked relative to this one.

Definition 4.14. A strategy w_i in a revelation mechanism is said to be *normalized* if $w_i(\bar{K}) = 0$.

We have in mind that $\bar{K}$ represents no change from an existing situation.[8]

Theorem 4.2. (Groves [1974]). For any Groves mechanism there is a unique normalized dominant strategy for each individual.

[8] This follows the convention adopted in chapter 3, where the "status quo" decision was written $d = 0$, and v_i measured the willingness-to-pay for the adoption of $d = 1$.

Proof. Suppose that there exists a dominant strategy, say v_i', such that $v_i' - v_i$ is not constant. Then, there exist $\varepsilon > 0$, $\alpha > 0$, $K^* \in \mathscr{K}$, $K^{**} \in \mathscr{K}$, such that:

$$\begin{aligned} v_i'(K^*) &= v_i(K^*) + \alpha \\ v_i'(K^{**}) &= v_i(K^{**}) + \alpha + \varepsilon \end{aligned} \tag{4.7}$$

Choose w_{-i} to be the upper-semicontinuous function defined by:

$$\begin{aligned} \sum_{j \neq i} w_j(K^*) &= -v_i(K^*) - \alpha \\ \sum_{j \neq i} w_j(K^{**}) &= -v_i(K^{**}) - \alpha - \frac{\varepsilon}{2} \\ \sum_{j \neq i} w_j(K) &= -\sup\left[\sup_{K \in \mathscr{K}} v_i(K), \sup_{K \in \mathscr{K}} v_i'(K)\right] - \alpha - \varepsilon \\ &\qquad \text{for } K \in \mathscr{K},\ K \neq K^*,\ K \neq K^{**} \end{aligned} \tag{4.8}$$

Clearly, the answer v_i leads to the project K^* and the answer v_i' leads to the project K^{**}. Moreover, we have:

$$v_i(K^*) + \sum_{j \neq i} w_j(K^*) > v_i(K^{**}) + \sum_{j \neq i} w_j(K^{**}) \tag{4.9}$$

therefore, v_i' is not a dominant strategy, a contradiction. All the dominant strategies must then differ from v_i by a constant. From theorem 4.1, they are indeed dominant strategies.

Finally, a normalized strategy for agent i is such that $w_i(\bar{K}) = 0$. Clearly, displacing w_i by any non-zero constant would destroy this normalization. The normalized dominant strategy is therefore unique. Q.E.D.

Corollary 4.1. A Groves mechanism is successful.

Proof. From theorem 4.2 and the definition of a Groves mechanism, we know that the decision taken maximizes the sum of the valuation functions when any dominant strategy is played. Q.E.D.

Corollary 4.2. The set of dominant strategies for each agent, i, in a successful, *SIIC* revelation mechanism is included in the set $\{w_i \mid w_i = v_i + \alpha, \alpha \in \mathbf{R}\}$. There is a unique normalized dominant strategy corresponding to $\alpha = 0$.

Proof. As in theorem 4.2 we show that a dominant strategy for agent i must be of the form $w_i = v_i + \alpha$, $\alpha \in \mathbf{R}$. The strategy $w_i = v_i$ is a dominant strategy since the mechanism is *SIIC*. As in theorem 4.2 it is also the only normalized dominant strategy. Q.E.D.

Having demonstrated that Groves mechanisms are *SIIC* and successful, we now turn to the converse question. Namely, are there any other *SIIC* and successful direct revelation mechanisms? This is the heart of this chapter, and in many ways it is the negative answer to this question that provides the impetus for many of the subsequent ideas we explored. By limiting the *SIIC* and successful mechanisms to this easily characterized class of functions, we can readily check whether some of the other properties of mechanisms that one might feel are desirable can also be satisfied by Groves mechanisms. If not, as is unfortunately often the case, we are immediately thrust into the possibility of trading off the properties of *SIIC* and successfulness against the attainment of these others. The effects on these tradeoffs of having large numbers of individuals in the system is the central idea of part III.

To demonstrate this converse result we utilize a property of direct revelation mechanisms which, for lack of a more memorable term, we call *transfer independence and compensation.* This, it will be shown, is equivalent to both *SIIC* and successfulness and the fact that the mechanism is a Groves mechanism.

Definition 4.15. A direct revelation mechanism, $\boldsymbol{DRM} = (V, f)$, is said to satisfy the property of *transfer independence and compensation* if

(i) $t_i(w)$ is independent of w_i at K^*, i.e., if for w_{-i}, w_i, w_i' such that

$$K^*(w_{-i}, w_i) = K^*(w_{-i}, w_i')$$

then

$$t_i(w_{-i}, w_i) = t_i(w_{-i}, w_i')$$

(ii) $t_i(w_{-i}, w_i) - t_i(w_{-i}, w_i') = \sum_{j \neq i} w_j(K^*) - \sum_{j \neq i} w_j(K^{*\prime})$ where K^* maximizes $\sum_{j \neq i} w_j(K) + w_i(K)$ over $\mathscr{K}$ and $K^{*\prime}$ maximizes $\sum_{j \neq i} w_j(K) + w_i'(K)$ over $\mathscr{K}$.

The first part of this definition states that the transfer depends on the individual's own statement only through its influence on the project that is chosen and not on the statement directly. The second part embodies the essential idea of the compensation principle regarding externalities: If an individual, by changing his response, alters the project adopted, the cost of making this switch to him, in terms of the private good, is exactly the cost that it imposes on the other members of society. Of course the true cost is not known – only stated costs are available. But if true responses are given, then part ii) of this property will cause private and social costs to coincide.

Lemma 4.1. A direct revelation mechanism is a Groves mechanism if and only if

(i) $d(w)$ maximizes $\sum_i w_i$ over $\mathscr{K}$

and

(ii) t_i satisfies the property of transfer independence and compensation for every i.

Proof. This is immediate on comparing the definition of a Groves mechanism with the stated properties.

We are now able to prove the main characterization theorem.

Theorem 4.3. An *SIIC* direct revelation mechanism is a Groves mechanism.

Proof. We consider in turn the negation of the two parts of the transfer independence and compensation property. If (i) fails there exist w_{-i}, w_i, w_i' which lead to the same K^* such that

$$t_i(w_{-i}, w_i) > t_i(w_{-i}, w_i') \tag{4.10}$$

Let $v_i = w_i'$.

Then,

$$t_i(w_{-i}, w_i) + v_i(K^*) > t_i(w_{-i}, v_i) + v_i(K^*), \tag{4.11}$$

and v_i is not a dominant strategy, a contradiction.

If (ii) fails, there exist w_{-i}, w_i, w_i' such that:

K^* maximizes $\sum_{j \neq i} w_j + w_i$ over $\mathscr{K}$

$K^{*\prime}$ maximizes $\sum_{j \neq i} w_j + w_i'$ over $\mathscr{K}$

and

$$t_i(w_{-i}, w_i) - t_i(w_{-i}, w_i') = \sum_{j \neq i} w_j(K^*) - \sum_{j \neq i} w_j(K^{*\prime}) + \varepsilon \quad \text{for some } \varepsilon > 0. \tag{4.12}$$

Let $\tilde{w}_i'$ be defined as

$$\begin{aligned} \tilde{w}_i'(K^*) &= -\sum_{j \neq i} w_j(K^*) \\ \tilde{w}_i'(K^{*\prime}) &= -\sum_{j \neq i} w_j(K^{*\prime}) + \varepsilon/2 \\ \tilde{w}_i'(K) &= -c \quad \text{for } K \neq K^* \text{ or } K \neq K^{*\prime} \text{ with } c > \max_{K \in \mathscr{K}} \sum_{j \neq i} w_j(K) \end{aligned} \tag{4.13}$$

$\tilde{w}_i'$ is upper semi-continuous.

Note that $\max \tilde{w}_i'(K) + \sum_{j \neq i} w_j(K)$ is solved at $K = K^{*\prime}$, and therefore by the first part of the proof

$$t_i(w_{-i}, w_i') = t_i(w_{-i}, \tilde{w}_i') \tag{4.14}$$

We have that

$$\begin{aligned} t_i(w_{-i}, w_i) - t_i(w_{-i}, \tilde{w}_i') &= \sum_{j \neq i} w_j(K^*) - \sum_{j \neq i} w_j(K^{*\prime}) + \varepsilon \\ &= -\tilde{w}_i'(K^*) + \tilde{w}_i'(K^{*\prime}) + \varepsilon/2 \end{aligned} \tag{4.15}$$

Therefore

$$t_i(w_{-i}, w_i) + \tilde{w}_i'(K^*) > t_i(w_{-i}, \tilde{w}_i') + \tilde{w}_i'(K^{*\prime}) \tag{4.16}$$

Hence, when $v_i \equiv \tilde{w}_i'$, the announcement of w_i will be superior which contradicts the fact that the mechanism is *SIIC*. Q.E.D.

Combining theorems 4.1 and 4.3 we establish the equivalence of *SIIC* direct revelation mechanisms and Groves mechanisms. Since Groves mechanisms are also successful, this establishes the fact that all successful *SIIC* direct revelation mechanisms are of this type.

The characterization above is incomplete for two reasons. We might inquire whether the set of satisfactory mechanisms can be enlarged either by placing further restrictions on the valuation functions or by allowing more complexity in the strategy space and decision making process than direct revelation makes possible. It will be shown that neither of these possibilities leads to satisfactory mechanisms that differ from Groves mechanisms in any essential respect. Before proving the relevant characterization theorems, we digress for a moment to argue that, in fact, there are good reasons to reject restrictions on the valuation functions beyond upper semi-continuity, even if these would expand the potential range of satisfactory mechanisms.

Although we have maintained separability as an hypothesis throughout, we must think of the private good as a composite commodity consisting of a weighted average of purchases of various private goods. The weights are, of course, the equilibrium prices of these goods. Thus, if the individual perceives a connection between the social decision taken, the transfers made and the equilibrium prices he will face for private goods, the valuation function should reflect the change in the indirect utility function for private goods that he associates with the prices induced by the social decision.

Without giving a detailed theory of how the individual constructs reasonable or rational expectations of price systems based on his own observations and prior beliefs, the difficulties and implications of making such a connection can still be discussed. It is well-known that equilibrium prices in a Walrasian model are typically non-unique, and further that the behavior of this set of prices is upper semi-continuous, but *not continuous* in the parameters of the model. Usually, these parameters include preferences, endowments and perhaps share ownerships and technologies. Here, the level of public goods that is selected by the collective decision process is also a parameter of the Walrasian equilibrium that would ensue. Since the indirect utility function is continuous in prices, it is therefore only upper semi-continuous in the level of the public good to be selected. Assuming that the valuation function is continuous in the level of public goods can only be justified by the presumption that the price system is continuous in the parameters, and therefore implicitly unique.[9]

Having thus defended the necessity for treating the more general case of upper semi-continuity, it is only a slight additional comfort to note that even if one would be willing to assume uniqueness of prices (or if the project were small so that local uniqueness would be sufficient[10]) we still cannot find any new satisfactory mechanisms.

For simplicity we will now assume that $\mathscr{K}$ is a compact subset of the real line.[11]

Theorem 4.4. If announced valuation functions are restricted to the class of continuous functions, the family of Groves mechanisms is identical to the set of all *SIIC* direct revelation mechanisms.

Proof. It is only required to show that $\tilde{w}_i'$ in the proof of theorem 4.3 can be chosen continuous and so that

(i) $K^{*\prime}$ maximizes $\tilde{w}_i' + \sum_{j \neq i} w_j$ over $\mathscr{K}$

[9] This point is made precise in Balasko [1975].

[10] Local uniqueness, unlike global uniqueness can be expected to hold in general, see Debreu [1970].

[11] Theorems 4.4, 4.5 and 4.6 can be generalized to the case of normal linear spaces of decisions, at the cost of notational complexity only.

(ii) $\tilde{w}_i'(K^*) = -\sum_{j \neq i} w_j(K^*)$

(iii) $\tilde{w}_i'(K^{*\prime}) = -\sum_{j \neq i} w_j(K^{*\prime}) + \frac{\varepsilon}{2}$

Let $\eta > 0$ such that $\eta < |K^* - K^{*\prime}|$ and

$$\tilde{w}_i'(K) = -\sum_{j \neq i} w_j(K) \quad \text{for } K \notin [K^{*\prime} - \eta, K^{*\prime} + \eta] \tag{4.17}$$

$$= -\sum_{j \neq i} w_j(K) + \frac{\varepsilon}{2}\left(1 - \frac{|K - K^{*\prime}|}{\eta}\right) \text{ for } K \in [K^{*\prime} - \eta, K^{*\prime} + \eta]$$

By construction $\tilde{w}_i'$ is continuous since $w_j, j \neq i$ are continuous and satisfy (ii) and (iii). Moreover, it is easy to see that (i) is also satisfied since we always have

$$\tilde{w}_i'(K) + \sum_{j \neq i} w_j(K) < \frac{\varepsilon}{2} \text{ for } K \neq K^{*\prime} \qquad \text{Q.E.D.} \tag{4.18}$$

Having shown that the restriction to continuous valuation functions does not change the basic characterization theorem, we may ask whether other, stronger, restrictions will affect our results. A natural condition to consider is some type of convexity postulate. For this, the space in which $\mathscr{K}$ is embedded must have a natural linear structure. For example, if $\mathscr{K}$ represents the levels of a variety of public projects which can be chosen continuously, then a suitable Euclidean space is relevant. If $\mathscr{K}$ represents discrete choices, or yes-no decisions, on various potential projects, a normalization or scaling of the project sizes is required before convexity can be discussed precisely.

Let us consider the project space[12] $\mathscr{K} = \mathbf{R}$ and let $V°$ be the family of valuation functions

$$v_i(K, \theta) = \theta_i K - \frac{K^2}{2} \quad i = 1, \ldots, N \tag{4.19}$$

where θ_i belongs to an open set Θ. Even if one restricts the class of preferences to the family $V°$, one does not obtain additional *SIIC* direct revelation mechanisms.

Theorem 4.5. If $\mathscr{K} = \mathbf{R}$ and announced valuation functions are restricted

[12] Although this project space is not compact, for the class of preferences to be explored below, a maximum of $\Sigma_i v_i$ will always exist. The only reason that compactness of $\mathscr{K}$ was assumed was to assure this existence. Therefore the present formalization is not different in any essential respect – see also, sections 5.3 and 5.5, theorem 5.12 and example 5.7.

o belong to $V°$, the family of Groves mechanisms is identical with the lass of all *SIIC* direct revelation mechanisms.

roof. We show that any *SIIC* ***DRM*** is a Groves mechanism. A ***DRM*** s such that the decision is taken by maximizing:

$$\sum_i \left(\theta_i K - \frac{K^2}{2} \right) \tag{4.20}$$

eading to

$$K^*(\theta) = \frac{1}{N} \sum_i \theta_i \tag{4.21}$$

et $\tilde{\theta}_i$ be the true preference parameter of agent i, $i = 1, \ldots, N$.

The transfer functions can be thought of as depending on θ directly, ince θ_i determines v_i. We denote them by $t_i(\theta)$, $i = 1, \ldots, N$. They must e such that the truth, $\tilde{\theta}_i$, is a dominant strategy for all possible values of his parameter. Thus $\tilde{\theta}_i$ maximizes

$$v_i(K^*(\theta), \tilde{\theta}_i) + t_i(\theta) \tag{4.22}$$

ver all possible $\theta_i \in \Theta$, for each value of θ_{-i}. This will hold only if

$$\frac{\partial v_i}{\partial K}[K^*(\tilde{\theta}_i, \theta_{-i}), \tilde{\theta}_i] \cdot \frac{\partial K^*}{\partial \theta_i}(\tilde{\theta}_i, \theta_{-i}) + \frac{\partial t_i}{\partial \theta_i}(\tilde{\theta}_i, \theta_{-i}) = 0 \tag{4.23}$$

r any θ_{-i}.

Since we require (4.23) for any $\tilde{\theta}_i$, it can be viewed as an identity. Integrating (4.23) with respect to θ_i yields then:

$$t_i(\theta_i, \theta_{-i}) = \int -\frac{\partial v_i}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_i} \mathrm{d}\theta_i + h_i(\theta_{-i}) \tag{4.24}$$

here h_i is an arbitrary function.

Using the fact that v_i is quadratic, (4.24) becomes:

$$t_i(\theta_i, \theta_{-i}) = -\frac{\theta_i^2}{2} \frac{(N-1)}{N^2} + \frac{\sum_{j \neq i} \theta_j}{N^2} \cdot \theta_i + h_i(\theta_{-i}) \tag{4.25}$$

ut

$$\begin{aligned} t_i(\theta_i, \theta_{-i}) &= \sum_{j \neq i} \left(\theta_j K^*(\theta) - \frac{K^*(\theta)^2}{2} \right) - \frac{(N+1)}{2N^2} (\sum_{j \neq i} \theta_j)^2 + h_i(\theta_{-i}) \\ &= \sum_{j \neq i} v_j(K^*(\theta), \theta_j) + h_i'(\theta_{-i}) \end{aligned} \tag{4.26}$$

appear to be the transfer functions of the Groves mechanisms. By theorem 4.1, these are *SIIC* ***DRM***s and hence the Groves mechanisms are identical with the class of all *SIIC* ***DRM***s.[13,14] Q.E.D.

The point of view taken in theorem 4.5 is interesting because it provides a constructive way of obtaining the Groves class. Furthermore, it will be useful when we ask a number of other questions, such as the characterization of successful expected utility maximizing mechanisms (see chapter 7) or the characterization of the class of coalitions which can be prevented from manipulating Groves mechanisms (see chapter 10).

The extension to multidimensional project spaces[15] is also immediate by taking for example the class of utility functions:

$$v_i(K, \theta) = \frac{K'QK}{2} + K'\theta \tag{4.27}$$

where $\theta \in \mathbf{R}^L$, $K \in \mathbf{R}^L$,

K' is the transpose of K, and

Q is a negative definite matrix of constants.

It should be emphasized that one of the great advantages of the mechanisms being discussed is their ability to deal successfully with situations involving non-convex preferences. Although some problems of a purely computa-

[13] This argument shows that, in the quadratic case, (4.24) characterizes the class of transfer functions generating *SIIC* ***DRM***.

Indeed, with regularity conditions, (4.24) characterizes the *SIIC* ***DRM***, for valuation functions $v_i(\theta, K)$, parameterized by θ in an open set Θ.

[14] The extension of theorem 4.5 to valuation functions parameterized by points in a multi-dimensional space can be sketched as follows.

First one integrates the system of differential equations obtained from the analog of (4.24). When θ_i is multidimensional one has to verify that the Poincaré conditions are satisfied ($\partial^2 t_i/\partial\theta_i'\partial\theta_i = \partial^2 t_i/\partial\theta_i\partial\theta_i'$) but this presents no problem. One can then directly argue from the derivatives of $t_i(\cdot)$ that the transfer functions belong to the Groves class since we know that they contain the Groves class.

[15] The theorem of Walker [1978] addresses this issue. It can be stated as follows: Let $\mathscr{K}$ be compact and let V_i be the set of all concave functions over $\mathscr{K}$ for each i. Then the set of all *SIIC* successful relevation mechanisms is identical with the set of Groves mechanisms.

Except for the difference in the conditions on $\mathscr{K}$, the multi-dimensional version of theorem 4.5 discussed in the text would yield the same conclusion as Walker's theorem. Clearly, however, once $\mathscr{K}$ is fixed, we can find a region of the parameter space Θ, such that for any $\theta_1, \ldots, \theta_N$ in Θ, the optimal project defined by the maximization of the sum of the quadratic functions (4.27) would lie in $\mathscr{K}$. Therefore, applying theorem 4.5 to the quadratic utilities defined over this parameter space implies Walker's result, but uses a much smaller set of concave functions over $\mathscr{K}$ than the full set.

Walker's proof follows our general method as used in theorem 4.3 – namely, it is a proof by contradiction using a false preference pattern – rather than the constructive method which we have found applicable in concave, differentiable cases.

tional nature will arise in the actual process of maximizing $\sum_i v_i$ over a multidimensional domain, the fact that such taste patterns can be handled is an unintentional by-product of the solution to the revelation problem. Therefore, it would be a step backwards, in a sense, to require convexity where it is not needed. The virtue of theorem 4.5 is that it shows that there is no need for doing so, apart from the computational difficulties which, it is well-known, are unavoidable in non-convex problems.

These theorems settle the first of the two questions mentioned above concerning potential extensions of our basic characterization result, theorem 4.3. The remaining issue, whether a superior form of mechanism can be achieved using techniques other than direct revelation, is treated in the next section.

4.5. Optimality and characterization theorems: Extensions

When larger strategy spaces are envisioned, the possibility that some individuals will have multiple dominant strategies must be faced. In the first two theorems below, we assume this away and thus restrict attention to mechanisms that have unique outcomes when dominant strategies are played. First we need a definition.

Definition 4.16. A revelation mechanism will be called *normalized* (relative to a null project, $\overline{K}$) if the set of allowable strategies is the set of strategies that are normalized (relative to $\overline{K}$) – that is, $w_i(\overline{K}) = 0$, for $i = 1, \ldots, N$.

Theorem 4.6. A successful *SIIC* normalized revelation mechanism is a normalized Groves mechanism.

Proof Since the revelation mechanism is successful *SIIC* and normalized, agents will answer their true valuation function v_i, $i = 1, \ldots, N$ by corollary 4.2. Since it is successful, we can say that the decision is taken by maximizing the sum of the answers. Therefore, it is a direct revelation mechanism. Hence, the result by theorem 4.3. Q.E.D.

Theorem 4.7. A satisfactory mechanism (S, f) which satisfies the property of uniqueness of dominant strategies is such that there exist functions

$$\psi_i: S_i' \to V_i, \quad i = 1, \ldots, N,$$

and a normalized Groves mechanism, $NGM = \{V, g\}$ such that

$$f(s) = g[\psi_1(s_1), \ldots, \psi_N(s_N)]$$

Proof. Let $\mathscr{D}_i(v_i)$ be the unique dominant strategy of agent i when the truth is v_i, $i = 1, \ldots, N$.

From the mechanism (S, f), we construct a normalized revelation mechanism (V, ϕ) as follows: Let

$$\phi(w) = f[\mathscr{D}(w)] \quad \text{for any } w \in V \tag{4.28}$$

where $V = \prod_i V_i$, the space of upper-semicontinuous normalized valuation functions.

The revelation mechanism (V, ϕ) is well-defined since $\mathscr{D}_i(v_i)$ is a singleton for $i = 1, \ldots, N$.

We want to show that (V, ϕ) is *SIIC*. Suppose it is not. For some v_i, w_{-i}, there exists $w_i \neq v_i$ with $\phi(w_{-i}, w_i)$ preferred by individual i to $\phi(w_{-i}, v_i)$, i.e., $f[\mathscr{D}_{-i}(w_{-i}), \mathscr{D}_i(w_i)]$ preferred to $f[\mathscr{D}_{-i}(w_{-i}), \mathscr{D}_i(v_i)]$ where

$$\mathscr{D}_{-i}(w_{-i}) = [(\mathscr{D}_1(w_1), \ldots, \mathscr{D}_{i-1}(w_{i-1}), \mathscr{D}_{i+1}(w_{i+1}), \ldots, \mathscr{D}_N(w_N)] \tag{4.29}$$

therefore,

$$u_i(\mathscr{D}_{-i}(w_{-i}), \mathscr{D}_i(w_i); \{S, f\}) > u_i(\mathscr{D}_{-i}(w_{-i}), \mathscr{D}_i(v_i); \{S, f\}) \tag{4.30}$$

which contradicts the fact that $\mathscr{D}_i(v_i)$ is a dominant strategy. Also, (V, ϕ) is successful since it has the same outcomes as (S, f). Therefore, it is a successful *SIIC* normalized revelation mechanism, and consequently a normalized Groves mechanism from theorem 4.4.

Finally, to show that $\mathscr{D}_i^{-1}$, $i = 1, \ldots, N$, exist as functions, we prove that $\mathscr{D}_i$ is one-to-one for $i = 1, \ldots, N$. Suppose $\mathscr{D}_i$ is not one-to-one. Then, there exist v_i and v_i' in V_i, with $v_i \neq v_i'$ such that $\mathscr{D}_i(v_i) = \mathscr{D}_i(v_i')$. By definition $v_i(\overline{K}) = v_i'(\overline{K}) = 0$, but since $v_i \neq v_i'$, there exists $K \in \mathscr{K}$, $K \neq \overline{K}$, such that $v_i(K) \neq v_i'(K)$.

Without loss of generality, let K^* such that

$$v_i(K^*) - v_i'(K^*) = \varepsilon \tag{4.31}$$

Let $A = \sup\,[\sup_{K \in \mathscr{K}} v_i(K), \sup_{K \in \mathscr{K}} v_i'(K)]$

We can choose v_{-i} such that:

$$\begin{aligned} &\sum_{j \neq i} v_j(K^*) = -v_i'(K^*) - \frac{\varepsilon}{2} \\ &\sum_{j \neq i} v_j(\overline{K}) = 0 \\ &\sum_{j \neq i} v_j(K) = -A - \varepsilon \end{aligned} \tag{4.32}$$

Then, clearly,

$$\sum_{j\neq i} v_j(K^*) + v_i(K^*) > \sum_{j\neq i} v_j(K) + v_i(K) \qquad \forall K \in \mathscr{K},\ K \neq K^* \tag{4.33}$$

and

$$\sum_{j\neq i} v_j(\bar{K}) + v_i'(\bar{K}) > \sum_{j\neq i} v_j(K) + v_i'(K) \qquad \forall K \in \mathscr{K},\ K \neq \bar{K}. \tag{4.34}$$

Therefore, we are able to construct v_i such that (v_{-i}, v_i) and (v_{-i}, v_i') should lead to different decisions under the Pareto criterion, and they do not, contradicting the successfulness of (V, ϕ).

Therefore, $\mathscr{D}_i$ is one-to-one for $i = 1, \ldots, N$. If we define $\psi_i = \mathscr{D}_i^{-1}$ from S_i' into V_i, then,

$$g[\psi_1(s_1), \ldots, \psi_N(s_N)] = f[\mathscr{D}_1 \circ \mathscr{D}_1^{-1}(s_1), \ldots, \mathscr{D}_N \circ \mathscr{D}_N^{-1}(s_N)] = f(s) \tag{4.35}$$

Q.E.D.

Without the condition of uniqueness of dominant strategies, other mechanisms are possible. However, it may be argued that the way in which these mechanisms differ from Groves mechanisms does not have any essential relation to either the resource allocation problem or the individual's incentive problems. The following example brings this out:

Example 4.1. Let $\mathscr{K} = \{0, 1\}$, $S_i = \mathbf{R}^2$, $s_i = (s_{i1}, s_{i2})$, $i = 1, \ldots, N$, and define the mechanism as follows:

$$\begin{aligned} d(s) &= 1 \quad \text{if} \quad \sum_i s_{i1} \geqq 0 \\ &= 0 \quad \text{if} \quad \sum_i s_{i1} < 0 \\ t_i(s) &= \sum_{j\neq i} s_{j1} + \sum_{j\neq i} s_{j2} \quad \text{if} \quad d(s) = 1 \\ &= \sum_{j\neq i} s_{j2} \quad \text{if} \quad d(s) = 0 \\ &\qquad \text{for } i = 1, \ldots, N. \end{aligned} \tag{4.36}$$

Then, the set of dominant strategies is:

$$\mathscr{D}_i(v_i) = \{(s_{i1}, s_{i2}) \in \mathbf{R}^2 \mid s_{i1} = v_i(1)\} \tag{4.37}$$

Therefore, this mechanism cannot be expressed as $g[\psi(s)]$ for any Groves mechanism since the functions t_i, $i = 1, \ldots, N$ are not constant over $\prod_i \mathscr{D}_i(v_i)$.

The force of this example is that since t_i does not depend on s_{j2} for $j \neq i$, the second part of j's strategy is really irrelevant. Because no overall resource constraint on the transfers has been introduced, j has no reason at all to worry about this irrelevant parameter of his strategy. When dealing with such mechanisms, therefore, the relevant comparison is with the concept of *extended* mechanisms of various types. These allow for set-valued outcomes. It is clear that by setting $s_{i2} = 0$ for $i = 1, 2$, we can completely eliminate the arbitrariness of the transfers due to this term. We will have converted the mechanism above into an ordinary Groves mechanism. When the multiplicity of dominant strategies is more complex, such a simple device as elimination of irrelevant dimensions of the strategy space may not work. Nevertheless, as the following theorem shows, there is always a Groves mechanism embedded in any satisfactory mechanism.

Theorem 4.8. Let (S, f) be a satisfactory mechanism. Then, there exist functions ψ_i from S_i' into V_i, $i = 1, \ldots, N$ and an extended normalized Groves mechanism (V, Φ) such that

$$\bigcup_{s' \in S(s)} f(s') = \Phi[\psi_1(s_1), \ldots, \psi_N(s_N)]$$

where $S(s) = \{s' \mid s \text{ and } s' \text{ belong to the same } \mathscr{D}(w)\}$.

Proof. It is a matter of routine to check that theorem 4.2, theorem 4.3, and theorem 4.6 are true for extended (direct) revelation mechanisms.[16] The proof then follows the lines of the proof of theorem 4.7 with some differences noted below. Now $\mathscr{D}_i(v_i)$ is the set of dominant strategies of agent i when the truth is v_i, $i = 1, \ldots, N$. We construct an extended normaliized revelation mechanism as follows. Let V_i be the set of normalized upper-semicontinuous valuation functions and let $\Phi(w) = f[\mathscr{D}(w)]$ for all $w \in V$. Φ is now a correspondence with two properties:

$$s^1 \in \mathscr{D}(v) \text{ and } s^2 \in \mathscr{D}(v) \tag{4.38}$$

implies that both $d(s^1)$ and $d(s^2)$ maximize $\sum_i v_i$; otherwise the mechanisms would not be successful.

Also,

$$v_i(d(s^1)) + T_i(s^1) = v_i(d(s^2)) + T_i(s^2). \tag{4.39}$$

Otherwise, there would exist $t_i \in T_i$ such that, without loss of generality:

[16] An extended revelation mechanism is an ***EDRM*** without the requirement that $D(w)$ be contained in the maximizers of $\Sigma_i w_i(\cdot)$. See definition 4.5.

$$v_i(d(s^1)) + t_i(s^1) > v_i(d(s^2)) + t_i(s^2) \tag{4.40}$$

and then s_i^2 would not be a dominant strategy for agent i.

As in theorem 4.7, it is shown that (V, Φ) is *SIIC*. By theorem 4.6, (V, Φ) is then an extended normalized Groves mechanism.

Now, if $v_i \neq v_i'$ then $\mathscr{D}_i(v_i) \cap \mathscr{D}_i(v_i') = \emptyset$. Suppose, on the contrary, that there exists $s_i \in \mathscr{D}_i(v_i) \cap \mathscr{D}_i(v_i')$. Then, v_i and v_i' may lead to the same project K^*. As in theorem 4.7, we can choose $\sum_{j \neq i} w_j$ such that v_i and v_i' should lead to different projects, contradicting the successfulness of (V, Φ).

It is therefore possible to define the functions $\psi_i = \mathscr{D}_i^{-1}$, $i = 1, \ldots, N$, which are such that:

$$\mathscr{D}_i \circ \mathscr{D}_i^{-1}(s_i) = S_i(s_i) = \{s_i' \mid s_i \text{ and } s_i' \text{ belong to the same } \mathscr{D}_i(v_i)\}. \tag{4.41}$$

Then,

$$\begin{aligned} \Phi[\psi_1(s_1), \ldots, \psi_N(s_N)] &= f[\mathscr{D}_1 \circ \mathscr{D}_1^{-1}(s_1), \ldots, \mathscr{D}_N \circ \mathscr{D}_N^{-1}(s_N)] \\ &= \bigcup_{s' \in S(s)} f(s') \quad \text{Q.E.D.} \end{aligned} \tag{4.42}$$

4.6. Applications: Successful and unsuccessful

We have tried to make the case above that Groves mechanisms are particularly attractive as a method for making resource allocation decisions, both public and private, whenever strategic considerations might stand in the way.

It can clearly be applied in the classical case of the free-rider problem, for a decision concerning the level and mixture of public expenditure programs. There are also interesting applications to private goods economies. The second-price auction described in chapter 3 for efficiently allocating an indivisible private good is one. A related application, also proposed by Vickrey [1961], has to do with an ordinary competitive market in which each buyer and each seller announce their schedule of demand or supply prices. If the usual competitive rules are followed and the number of economic agents is small, the incentives for a truthful revelation of these price schedules are not present. A natural consequence of the above methods, viewing the set $\mathscr{K}$ as the vector of allocations of the commodity in question, is to choose those quantities which would arise under the competitive rules, but to use different prices for different agents. The price relevant to each trader is the one that *would have been*

the equilibrium price in a competitive market in which his demand or supply schedule had been deleted.

It is unfortunate, however, that trying to use this idea in several highly desirable directions turns out to involve necessary contradictions to the assumptions used above. We will proceed with the rather sad task of this section as follows: Three desirable applications for this problem are discussed. They are the idea that the income distribution is itself a public good, and therefore an optimal income distribution might be found through incentive compatible means; the fact that the same physical public projects can be financed in a variety of ways and the decision about the financing plan could be combined with the selection of the project; and finally, the problem of introducing distributional considerations into the analysis through adopting other welfare criteria than the maximization of the sum of utilities, equally weighted. In each case, we show why Groves mechanisms cannot circumvent the difficulty.

In the next chapter we will go on to consider whether Groves mechanisms or suitable generalizations might be applied in economic environments not satisfying the restrictions we have used thus far. The answer is that the assumptions which preclude successful treatment of the problems above are essential for the incentive problem, as long as dominant strategy outcomes are required. Some generalization is, however, possible in other directions. Combining the results of this section with the next chapter therefore delineates the range of valid application of Groves procedures.

4.6.1. *The income distribution as a public good*

Ethical considerations, when incorporated into individual preference patterns, can be viewed under certain circumstances as converting the income distribution itself into a public good. It is, after all, a common variable in everyone's economic environment and the actions of any one agent affect it in a negligible way. Arguing on this basis, Thurow [1971] and Brown, Fane and Medoff [1973] have derived an optimal tax/subsidy system, that is, a tax/subsidy system capable of attaining a Pareto optimum for any set of utility functions for the individuals. Moreover, this system will always achieve a resource allocation that is superior to the original distribution for every individual. It is natural to ask, whether the free-rider problem poses serious difficulties for these redistributive schemes. By feigning an unconcerned attitude regarding the welfare of others, the individual would avoid taxation for redistributive purposes and would thus experience a higher utility at this distortive Pareto optimum than at the honestly attained one.[17] One might hope to elicit preferences for equality via a Groves mechanism and then to attain the optimal income distribution.

When individuals' preferences are concerned with the income distribution in the population, we can, in general, express these tastes as,

$$u_i(t) = v_i(t) + t_i, \qquad i = 1, \ldots, N \tag{4.43}$$

where $t = (t_1, \ldots, t_N)$ is the vector of transfers to all agents. Written in this way, the valuation function over "public projects", which are now just specifications of transfers, are, in particular, dependent on the agent's own transfer. This violates the assumption of separability, and we shall show in the next chapter that satisfactory mechanisms for this class of preferences do not exist. We therefore turn our attention to valuation functions that depend only on the transfers to the other agents.

The question we ask is whether, when each agent can announce his preferences for the income distribution in this way, the optimal income distribution can be attained via the imposition of a Groves mechanism. By obtaining a negative answer to this question, we will have proven, by virtue of the characterization theorems of this chapter, that there is no incentive compatible method for determining the optimal income distribution.

Before demonstrating this, via a counterexample, we will restrict the class of allowable preference functions further, in an economically interesting way. Each agent may be supposed to have preferences over the *distribution* of others' incomes, but not over specifically *which* individual among the others attains *which* income level. This type of symmetry condition seems very natural in the present framework. By proving that even this class of preferences cannot be satisfactorily optimized by Groves mechanisms, we strengthen the negative character of the result obtained.

We will argue as follows: For given preferences and initial endowments, we can find the optimal system of transfers within some specified feasible set. To implement these as consistent outcomes of a Groves mechanism is to fix the values of the arbitrary h-functions at the truthful strategy, which is the one that would be played. Now change the income distribution preferences of a single agent. In general, this will change the optimal transfers to all agents, and consequently their valuations of the transfer

[17] Part of the problem may be inherent in the formulation of utility as dependent on inequity of the income distribution instead of on the "utility distribution", but we will not delve into the matter in this regard.

system achieved. In this new environment, we can again calculate the h-functions required to implement the optimal transfers by a Groves mechanism. Since everyone else is still playing the same strategy, the implied value of the h-function for the agent whose tastes have been varied should be the same. By demonstrating that it must be different, we will have established the desired counterexample to the fact that optimal income distributions, in this sense, can be achieved via a satisfactory procedure.

The details of the construction follow:

Example 4.2. There are three agents, $i = 1, 2, 3$. The set of possible transfers is

$$\mathscr{K} = \{(t_1, t_2, t_3) \mid t_1 + t_2 + t_3 = 0,\ t_1 = t_2\}. \tag{4.44}$$

This means that we are considering purely redistributive schemes, in which, perhaps for some institutional reasons, the same transfers must apply to agents 1 and 2, even though their initial endowments are not necessarily equal.[18] Initial endowments are denoted $\overline{x}_i$. It will be useful to define

$$\begin{aligned} \Delta_{12} &= \overline{x}_1 - \overline{x}_2 \\ \Delta_{13} &= \overline{x}_1 - \overline{x}_3 \end{aligned} \tag{4.45}$$

These parameters define the inequality present in the initial allocation.

Valuation functions are

$$\begin{aligned} v_1(t_2, t_3) &= -(t_3 - t_2 + \Delta_{12} - \Delta_{13})^2 \\ v_2(t_1, t_3) &= -(t_1 - t_3 + \Delta_{13})^2 \\ v_3(t_1, t_2) &= -(t_1 - t_2 + \Delta_{12})^2 \end{aligned} \tag{4.46}$$

Note that they are consistent with our hypothesis that each agent's preferences about the others' incomes is a function only of the distribution of these magnitudes, and is not specifically related to that of any particular individual.

One can calculate straightforwardly the system of transfers in (4.44) that maximizes $\sum_i v_i$. These are given by

$$\begin{aligned} t_1^* = t_2^* &= \frac{\Delta_{12} - 2\Delta_{13}}{6} \\ t_3^* &= \frac{-2\Delta_{12} + 4\Delta_{13}}{6} \end{aligned} \tag{4.47}$$

[18] Similar restrictions often arise in optimal tax theory, see Sandmo [1976] for a useful survey of this literature.

Using these transfers, the valuation functions of agents 2 and 3, and the definition of Groves mechanisms we see that agent 1's transfer must satisfy

$$t_1^* = \frac{-5\Delta_{12}^2}{4} + h_1(v_2, v_3)$$

Thus, from (4.47) and (4.48), we obtain:

$$h_1(v_2, v_3) = \tfrac{5}{4}\Delta_{12}^2 + \frac{\Delta_{12}}{6} - \frac{2\Delta_{13}}{6} \qquad (4.49)$$

Now, consider the alternative valuation function for agent 1 given by

$$v_1'(t_2, t_3) = -2(t_3 - t_2 + \Delta_{12} - \Delta_{13})^2 \qquad (4.50)$$

This produces the optimal transfers

$$t_1^{*\prime} = t_2^{*\prime} = \frac{2\Delta_{12} - 3\Delta_{13}}{9}$$
$$t_3^{*\prime} = \frac{-4\Delta_{12} + 6\Delta_{13}}{9} \qquad (4.51)$$

and, in a manner analogous to the above, we find that

$$h_1(v_2, v_3) = \tfrac{13}{9}\Delta_{12}^2 + \frac{2\Delta_{12}}{6} - \frac{2\Delta_{13}}{6}. \qquad (4.52)$$

Therefore, except in the coincidental case where $\frac{13}{9}\Delta_{12} + \frac{2}{6} = \frac{5}{4}\Delta_{12} + \frac{1}{6}$, equations (4.49) and (4.52) are inconsistent. Since v_2 and v_3 are unchanged, the h-function used in the Groves mechanism must take the same value at these truthful preference revelations in each instance. Hence these transfers cannot be the result of any Groves mechanism.

4.6.2. *The choice among financing plans*

In the discussion of chapter 3 we have argued that private preferences for alternative methods of financing public projects can be embodied in preferences for projects themselves, by considering a larger set of "projects", namely those derivable by looking at the physical projects and their financing simultaneously. While no slight-of-hand was intended, the impression derived from our analysis could be greatly misleading if one were to assume that incentive compatible mechanisms can be arranged so as to simultaneously allocate the costs of the project itself. To see this, one need only consider the situation in which there is one costly project

possible, in physical terms, and the status quo, or "no project". The cost of the project is C, and this is to be covered by payments of each individual, c_i, $i = 1, \ldots, N$, such that $\sum_i c_i = C$. Suppose the ith individual is willing to pay v_i' for the project itself. Including the financing plan $(c_1, \ldots, c_N)$, he is willing to pay

$$v_i = v_i' - c_i \qquad (4.53)$$

Letting the set of allowable projects, $\mathscr{K}$, be $\{0, 1\} \times \{(c_1, \ldots, c_N) \mid \sum_i c_i = C\}$, and instituting a Groves mechanism, it is clear that if $\sum_i v_i' > C$, the maximizing elements of $\sum_i v_i$ in $\mathscr{K}$ will be $1 \times \{(c_1, \ldots, c_N) \mid \sum_i c_i = C\}$. That is, if accepting the project is the efficient decision, it will emerge from the Groves mechanism, but $\sum_i v_i$ will be a constant with respect to the financing plan. Any selection from this optimizing correspondence will be purely arbitrary, since the mechanism makes no distinction among financing plans. Thus the problem of choosing an allocation of costs for public projects remains outside the realm of public decisions that can effectively be made using dominant strategy mechanisms.

4.6.3. *Distributional considerations and the neutrality of dominant strategy mechanisms*

We have shown that Groves mechanisms can achieve Pareto optimal outcomes and individuals will have dominant strategies for every possible specification of the economic environments within the class we consider. Their great advantage is that the planner need have no information about the environment, save the fact that it satisfies the conditions necessary for the employment of the mechanism. Presumably the specific mechanism to be used is chosen, and then, when applied to a given environment, it produces a well-defined resource allocation. But it is possible to reverse this temporal sequence, asking instead whether a Groves mechanism exists capable of attaining a given Pareto optimum. This concept, termed "neutrality" by Champsaur[19] [1976], expresses the fact that the nature of the class of procedures itself does not bias it towards, or away from, any specific allocations in the optimal set.

[19] Champsaur [1976] studied the neutrality problem in the context of dynamic planning procedures, but the definition carries over straightforwardly to the static context we consider. In chapter 15 we consider such dynamic mechanisms, played iteratively under the assumption that behavior is myopic. Chapter 5 also discusses the neutrality issue in some static mechanisms without the assumption of unlimited transfers.

For the purpose of the present set of mechanisms, this question can be trivially answered in the affirmative. Since the allowable transfers are unrestricted, any Pareto optimum involves a project choice $K \in \mathscr{K}$ which maximizes $\sum_i v_i(K)$. Any private goods allocation can then be attained by adding, to a fixed choice of h-functions, the differences between the transfers that these h-functions would induce at the known preference pattern and those specified by the desired optimal allocation. In models where transfers are unrestricted, the class of Groves mechanisms is neutral. The case where transfers are constrained to maintain non-negative consumptions for each individual is treated in the next chapter.

Another type of distributional consideration involves allowing the agents to have different degrees of influence in determining the decision. That is, rather than weighting their utility functions equally, we use weights, λ_i, in the objective function to determine the project. Such a scheme, if it can succeed in maximizing $\sum_i \lambda_i v_i(K)$, will not be selecting Pareto optima relative to the preferences v_i except perhaps by coincidence. It may be defended on the grounds that it provides some, imperfect, way of compensating for an adverse initial income distribution. Alternatively, this method may be used to reflect the fact that the planner knows that some agents, or groups of agents, have much better information than others concerning the true characteristics or consequences of adopting the project. In such situations, with imperfect personal information,[20] the concept of Pareto optimality ex ante as defined by the preferences v_i may not be the desired welfare criterion. Weighting these as described is one way of approaching the goal of *ex post* optimality in imperfect information settings.

Accepting this idea, it is easy to see that the following modification of Groves mechanisms suffices to insure that $\sum_i \lambda_i v_i(K)$ will be maximized for the true preferences v_i, as long as $\lambda_i > 0$ for all i:

$$d(w) \text{ maximizes } \textstyle\sum_j \lambda_j w_j(K) \text{ over } \mathscr{K}$$

$$t_i(w) = \frac{\sum_{j \neq i} \lambda_j w_j(d(w))}{\lambda_i} + h_i(w_{-i}). \tag{4.54}$$

[20] See chapter 13, where endogenous improvements in imperfect information are analyzed the context of incentives in large populations.

4.6.4. Unobservable characteristics and distributional considerations in dominant strategy mechanisms

We can see from the previous subsection that if the planner desires to base the distributional weights λ_i on observable characteristics of the agents, this can be done directly. A more subtle problem arises if these characteristics are unobservable and must be discovered in the course of the social decision process. For example, wealth may be hard to measure, and yet the planner may want to give greater weight to the preferences of poorer individuals as an implicit form of income redistribution.

If one were to ask agents, directly, to reveal these characteristics, there would be no incentive not to maximize one's influence on the outcome by lying. It is natural, therefore, in our context, to try to include the characteristics as an argument of the transfer function. Unfortunately this type of mechanism cannot remain incentive compatible in both the dimensions of the preferences and these characteristics.

Another point of view relevant to the consideration of costly projects is that the cost shares, which we define ex ante in the description of the project itself, might be adjusted to reflect these characteristics. This can be treated by following our original practice of choosing a particular distribution of costs and then adjusting these cost shares by making the transfer function dependent on the characteristics. From a formal standpoint, this problem is identical to that of eliciting them for the purpose of adjusting the distributional weights as mentioned above.

Let us assume that the ith agent's characteristic is a real number $\tilde{\eta}_i$, and let $\tilde{\eta} = (\tilde{\eta}_1, \ldots, \tilde{\eta}_N)$ be the vector of characteristics of the members of the population. The projects space is assumed to be the real line, $\mathscr{K} = \mathbf{R}$.

Let $K^*(\theta, \eta)$ be the decision function depending on the announced values of η and on the vector $\theta = (\theta_1, \ldots, \theta_N)$ where θ_i is the announced value of a parameter of the preferences of agent i, $v(\cdot, \tilde{\theta}_i)$ over $\mathscr{K}$. For example, $K^*(\theta, \eta)$ may be derived from the maximization of

$$\sum_i \lambda(\eta_i) v(K, \theta_i) \tag{4.55}$$

which would correspond to the first interpretation given above.

The maximization of the utility function

$$v(K^*(\theta, \eta), \tilde{\theta}_i) + t_i(\theta, \eta) \tag{4.56}$$

by agent i leads to the necessary first-order conditions

$$\frac{\partial v}{\partial K}(K^*(\theta,\eta),\tilde{\theta}_i)\frac{\partial K^*}{\partial \theta_i}(\theta,\eta)+\frac{\partial t_i}{\partial \theta_i}(\theta,\eta)=0$$
$$\frac{\partial v}{\partial K}(K^*(\theta,\eta),\tilde{\theta}_i)\frac{\partial K^*}{\partial \eta_i}(\theta,\eta)+\frac{\partial t_i}{\partial \eta_i}(\theta,\eta)=0 \tag{4.57}$$

If the true values $\theta_i = \tilde{\theta}_i$, $\eta_i = \tilde{\eta}_i$ are to be dominant strategies for all possible values of these parameters, (4.57) must hold as an identity when evaluated at the truthful point. Thus,

$$\frac{\partial t_i}{\partial \theta_i}(\tilde{\theta},\tilde{\eta}) = -\frac{\partial v}{\partial K}(K^*(\tilde{\theta},\tilde{\eta}),\tilde{\theta}_i)\frac{\partial K^*}{\partial \theta_i}(\tilde{\theta},\tilde{\eta})$$
$$\frac{\partial t_i}{\partial \eta_i}(\tilde{\theta},\tilde{\eta}) = -\frac{\partial v}{\partial K}(K^*(\tilde{\theta},\tilde{\eta}),\tilde{\theta}_i)\frac{\partial K^*}{\partial \eta_i}(\tilde{\theta},\tilde{\eta}) \tag{4.58}$$

The equality of the cross-derivatives of $t_i(\cdot\,,\cdot)$ would imply

$$\frac{\partial^2 v}{\partial K \partial \theta_i}\cdot\frac{\partial K^*}{\partial \eta_i} \equiv 0 \tag{4.59}$$

The identity (4.59) implies that v must be additively separable in θ_i whenever η_i genuinely influences the decision. Since we have assumed that the goal of the process is precisely to incorporate $\tilde{\eta}_i$ into the decision-making process, and that the relative evaluations of projects are influenced by θ_i, (4.59) is contradicted. This proves, in differentiable environments, the impossibility of eliciting both preferences and distributional characteristics.

Chapter 5

ON THE LIMITATIONS OF DOMINANT-STRATEGY MECHANISMS

5.1. Introduction

Despite the negative results of the final section in the last chapter, it must be said that the potential application of the devices we study, even in the limited range of cases for which they are designed, represents a significant advance over the previous state of the art in the economics of incentives. In this chapter we study the question of whether the same class of mechanisms can be extended to cover more complex economic environments and other types of strategic behavior.

First we study mechanisms in economies with non-zero income effects. Then, we ask whether restrictions can be placed on the mechanism that will make the sum of the transfers among individuals identically zero. If this could be done, the mechanism would then be applicable to private goods situations such as the auction or counter-speculation ideas discussed in chapter 3. Next, we explore the potential application of dominant strategy mechanisms to environments without income effects but with some lower bounds on the magnitudes of possible transfers to individuals. Finally, the possibility of designing mechanisms to avoid manipulation by coalitions is treated.

In all cases except the bounded transfer problem, our results are primarily negative. We return to these issues in chapters 11, 9 and 10, respectively, where we explore potential approximate solutions based on large numbers arguments.

5.2. Separability

The most severe restriction of the previous chapters is that of additive separability of tastes, or the absence of income effects, which we will refer to as separability throughout this chapter. If it is desired to use incentive-compatible mechanisms in a model where the public decision taken will affect the demand for other commodities, we must weaken the separability assumption. Unfortunately this cannot be done, at least if we want to allow the full class of separable preferences to remain in the domain of the process. The possibility of using stepwise schemes with myopic agents may allow us to circumvent this important obstacle. Because of the complexity of the dynamic processes involved, we postpone this until the final part of the book. Presently we simply show why the Groves mechanisms break down in many cases of interest, without separability.

Although exact solutions are not possible, we will demonstrate in this section that asymptotic results on the successfulness of some mechanisms can be achieved when a sequence of non-separable preference patterns converges to a separable relation. This may be useful either when income effects are weak or the project is a small one.

Let us consider the case of a fixed-size project, $\mathscr{K} = \{0, 1\}$, with the normalization $\overline{K} = 0$ so that separable preferences are defined by

$$\begin{aligned} u_i(0, t_i) &= t_i \\ u_i(1, t_i) &= v_i + t_i, \end{aligned} \qquad (5.1)$$

where v_i is a constant as usual.

Without separability, v_i would depend on t_i and

$$u_i(1, t_i) = v_i(t_i) + t_i. \qquad (5.2)$$

The first problem is to define successfulness in a world without separability. Previously, the combination of separability of utility and transferability of resources implied that all Pareto optima could be found by maximizing $\sum_i v_i(\cdot)$ over $\mathscr{K}$. Indeed this was equivalent to maximizing $\sum_i u_i(K, t_i) - \sum_i t_i$ simultaneously with respect to K and t_i, for the transfer variables would not affect the maximum with respect to K, and without constraints on $\sum_i t_i$, there was no force to the maximization with respect to these variables. The usual prescription of solving

$$\max_{K \in \mathscr{K}} \sum_i \lambda_i u_i(K, t_i) - \sum_i t_i \qquad (5.3)$$

for arbitrary non-negative weights λ_i, in order to trace out all Pareto optima would not lead to determinate results in this case.

Let V be the set of all normalized valuation functions (of t now, not of K) that are in the domain of the mechanism. Only revelation mechanisms will be treated here. They map the N-fold product of V with itself, V^N, into $\{0, 1\} \times \mathbf{R}^N$: $f(w(\cdot)) = (d(w(\cdot)), t_1(w(\cdot)), \ldots, t_N(w(\cdot)))$.

Definition 5.1. A revelation mechanism will be called *successful over V* if whenever the announced valuation functions are $w(\cdot) = (w_1(\cdot), \ldots, w_N(\cdot)) \in V^N$ the decision taken $d(w(\cdot))$ maximizes

$$\sum_i u_i(t_i(w(\cdot)))$$

over all elements of $\mathscr{K}$.

This definition bears some discussion. The mechanism produces simultaneously a decision about K and a pattern of transfers. Under the criterion of maximizing the sum of utilities, the optimal choice of K depends on the transfers. We require only that this choice be taken optimally given the transfer pattern induced.

This is a weaker requirement than the maximization of the sum of utilities, minus $\sum_i t_i$, with respect to both K and t, and hence does not really correspond to our notion of Pareto optimality in this sense. However, since we are about to demonstrate that even K alone cannot be optimized for all environments in V, whenever V is larger than the set of all constant functions, we are negating the possibility of maximizing $\sum_i u_i$ at the same time. On the other hand we have not explicitly ruled out the possibility that $\sum_i \lambda_i u_i - \sum_i t_i$ could be maximized for some set of λ_i. In principle, Pareto optima might still be attainable. But this too can be shown to be impossible, using methods analogous to those of section 4.6.

We present below two cases in which by including a class of non-constant functions in the allowable set of valuation functions, the existence of mechanisms that are successful in the above sense will be precluded.

Theorem 5.1. Let V contain all constant functions and the class of step functions with values a and b, $b > a$, for transfers above or below the level t^*. Then there is no revelation mechanism that is strongly individually incentive compatible and successful over V.

Proof. We prove the theorem by providing a counterexample. Consider the following two agents: agent 1 is such that

$$\begin{aligned} u_1(0, t_1) &= t_1 \\ u_1(1, t_1) &= t_1 + v_1 \end{aligned} \qquad (5.4)$$

He has a separable utility function. Agent 2 is such that

$$\begin{aligned} u_2(0, t_2) &= t_2 \\ u_2(1, t_2) &= t_2 + v_2 \quad \text{if } t_2 < t^* \\ &= t_2 + v_2' \quad \text{if } t_2 \geqq t^* \quad \text{with } v_2 < v_2' \end{aligned} \qquad (5.5)$$

Agent 2's evaluation of the project increases when the transfer he receives reaches a threshold value. Here the ***DRM*** associates to a value (v_1, v_2, v_2', t^*) a vector $(d, t_1, t_2) \in \{0, 1\} \times \mathbf{R} \times \mathbf{R}$ where d is the decision and (t_1, t_2) are the transfers received.

Let us consider a particular combination (v_2, v_2', v_1) which remains fixed for the rest of the argument and such that:

$$v_2 + v_1 < 0 \qquad v_2' + v_1 > 0 \qquad v_1 < 0 \tag{5.6}$$

and let t^* vary.

Successfulness means, in this case that

$$\begin{aligned} &\text{if } t_2 < t^* \text{ then } d = 0 && \text{since } v_2 + v_1 < 0 \\ &\text{if } t_2 \geqq t^* \text{ then } d = 1 && \text{since } v_2' + v_1 > 0 \end{aligned} \tag{5.7}$$

We prove that there exists no successful *SIIC* ***DRM*** by contradiction in a sequence of lemmas.

Lemma 5.1. It is not true that the project is accepted for all t^*.

Proof. Whenever the project is accepted $t_2 \geqq t^*$ by successfulness. Then, there would be an incentive to set t^* high to force a high t_2 (since $u_2(1, t_2)$ is increasing in t_2), contradicting *SIIC*, for agent 2. Q.E.D.

Lemma 5.2. It is not true that the project is rejected for all t^*.

Proof. If so, the transfer to agent 2 would be given by a function $t_2(t^*)$ which would be everywhere below t^*. Clearly, no constant function has this property. Let $t^{*\prime}$ and $t^{*\prime\prime}$ be such that $t_2(t^{*\prime}) < t_2(t^{*\prime\prime})$; then if the true tastes correspond to $t^{*\prime}$, the individual has an incentive to use $t^{*\prime\prime}$ instead (since $u_2(0, t_2)$ is increasing in t_2). Q.E.D.

Let T^* be the set of t^* that lead to acceptance and T^{*c} be the complement of that set. Both T^* and T^{*c} are non-empty.

Lemma 5.3. The transfers to individual 2 must be constant on T^* and T^{*c}.

Proof. If not (say on T^*), then if the $t^* \in T^*$ were associated with a lower transfer than some other $t^{*\prime} \in T^*$, the statement $t^{*\prime}$ would be better than t^* since the transfer would be higher and since $u_2(1, t_2)$ is increasing in t_2, contradicting *SIIC* ,for agent 2.

In the case of $T^{*\prime}$ the same argument holds. Q.E.D.

Lemma 5.4. Take $t^* \in T^*$ and $t^{*\prime} \in T^{*c}$, then $t_2(t^*) = t_2(t^{*\prime}) + v_1$.

Proof. First, let us compare the transfer with the statement (v_2, v_2', t^*) to the transfer with the constant statement w_2, such that $w_2 \geqq -v_1$. These transfers must be equal for if they were higher at (v_2, v_2', t^*), then (v_2, v_2', t^*) would be answered instead of w_2 when w_2 was true and vice versa (since $u_2(1, t_2)$ is increasing).

By a similar argument, we note that the transfer at (v_2, v_2', t^*) must equal that at constant statements w_2' such that $w_2' < -v_1$.

Then, by theorem 4.3, we know that the transfer at $w_2 \geqq -v_1$ is equal to that at $w_2' < -v_1$, plus v_1. Q.E.D.

We are now in a position to complete the proof of theorem 5.1.

Proof. By the above lemmas, we know that there exists a number $\bar{t}$ such that the transfer to agent 2 throughout the region T^* is $\bar{t} + v_1$ and the project is accepted, and throughout $T^{*\prime}$ it is $\bar{t}$ and the project is rejected. Moreover, since $v_1 < 0$, $\bar{t} + v_1 < \bar{t}$. Since the mechanism is successful, $t^* \in T^*$ implies that $t^* \leqq \bar{t} + v_1$ for if $t^* > \bar{t} + v_1$, then the preferences at $\bar{t} + v_1$ would lead to rejection rather than acceptance. Likewise, $t^{*\prime} \in T^{*c}$ implies $t^{*\prime} > \bar{t}$.

Therefore, letting $t^* \in (\bar{t} + v_1, \bar{t}]$, we obtain a contradiction to both the statements $t^* \in T^*$ and $t^* \in T^{*c}$, and hence we contradict the fact that the mechanism produces a well-defined outcome. Q.E.D.

To show that the discontinuous nature of step functions is not the critical feature of the above impossibility result, we will prove a different version using continuous strictly monotone functions. Negative results on separability require only that the class of functions be sufficiently rich, to permit advantageous manipulation.

Theorem 5.2. Let V include all constant valuation functions, one strictly monotone function, $\bar{v}(\cdot)$, and all shifts of $\bar{v}(\cdot)$ by a constant. Then there is no *SIIC* and successful revelation mechanism over V.

Proof. Since V contains all constant valuation functions, we know by theorem 4.2 that a mechanism successful over this class has transfer rules that can be written as

$$\begin{aligned} t_i(w) &= \sum_{j \neq i} w_j + h_i(w_{-i}) && \text{if } \sum_i w_i \geqq 0 \\ &= h_i(w_{-i}) && \text{if } \sum_i w_i < 0. \end{aligned} \tag{5.8}$$

Consider the case of two individuals, $i = 1, 2$.

Clearly, whenever one ($i = 1$) of them plays as his strategy a constant function the transfer to the other ($i = 2$) can depend only on the project selected and the level of this constant. Therefore, even if the second individual is playing $\bar{v}(\cdot)$, his transfer is given by (5.8). Let the levels of constants played by $i = 1$ for which the project is accepted when $i = 2$ plays $\bar{v}(\cdot)$, be

$$V_1 = \{v_1 \in \mathbf{R} \mid d(v_1, \bar{v}(\cdot)) = 1\} \tag{5.9}$$

The rejection set is

$$V_1^c = \{v_1 \in \mathbf{R} \mid \mathrm{d}(v_1, \bar{v}(\cdot)) = 0\} \tag{5.10}$$

Because of the assumed successfulness of the mechanism, we must have

$$\bar{v}(v_1 + h_2(v_1)) + v_1 \geqq 0 \qquad \text{for all } v_1 \in V_1 \tag{5.11}$$

$$\bar{v}(h_2(v_1)) + v_1 < 0 \qquad \text{for all } v_1 \in V_1^c \tag{5.12}$$

where $h_2(v_1)$ is the h-function for individual 2 as a function of individual 1's announced (constant) valuation function.

If the mechanism is to be decisive, the sets V_1 and V_1^c defined in (5.9), (5.10) must form a genuine partition of the real line. Thus, using (5.11), (5.12),

$$\{v_1 \mid \bar{v}(v_1 + h_2(v_1)) + v_1 \geqq 0\} \cup \{v_1 \mid \bar{v}(h_2(v_1)) + v_1 < 0\} = \mathbf{R} \tag{5.13}$$

is required. Let

$$\begin{aligned} \phi(v_1) &= \bar{v}(v_1 + h_2(v_1)) + v_1 \\ \phi'(v_1) &= \bar{v}(h_2(v_1)) + v_1 \end{aligned} \tag{5.14}$$

We treat the case of $\bar{v}$ monotone increasing; the case of $\bar{v}$ decreasing is completely analogous. Under this condition

$$\phi(v_1) - \phi'(v_1) < 0 \qquad \text{for } v_1 < 0 \tag{5.15}$$

Consider the valuation functions $\bar{v} + \alpha$, $\alpha \in \mathbf{R}$, which are all admissible valuations under the hypothesis of the theorem, and the corresponding functions. We will show that there exists a value of α for which (5.13) fails, and hence for which V_1 and V_1^c do not form a partition of $\mathbf{R}$.

$$\begin{aligned} \phi_\alpha(v_1) &= \phi(v_1) + \alpha \\ \phi'_\alpha(v_1) &= \phi'(v_1) + \alpha \end{aligned} \tag{5.16}$$

Take any $v_1 < 0$ and

$$\alpha = -\frac{\phi(v_1) + \phi'(v_1)}{2}.$$

Thus

$$\begin{aligned} \phi_\alpha(v_1) &= \frac{\phi(v_1) - \phi'(v_1)}{2} \\ \phi'_\alpha(v_1) &= -\left(\frac{\phi(v_1) - \phi'(v_1)}{2}\right). \end{aligned} \tag{5.17}$$

Using (5.15),

$$\phi_\alpha(v_1) < 0 < \phi'_\alpha(v_1). \tag{5.18}$$

Therefore if the valuation functions are $\bar{v} + \alpha$, for $i = 1$ and v_1 for $i = 2$, we have $v_1 \notin V_1$ and $v_1 \notin V_1^c$, contradicting the decisiveness of the mechanism. Q.E.D.

Thus, separability of all utility functions is not a requirement for the existence of a successful mechanism, but whenever the class of non-separable functions is sufficiently rich, they will be precluded. The interested reader can verify this impossibility for other parametric families of non-separable functions by techniques similar to those above.

Although a general solution is not available when allowable preferences are not separable, a mechanism related to the pivotal mechanism can be used to obtain a good approximation to the optimal decision in situations when income effects are small but not zero. It is not a revelation mechanism, but rather individuals are asked to respond with single number instead of the entire function which would be necessary to convey complete information in this case. We note in passing that the process thereby economizes on some of the informational requirements which, while outside the realm of this study, may have been severe obstacles nevertheless.

The idea of the mechanism is to elicit $v_i(0)$, the willingness to pay in the absence of transfers, and take the decision according to the sign of the sum of these answers, even though the actual transfers made will result in ex post valuations of $v_i(t_i)$ which may imply that the opposite decision is optimal. We will show that, although there is no dominant strategy, the optimal strategy for any individual given expectations about the statements to be made by the others, converges to the truthful value of $v_i(0)$ when $v_i(t_i)$ approaches a constant function. Then, under some conditions on the joint distribution of $v_i(0)$ and the derivative $(\mathrm{d}/\mathrm{d}t_i)v_i(0) = v'_i(0)$ across the members of the population, we find that the decision taken by this procedure will be a Pareto optimum with arbitrarily high probability. We recall that the Clarke mechanism is defined formally by

$$
\begin{aligned}
S_i &= \mathbf{R} & & i = 1, \ldots, N \\
d(w) &= 1 & & \text{iff } \sum_i w_i \geqq 0 \\
t_i(w) &= \sum_{j \neq i} w_j + \min(-\sum_{j \neq i} w_j, 0) & & \text{if } \sum_i w_i \geqq 0 \\
&= \min(-\sum_{j \neq i} w_j, 0) & & \text{if } \sum_i w_i < 0.
\end{aligned} \tag{5.19}
$$

Consider a typical individual whose willingness to pay function is $v_i(t_i)$. Let the sum of the responses of all of the other agents be x, and the individual's subjective distribution of x be $P(\cdot)$, with density $p(\cdot)$. Suppose he contemplates a response of $w_i = v_i(0) + \delta$. We can calculate the excess gain from the answer w_i compared to the answer $v_i(0)$. We divide the analysis into several cases according to the region in which δ lies. Assume first that $v_i(0) = v_i > 0$.

Case I

$\delta > 0$

Subcases	*excess gain*
(i) $x > 0$	0
(ii) $\begin{cases} x < 0 \\ w_i + x \geqq 0 \\ v_i + x < 0 \end{cases}$	$x + v_i(x)$
(iii) $\begin{cases} x < 0 \\ w_i + x < 0 \\ v_i + x < 0 \end{cases}$	0

Case II

$-v_i < \delta < 0$

subcases	*excess gain*
(i) $x > 0$	0
(ii) $\begin{cases} x < 0 \\ w_i + x < 0 \\ v_i + x \geqq 0 \end{cases}$	$-x - v_i(x)$
(iii) $\begin{cases} x < 0 \\ w_i + x < 0 \\ v_i + x < 0 \end{cases}$	0

Case III

$\delta < -v_i < 0$

subcases	*excess gain*
(i) $\begin{cases} x > 0 \\ w_i + x \geqq 0 \\ v_i + x \geqq 0 \end{cases}$	0
(ii) $\begin{cases} x > 0 \\ w_i + x < 0 \\ v_i + x \geqq 0 \end{cases}$	$-x - v_i(0)$
(iii) $\begin{cases} x < 0 \\ w_i + x < 0 \\ v_i + x \geqq 0 \end{cases}$	$-x - v_i(x)$
(iv) $\begin{cases} x < 0 \\ w_i + x < 0 \\ v_i + x < 0 \end{cases}$	0

We can therefore compute the expected excess gain for each of the three cases by integrating the excess gain over the indicated intervals for x.

$$\text{Case } I: \quad \int_{-v_i-\delta}^{-v_i} [x + v_i(x)]\, dP(x) \tag{5.20}$$

$$\text{Case } II: \quad \int_{-v_i}^{-v_i-\delta} (-x - v_i(x))\, dP(x) \tag{5.21}$$

$$\text{Case } III: \quad \int_{-v_i}^{0} [-x - v_i(x)]\, dP(x) + \int_{0}^{-v_i-\delta} [-x - v_i(0)]\, dP(x) \tag{5.22}$$

We present the solution to agent i's problem by optimizing with respect to δ within each interval and then finding the global optimum by comparing the values of the objective function at these candidate points. We use the approximation[1]

$$v_i(t) = t v_i'(0) + v_i(0) \tag{5.23}$$

[1] One can generalize this by taking higher-order approximations, but the general character of the results would be unaffected.

n the differentiation. We assume that $v_i'(0) > -1$. If this condition were iolated, the individual would actually benefit from giving up transfers, o this is not a real restriction. Then we will let $v_i'(0) \to 0$ in order to show he asymptotic optimality of $v_i(0)$, as desired.

In Case I we find that

$$\begin{aligned} \delta^* &= 0 && \text{if } v_i'(0) \geqq 0 \\ \delta^* &= -\frac{v_i'(0)v_i(0)}{v_i'(0)+1} && \text{if } v_i'(0) < 0. \end{aligned} \tag{5.24}$$

n Case II,

$$\begin{aligned} \delta^* &= -\frac{v_i'(0)v_i(0)}{v_i'(0)+1} && \text{if } v_i'(0) \geqq 0 \\ \delta^* &= 0 && \text{if } v_i'(0) < 0 \end{aligned} \tag{5.25}$$

nd in case III,

$$\delta^* = -v_i \tag{5.26}$$

t is easy to see that case III results in a negative gain when compared to he response of $w_i = v_i(0)$, and it can be eliminated.

For $v_i'(0) \geqq 0$, the optimized expected gain in Case I is zero, while in Case II it is

$$-\int_{-v_i(0)}^{-v_i(0)\cdot\left(\frac{1}{1+v_i'(0)}\right)} [v_i(0) + x(1 + v_i'(0))]\,\mathrm{d}P(x) \tag{5.27}$$

s $v_i'(0)$ approaches 0, this is shown to be positive, indicating that some ownward distortion from $v_i(0)$ is optimal. By the mean value theorem, or some $\lambda \in [0, 1]$ we have that (5.27) is equal to

$$-\alpha v_i(0)[v_i(0) + (1 + v_i'(0))(-v_i(0) + \lambda\alpha v_i(0))]p(-v_i(0) + \lambda\alpha v_i(0)) \tag{5.28}$$

vhere $\alpha = v_i'(0)/(1 + v_i'(0))$

This can be reduced further to

$$-\alpha^2(v_i(0))^2(1 + v_i'(0))(-1 + \lambda)p(-v_i(0) + \lambda\alpha v_i(0)) \tag{5.29}$$

vhich is positive, as desired.

By symmetry, analogous calculations yield that the optimal statements or $|v_i'(0)|$ sufficiently small are given by table 5.1.

Table 5.1

	$v_i'(0) \geqq 0$	$v_i'(0) < 0$
$v_i(0) \geqq 0$	$v_i(0)\left(1 - \frac{v_i'(0)}{v_i'(0)+1}\right)$	$v_i(0)\left(1 - \frac{v_i'(0)}{v_i'(0)+1}\right)$
$v_i(0) < 0$	$v_i(0)$	$v_i(0)$

It is of interest to note that agents opposing the project will not distort their preferences. Clearly, in every instance, $w_i \to v_i(0)$ as $v_i'(0) \to 0$.

Finally, in order to show that the decision taken is a Pareto optimum in the limit, we observe that there are two sources of error for the decision. First, $\sum_i w_i$ may differ from $\sum_i v_i(0)$ and second, $\sum_i v_i(0)$ may differ from $\sum_i v_i(t_i)$. If $v_i'(0)$ is the same for all individuals, $\sum_i w_i$ will be biased through distortions of the project's supporters. Since $\sum_i v_i(0) \to \sum_i v_i(t_i)$, because $v_i'(0)$ is small for all i and $|t_i|$ is bounded by something just slightly larger than $v_i(0)$ for each i (namely w_i), we know that for $|v_i'(0)|$ sufficiently small $\sum_i v_i(0)$ and $\sum_i v_i(t_i)$ will have the same sign unless $\sum_i v_i(0)$ is zero. Thus so will $\sum_i w_i$ and $\sum_i v_i(t_i)$, subject to the same qualification. When $\sum_i v_i(0)$ the decision is a matter of indifference so that the procedure is close to successful in any case.

There is another sense in which separability can be overcome asymptotically without assuming that taste patterns are converging. This involves the idea that the project which is considered to be continuously variable in size is to be evaluated at progressively smaller levels. Letting $u_i(K, t_i)$ be the underlying utility function of the transfer t_i, and the project's size, K,

$$v_i(K, t_i) = u_i(K, t_i) - u_i(0, t_i) \tag{5.30}$$

As $K \to 0$ we have that

$$v_i(K, t_i) \sim K\frac{\partial u_i(0, t_i)}{\partial K} \tag{5.31}$$

and

$$\frac{\partial v_i(K, t_i)}{\partial t_i} = \frac{\partial u_i(K, t_i)}{\partial t_i} - \frac{\partial u_i(0, t_i)}{\partial t_i} \sim K\frac{\partial^2 u_i(0, t_i)}{\partial K\,\partial t_i} \tag{5.32}$$

Thus, if $v_i'(t)$ is thought of at smaller and smaller levels of the project, it is converging to zero. The analysis above can therefore be applied to conclude that either the asymptotic decision taken is the correct one, or that it is virtually a matter of indifference.

5.3. Budget balance

As noted in chapter 3, the pivotal mechanism has the property that it will generally cause some taxes to be collected from the participants. Thus, despite its other desirable properties, it will generally not attain Pareto optimal allocations, for the actual N participants, ignoring the welfare of the government. Of course one could rebate these taxes in some way and thereby avoid the waste of resources that would otherwise result, but this would destroy the dominant strategy property. The consequences of doing so are explored in chapter 9.

One might hope that it is possible to use a Groves mechanism other than the pivotal mechanism that would balance the budget in all situations and would retain the property of successfulness. The goal of this section is to show that this is generally impossible. By virtue of the characterization theorem, this reduces to showing that for every collection of functions $\{h_i(w_{-i})\}$ defining a Groves mechanism, the sum of transfer payments would be non-zero for some vector of stated preferences $(w_1, \ldots, w_N)$. We will prove this result in the case of a fixed size project: $\mathscr{K} = \{0, 1\}$ to ease the notation. The results are, however, perfectly general as long as the valuation functions are unrestricted.

Theorem 5.3. There exists no Groves mechanism such that $\sum_i t_i(w) = 0$ for all $w \in \mathbf{R}^N$.

Proof. Suppose, to the contrary, that there existed a collection of functions $\{h_i(w_{-i})\}_{i=1,\ldots,N}$ such that $\sum_i t_i(w) = 0$, for all $w \in \mathbf{R}^N$. Consider first the case of $N = 2$. Let w_1^+, w_1^-, w_2 be such that

$$\begin{aligned} w_1^+ + w_2 &> 0 \\ w_1^- + w_2 &< 0 \end{aligned} \tag{5.33}$$

By the assumptions on $\{h_i(\cdot)\}$ and the definition of a Groves mechanism we have that

$$\begin{aligned} w_1^+ + w_2 + h_1(w_2) + h_2(w_1^+) &= 0 \\ h_1(w_2) + h_2(w_1^-) &= 0 \end{aligned} \tag{5.34}$$

Subtracting, we have

$$w_1^+ + h_2(w_1^+) - h_2(w_1^-) = -w_2 \tag{5.35}$$

Clearly, the left-hand side of (5.35) is independent of w_2. By varying w_2 within the set such that $w_1^+ + w_2 > 0$, to preserve the hypotheses of this

and $w_1^- + w_2 < 0$

construction, we obtain a contradiction to the invariance of the left-hand side.

We now consider $N \geqq 3$. Let $w_1^0, \ldots, w_N^0$ and $\beta > 0$ be such that

$$-(N-1)\beta < \sum_i w_i^0 < -(N-2)\beta \tag{5.36}$$

Let $h_k\,(O_{l_1}, \ldots, {}_{l_{\alpha^*}})$ be the value of the function h_k at the point where the ith argument w_i, $i \neq k$, is

$$\begin{aligned} &w_i^0 + \beta \quad &&\text{for } i \neq l_\alpha, \alpha = 1, \ldots, \alpha^* \\ &w_i^0 &&\text{for } i = l_\alpha, \alpha = 1, \ldots, \alpha^* \end{aligned}$$

Since $\sum_i w_i^0 + (N-1)\beta > 0$, we have

$$\sum_{i \neq N} h_i(O_N) + h_N(w_1^0, \ldots, w_{N-1}^0) = -(N-1)\Big[\sum_i w_i^0 + (N-1)\beta\Big] \tag{5.37}$$

$(w_1^0 + \beta, \ldots, w_{N-1}^0 + \beta)$

The idea of the proof is to make a succession of substitutions in the left-hand side of (5.37), leaving the right hand side unchanged in such a way that, at the end, the right hand side will depend on w_N^0 and the left-hand side will be independent of w_N^0. Specifically, the left-hand side will be reduced to terms involving only the function h_N, which is, of course, independent of w_N^0. Then, as in the case of $N = 2$ above, a contradiction can be derived.

These substitutions all involve the $(N-1)$ terms in the summation on the left hand side of (5.37), each of which has as its arguments one entry of the form w_k^0, for $k = N$, and $(N-2)$ entries of the form $w_k^0 + \beta$, for $k \neq N$, $k \neq i$. We first replace these with a summation of terms each of which has two entries w_k^0 and $N-3$ entries with $w_k^0 + \beta$. Note that since

$$\sum_i w_i^0 + (N-2)\beta < 0 \tag{5.38}$$

we can write, for each $j = 1, \ldots, N-1$,

$$\sum_{i \neq N} h_i(0_{j,N}) + h_N(0_j) = 0 \tag{5.39}$$

This can be written alternatively as

$$h_j(0_N) = -h_N(0_j) - \sum_{\substack{i \neq N \\ i \neq j}} h_i(O_{j,N}). \tag{5.40}$$

Substituting (5.40) into (5.37) for every $j = 1, \ldots, N-1$ we obtain a summation of terms on the left-hand side that involves $h_i(\cdot)$ evaluated at points with two entries of w_k^0, for $k = j, N$, and the other entries $w_k^0 + A$. Explicitly, (5.37) can now be written as

$$\sum_{i \neq N} \sum_{\substack{j \neq N \\ j \neq i}} h_j(O_{i,N}) = h_N(w_1^0, \ldots, w_{N-1}^0) + (N-1)\left[\sum_i w_i^0 + (N-1)\beta\right] - \sum_{i \neq N} h_N(O_i) \quad (5.41)$$

Continuing in this way we use

$$\sum_i w_i^0 + (N-3)\beta < 0 \quad (5.42)$$

to write, for each pair of indices $(j_1, j_2), j_1, j_2 \leqq N-1, j_1 \neq j_2$

$$\sum_{i \neq N} h_i(0_{j_1,j_2,N}) + h_N(0_{j_1,j_2}) = 0 \quad (5.43)$$

and rewrite this as

$$h_{j_1}(O_{j_2,N}) + h_{j_2}(O_{j_1,N}) = -h_N(O_{j_1,j_2}) - \sum_{\substack{i \neq N \\ i \neq j_1 \\ i \neq j_2}} h_i(O_{j_1,j_2,N}) \quad (5.44)$$

By summing the equations of the form (5.44) for $1 \leqq j_1 < j_2 \leqq N-1$, we obtain an equation in which the left hand side is a multiple of the left hand side in (5.41). Substituting in (5.41) we obtain a relation in which the terms involving h_i, $i \neq N$ are all of the form $h_i(O_{j_1,j_2,N})$, $1 \leqq j_1 < j_2 \leqq N-1$. Continuing in this way we arrive at a situation, finally, where all the terms in h_i, $i \neq N$, are evaluated at $O_{\iota_1,\ldots,\iota_{N-1}}$, that is at (w_j^0), $j \neq i$. Moreover their coefficients will all be the same, by the symmetry of the argument. Since $\sum_i w_i^0 < 0$, the project is rejected at the statements $(w_1^0, \ldots, w_N^0)$. Therefore,

$$\sum_{i \neq N} h_i(w_{-i}^0) = -h_N(w_{-N}^0). \quad (5.45)$$

Substituting the appropriate multiple of (5.45) we are left with a relation involving h_N, but none of the other h-functions. These terms in h_N are being evaluated at a variety of points, but of course they do not involve w_N^0 which is never an argument of h_N. This construction completes the proof since $-(N-1)[\sum_i w_i^0 + (N-1)\beta]$, does depend on w_N^0 and moreover, w_N^0 can be perturbed slightly without affecting the validity of our only assumed relation (5.36). Q.E.D.

The impossibility of finding balanced dominant-strategy mechanisms that produce optimal decisions necessitates facing the prospect of giving up either the dominant strategy property or the optimality of the decision if the exact feasibility of the mechanism is to be maintained. The extent to which this can be done in an approximate sense, without doing excessive violence to either of these desirata is the subject of chapter 9.

By placing restrictions beyond separability on the class of allowable

preference relations, it may be possible to construct balanced Groves mechanisms. Groves and Loeb [1975] have given an example of this type of construction in the special case where all individuals have quadratic utility functions over the project space $\mathscr{K} = \mathbf{R}_+$ of the form,

$$v_i(K) = -\tfrac{1}{2}K^2 + \theta_i K \qquad i = 1, \ldots, N \quad (5.46)$$

where θ_i is any non-negative real number (see section 4.4, in particular, theorem 4.5).

Let $\theta = (\theta_1, \ldots, \theta_N)$ and let $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_N)$. As shown in the proof of theorem 4.5, honest preference revelation implies, for each $i = 1, \ldots, N$,

$$t_i(\theta_i, \theta_{-i}) = \tfrac{1}{2}\left(\frac{1}{N^2} - \frac{1}{N}\right)\theta_i^2 + \frac{1}{N^2}\theta_i \sum_{j \neq i} \theta_j + h_i'(\theta_{-i}) \quad (5.47)$$

where $h_i'(\theta_{-i})$ are arbitrary functions of θ_{-i}.

We know, by virtue of theorem 4.5 that,

$$t_i(\theta_i, \theta_{-i}) = \sum_{j \neq i} w_j(K^*(\theta)) + h_i(\theta_{-i}) \qquad i = 1, \ldots, N \quad (5.48)$$

for some set of functions $h_i(\theta_{-i})$, $i = 1, \ldots, N$.

Here, we have

$$\begin{aligned}\sum_{j \neq i} w_j(K^*(\theta)) &= -\frac{N-1}{2}[K^*(\theta)]^2 + \sum_{j \neq i} \theta_j K^*(\theta) \\ &= \frac{1}{N} \sum_{j \neq i} \theta_j \cdot \left(\sum_k \theta_k\right) - \frac{N-1}{2N^2}\left(\sum_l \theta_l\right)^2.\end{aligned} \quad (5.49)$$

Rewriting (5.49), separating out the terms in θ_i and θ_i^2, we have

$$\begin{aligned}\sum_{j \neq i} w_j(K^*(\theta)) = &-\frac{N-1}{2N^2}\theta_i^2 + \frac{1}{N^2}\theta_i \sum_{j \neq i} \theta_j \\ &+ \frac{1}{N}\left(\sum_{j \neq i} \theta_j\right)^2 - \frac{N-1}{2N^2}\left(\sum_{j \neq i} \theta_j\right)^2\end{aligned} \quad (5.50)$$

Combining (5.50), (5.47) and (5.48) we can write $h_i(\theta_{-i})$ as

$$h_i(\theta_{-i}) = -\left(\frac{1}{2N^2} + \frac{1}{2N}\right)\left(\sum_{j \neq i} \theta_j\right)^2 + h_i'(\theta_{-i}) \quad (5.51)$$

We want to choose the arbitrary functions $h_i'(\theta_{-i})$ so that the mechanism is balanced. That is, we want

$$0 \equiv \sum_i t_i(\theta_i, \theta_{-i}) = \tfrac{1}{2}\left(\frac{1}{N^2} - \frac{1}{N}\right)\sum_i \theta_i^2 + \frac{1}{N^2}\sum_l \theta_l \sum_{j \neq l} \theta_j + \sum_i h_i'(\theta_{-i}) \quad (5.52)$$

Let us seek a solution to (5.52) such that $h_i'(\theta_{-i}) = h'(\theta_{-i})$ for all i. This type of symmetric solution will be of interest because it means that a balanced mechanism can be found that treats individuals in an anonymous way, and anonymity is regarded as a desirable property of social choice procedures. And because $h'(\theta_{-i})$ is not a function of θ_i, we know that it can be written as the sum two types of terms: those in θ_j^2 for $j \neq i$ and those in $\theta_j\theta_l$ for $j \neq l, j \neq i, l \neq i$. By symmetry, the weights of the terms within each of these types must be equal. Hence

$$h'(\theta_{-i}) = \alpha \sum_{j \neq i} \theta_j^2 + \beta \sum_{\substack{j \neq i \\ l \neq i \\ j \neq l}} \theta_j \theta_l \tag{5.53}$$

To find α and β we sum (5.53) and use the identity (5.52):

$$\begin{aligned} \sum_i h'(\theta_{-i}) &= \alpha \sum_i \sum_{j \neq i} \theta_j^2 + \beta \sum_i \sum_{\substack{j \neq i \\ l \neq i \\ j \neq l}} \theta_j \theta_l \\ &= (N-1)\alpha \sum_i \theta_i^2 + (N-2)\beta \sum_l \sum_{j \neq l} \theta_j \theta_l \end{aligned} \tag{5.54}$$

Identifying the coefficients on the right hand side of (5.54) with the corresponding terms in (5.52) we have

$$\begin{aligned} \alpha &= \frac{1}{2N^2} \\ \beta &= \frac{-1}{N^2(N-2)} \end{aligned} \tag{5.55}$$

Thus we obtain, finally, the expression for the functions $h_i(\theta_{-i})$ defining a balanced Groves mechanism for this class of environments by using (5.51) and (5.55):

$$h_i(\theta_{-i}) = -\frac{1}{2N^2} \sum_{j \neq i} \theta_j^2 - \frac{1}{N^2(N-2)} \sum_{\substack{j \neq i \\ l \neq i \\ j \neq l}} \theta_j \theta_l. \tag{5.56}$$

Note that this mechanism is well defined only in the case $N \geqq 3$. We can show that it is impossible to find a balanced, *SIIC*, successful mechanism for the case of $N = 2$ even with quadratic utility.[2] Write the conditions for *SIIC* as

[2] Hurwicz [1972] has given a general impossibility theorem for the $N = 2$ case using a slightly different class of environments.

$$\begin{aligned} \frac{\partial t_1}{\partial \theta_1} &= -\frac{\partial v_1}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_1} \\ \frac{\partial t_2}{\partial \theta_2} &= -\frac{\partial v_2}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_2} \end{aligned} \tag{5.57}$$

Balancedness implies

$$\frac{\partial t_1}{\partial \theta_1} + \frac{\partial t_2}{\partial \theta_1} = 0 \tag{5.58}$$

and

$$\frac{\partial t_1}{\partial \theta_2} + \frac{\partial t_2}{\partial \theta_2} = 0$$

From (5.57) and (5.58) we obtain

$$\begin{aligned} \frac{\partial t_1}{\partial \theta_1} &= \frac{-\partial v_1}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_1} \\ \frac{\partial t_1}{\partial \theta_2} &= \frac{+\partial v_2}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_2} \end{aligned} \tag{5.59}$$

which in the quadratic case reduces to

$$\frac{\partial t_1}{\partial \theta_1} = \frac{(\theta_2 - \theta_1)}{2}$$

and

$$\frac{\partial t_1}{\partial \theta_2} = \frac{(\theta_2 - \theta_1)}{2} \tag{5.60}$$

Clearly $t_1(\theta_1, \theta_2)$ does not satisfy the Poincaré condition for integrability, $\partial^2 t_1/\partial\theta_1\partial\theta_2 = \partial^2 t_1/\partial\theta_2\partial\theta_1$, and thus there is no differentiable function $t_1(\theta_1, \theta_2)$, yielding an *SIIC* and successful, balanced mechanism for quadratic environments, when $N = 2$.

To conclude this section we give a necessary and sufficient condition for a class of valuation functions to admit a balanced Groves mechanism.

Let us assume for $i = 1, \ldots, N$ that Θ_i is an open interval in $\mathbf{R}$ and v_i: $\mathring{\mathbf{R}}_+ \times \Theta_i \to \mathbf{R}$ is an $N+1$ differentiable function such that for any $\theta \in \Theta = \Pi_i \Theta_i$, there exists $K^*(\theta) \in \mathring{\mathbf{R}}_+$, for which

(i) $\sum_i v_i(K^*(\theta), \theta_i) = \max_{K>0} \sum_i v_i(K, \theta_i)$

(ii) $K^*(\theta)$ is continuously differentiable.

Theorem 5.4. Under assumption 1, there exists a balanced satisfactory mechanism for the class of valuation functions $V = \{v_1(\cdot, \theta_1), \ldots, v_N(\cdot, \theta_N) \mid \theta \in \Theta\}$, if and only if,

$$\sum_i \frac{\partial^{N-1}}{\partial \theta_{-i}} \left[\frac{\partial v_i}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_i} \right] \equiv 0.^3 \tag{5.61}$$

Proof. Suppose there exists a balanced satisfactory mechanism with respect to V. Then,

$$\sum_i t_i(\theta) \equiv 0$$

or from section 4.4

$$\sum_i \left[- \int \frac{\partial v_i}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_i} \mathrm{d}\theta_i + h_i(\theta_{-i}) \right] \equiv 0 \tag{5.62}$$

Differentiating (5.62) with respect to $\theta_1, \ldots, \theta_N$, we obtain

$$\sum_i \frac{\partial^{N-1}}{\partial \theta_{-i}} \cdot \frac{\partial v_i}{\partial K} \cdot \frac{\partial K^*(\theta)}{\partial \theta_i} \equiv 0 \tag{5.63}$$

which establishes necessity.

Reintegrating (5.63) successively with respect to $\theta_1, \ldots, \theta_N$ regenerates (5.61). Hence sufficiency. Q.E.D.

The interest of theorem 5.4 is to permit a direct check of the possibility of balancedness for a given class of valuation functions without actually attempting to construct the transfers. It is for example easy to verify the nonexistence of balanced Groves mechanisms for the admissible classes

$$V_1 = \{\theta_i K - \log K, \theta_i \in \Theta_i\}$$
$$V_2 = \{\theta_i \log K - K, \theta_i \in \Theta_i\}$$
$$V_3 = \{\theta_i e^{-K} + K, \theta_i \in \Theta_i\}$$
$$V_4 = \{\theta_i \log K - K^2, \theta_i \in \Theta_i\}$$
$$V_5 = \left\{\theta_i K - \frac{K^3}{3}, \theta_i \in \Theta_i\right\} \quad \text{etc.}$$

and to verify the existence of a balanced Groves mechanism for the quadratic class exhibited above (for $N > 2$).

[3] $\dfrac{\partial^{n-1}}{\partial \theta_{-i}} = \dfrac{\partial^{n-1}}{\partial \theta_1, \ldots, \partial \theta_{i-1}, \partial \theta_{i+1}, \ldots, \partial \theta_N}$.

5.4. Lower boundedness

One of the most troublesome features of Groves mechanisms is the potential for the reduction of welfare of an individual vis à vis the null project, if this is the appropriate benchmark level to use. A fuller discussion of the individual rationality question will be taken up in the next chapter. For the present we will concentrate not on the level of utility attained but rather on the individual's level of private goods before and after the employment of the mechanism. We have not emphasized the initial situation before because of the assumption that the transfers could be of any magnitude, positive or negative. The initial ownership of this resource would add a constant to all utility functions and this could justifiably be ignored.

It is reasonable to assume, however, that the agents may rather have limited means. That is, initial endowments are non-negative numbers $\bar{x}_i$, $i = 1, \ldots, N$, and transfers must satisfy

$$t_i \geqq -\bar{x}_i \qquad \text{for } i = 1, \ldots, N \tag{5.65}$$

for feasibility to be insured.

Two remarks are in order: First, there is a slight abuse of language now in continuing to call $v_i(K)$ the "willingness-to-pay" of individual i for project K, vis à vis the null project. It does describe the fact that he would be indifferent between receiving $v_i(K)$ units of the private good with the null project accepted and having project K accepted, with no transfer taking place. However it does not mean that $v_i(K)$ can be taken away from the agent in return for project K being accepted because he may not have that much endowment. He may be willing, but not able, to pay. Having brought out the asymmetry of jargon in use for this case, we will proceed with calling it "willingness-to-pay".

The second remark is more to the heart of the matter at hand. Without bounded consumptions, all Pareto optima in the economy with separable utility functions can be found by maximizing $\sum_i v_i(\cdot)$ and selecting $(t_1, \ldots, t_N)$ arbitrarily, as observed above. With consumption bounded below, however, there may be some Pareto optimal (K, t) combinations such that $\sum_i v_i(K)$ is not maximal. These will occur when some agents are at the boundary of their consumption sets: $t_i = -\bar{x}_i$. Therefore, direct revelation mechanisms would not necessarily select a feasible point if they proceeded to maximize $\sum_i w_i(\cdot)$ as usual, and attain the required utility levels for the agents in the economy. An example of this follows:

Example 5.1. Let $\mathscr{K} = \{0, 1\}$; the project is to be either accepted, $K = 1$ or rejected, $K = 0$. For simplicity we normalize by setting $v_i(0) = 0$.

There are two individuals, $i = 1, 2$, whose characteristics are described by

$$\begin{aligned} \bar{x}_1 &= 0, \quad v_1 = 5 \\ \bar{x}_2 &= 10, \quad v_2 = -1 \end{aligned} \tag{5.66}$$

Consider the social state described by the initial resource allocation and rejection of the project. This gives utilities of 0 and 10 to the individuals 1 and 2 respectively and the net revenue generated by the process is zero. Note that $\sum_i v_i = 4$ and hence the sum of the willingnesses-to-pay would indicate that the project should be accepted. However, consistent with net revenues at a zero level, we cannot find a Pareto superior allocation and decision. To see this, we think of maximizing $u_2(\text{d}, x_2)$ subject to $u_1(\text{d}, x_1) \geqq 0$ and $x_1 + x_2 \leqq 10$. If the project is rejected, then $x_1 = 0$ is necessary for if $x_1 > 0$, then $x_2 < 10$ and hence individual 2's utility is not being maximized subject to these constraints. If the project is accepted, then $u_1(\text{d}, x_1) \geqq 5$ for all feasible values of x_1. Hence although $x_2 = 10$, maximizes $u_2(\text{d}, x_2)$ given the constraints, we have $u_2(\text{d}, 10) = 9$ whenever the project is accepted. Therefore there is no Pareto superior allocation to the initial allocation and rejection of the project. Without consumption bounds, of course, acceptance of the project would allow the initial allocation to be dominated, without violating the net revenue constraint.

In light of this example, the concept of successfulness of the mechanism must be modified if we are to study the means for attaining Pareto optima, when consumption bounds are taken into account. In particular, explicit attention must now be paid to the level of revenues collected by the mechanism.

Definition 5.2. A *feasible social state* is:

$$a = (K, x_1, \dots, x_N, \sum_i (\bar{x}_i - x_i))$$

such that $x_i \geqq 0$ for all i, and $K \in \mathscr{K}$.

The last entry represents the net transfer from the individuals to the decision-maker. It is natural to suppose that the utility of the decision-maker is strictly increasing in this quantity, as revenues are usable for other purposes or may be substituted for other taxes being collected.[4] Without

[4] Of course, if individuals were to recognize this dependence, the favorable incentive properties of the mechanisms to be studied would be destroyed – see, however, chapter 9 for approximation results in this regard.

fear of confusion we can write the utility of individual i in social state a as $u_i(a) = u_i(K, x_i)$.

The utility of the decision-maker is then:

$$u_0(a) = \sum_i (\bar{x}_i - x_i) \tag{5.67}$$

The standard definition of Pareto optimality for the $N + 1$ agents $i = 0, 1, \dots, N$ coincides with the definition we have been using, (see Definition 3.1).

The principle difference between mechanisms as presently studied and those used in chapters 3 and 4 is that the transfers induced by the mechanism must be compatible with the existing endowment distribution in the sense that the final allocations of the private good must all be non-negative. This can be interpreted in either of two ways. If endowments were observable by the planner in advance of his choice of a specific mechanism, he might be able to arrange the mechanism so as to insure that a non-negative consumption level would result. This context is not without interest for it may well be the case that endowments are real physical entities whereas preferences are a more elusive concept. Nevertheless, taking this view of the construction of the mechanism would set up incentives[5] for the disguising of endowments in order to influence the mechanism selected. Another difficulty with this idea is that even if endowments cannot be misrepresented, it may not be feasible for the planner to arrange the mechanism in a way that does not involve some risk of a negative consumption level, unless further restrictions can be placed on preferences. Consider any Groves mechanism. We may think of this as the pivotal mechanism plus some function of the others' statements. The agents' pivotal payments are bounded above by the absolute value of the sum of these statements. In order to guarantee the non-negativity of consumptions, it will be necessary to choose the additional function so that it is bounded below by this quantity. Such a mechanism would place a heavy burden on the resources of the planner. Not having bounded his transfer below, we can use this method to insure non-negative transfers for the other agents – but if the resources at his command are limited as well, it may not be possible to create satisfactory mechanisms.

The other approach to this problem is to fix a mechanism in the absence of knowledge about endowments. One can rely on the individuals themselves to restrict their strategies in such a way as to guarantee non-negativity

[5] See chapter 4, section 4.6.4, where the unobservable characteristic may be interpreted as the endowment itself.

of their consumptions. The mechanism chosen must then be such that it is always possible for each agent, regardless of his endowment, to choose some strategy that will always lead to a feasible outcome. This behavioral assumption – that individuals will never choose a strategy for which their outcome function is not well-defined – forms the basis of our analysis in this section. The methodology of having to select a mechanism before any information on the endowment is known is in keeping with the spirit of our overall framework; and our results focus on the implications of these restrictions on consumption, for the possibility of overcoming adverse incentives in the decision making process.

Consider a Groves mechanism (S, f).

Definition 5.3. A strategy w_i is called *admissible* if, for every w_{-i}, $f(w_{-i}, w_i) = (K, x_1, \ldots, x_N, \sum_i(\bar{x}_i - x_i))$ is such that $x_i \geqq 0$.

The set of all strategies admissible for agent i is denoted S_i^a.

We can either view the mechanism as being defined directly on the admissible sets, if endowments are observable, or we can assume that individuals restrict themselves to this set of strategies. In either way the mechanism specifies a function f from $\mathbf{R}^N$ into the social states, and a set of subsets of $\mathbf{R}^N$, S_i^a, $i = 1, \ldots, N$, which are the admissible strategy spaces under the mechanism f. The agents will not need to compare their selection to any of the stated willingness-to-pay corresponding to inadmissible strategies. This suggests the following definition for dominance in the present context.

Definition 5.4. An admissible strategy w_i, is called *admissible dominant* (or just *dominant* for brevity) if, for every other admissible strategy, $w_i' \in S_i^a$

$$u_i(f(w_{-i}, w_i)) \geqq u_i(f(w_{-i}, w_i'))$$

for all w_{-i}.

Note, however, that the strategy $w_i = v_i$ may not be admissible for some mechanisms and some individuals. There may nevertheless exist other dominant strategies.

Definition 5.5. A mechanism will be called *admissible satisfactory* (or just *satisfactory*, for brevity) if

(i) for every i, there exists an admissible dominant strategy, w_i^*

and

(ii) $f(w^*) = f(w_1^*, \ldots, w_N^*)$ is a Pareto optimum.

The goal of this section is to study the class of mechanisms that are satisfactory in the presence of consumption bounds. It will be shown that some, but not all, of the mechanisms of the Groves class remain satisfactory in this context as well.

Let us consider the special case in which, for all i,

$$h_i(w_{-i}) = \min\left(-\sum_{j \neq i} w_j, 0\right). \tag{5.68}$$

the *pivotal mechanism*. Under this mechanism we always have $x_i \leqq \bar{x}_i$. The strategy w_i is pivotal when $w_i \sum_{j \neq i} w_j < 0$ and $|w_i| > |\sum_{j \neq i} w_j|$, because the presence of individual i changes the social decision.

Theorem 5.5. For the mechanism defined by (5.68), the strategies

$$\begin{aligned} \hat{w}_i(\bar{x}_i, v_i) &= v_i - \max(v_i - \bar{x}_i, 0) \qquad v_i \geqq 0 \\ &= v_i - \min(v_i + \bar{x}_i, 0) \qquad v_i < 0 \end{aligned} \tag{5.69}$$

are unique dominant strategies.

Proof. Consider first the case in which $v_i \geqq 0$. Note that $u_i(w_{-i}, \hat{w}_i)$ is defined for all w_{-i}. However, for $|\tilde{w}_i| > \bar{x}_i$, $u_i(w_{-i}, \tilde{w}_i)$ is not defined whenever w_{-i} is such that $|\tilde{w}_i| > |\sum_{j \neq i} w_j| > \bar{x}_i$ and $\tilde{w}_i \cdot \sum_{j \neq i} w_j < 0$ because then $x_i = \bar{x}_i - |\sum_{j \neq i} w_j| < 0$.

For $|\tilde{w}_i| < \bar{x}_i$, $u_i(w_{-i}, \tilde{w}_i)$ is defined for all w_{-i}, and hence these values of $\tilde{w}_i$ are admissible strategies.

If $v_i \leqq \bar{x}_i$, then $\hat{w}_i = v_i$ and theorem 4.1 establishes the dominance of $\hat{w}_i$.

If $v_i > \bar{x}_i$, so that $\hat{w}_i = \bar{x}_i$, and $\tilde{w}_i$ is any other admissible strategy, we have the following cases:

The value of w_{-i} is such that:

(i) $\hat{w}_i$ is pivotal but $\tilde{w}_i$ is not

(ii) $\hat{w}_i$ and $\tilde{w}_i$ are both pivotal

(iii) neither $\hat{w}_i$ nor $\tilde{w}_i$ are pivotal.

Case (i)

Since $\hat{w}_i = \bar{x}_i > 0$ is pivotal, it must be that $\sum_{j \neq i} w_j < 0$, and $\bar{x}_i > -\sum_{j \neq i} w_j$. Thus:

$$u_i(f(w_{-i}, \hat{w}_i)) = v_i + \bar{x}_i + \sum_{j \neq i} w_j \tag{5.70}$$

and

$$u_i(f(w_{-i}, \tilde{w}_i)) = \bar{x}_i \tag{5.71}$$

Using $v_i > \overline{x}_i > -\sum_{j\neq i} w_j$, the former outcome dominates the latter.

Case (ii)

We must have $\hat{w}_i \geqq \tilde{w}_i > 0$, $\sum_{j\neq i} w_j < 0$, $w_i > -\sum_{j\neq i} w_j$.

Therefore the outcome associated with $(w_{-i}, \hat{w}_i)$ is the same as that associated with $(w_{-i}, \tilde{w}_i)$.

Case (iii)

Similarly, the two outcomes are identical.

The case of $v_i < 0$ is analogous and therefore $\hat{w}_i$ is the unique dominant strategy. Q.E.D.

Up to now we have not had to analyze the method by which a decision is taken in the event that the project that maximizes stated willingness to pay is not determined uniquely. Without restrictions on consumption sets, the chosen project, however selected, would result in a Pareto optimum. The following example shows that this is no longer the case in the present context.

Example 5.2. We use the pivotal mechanism (5.68) and consider the environment

$$\begin{aligned} \overline{x}_1 &= 5 \qquad v_1 = -6 \\ \overline{x}_2 &= 10 \qquad v_2 = +5 \end{aligned} \tag{5.72}$$

By virtue of theorem 5.4, the stated preferences would be $w_1 = -5$, $w_2 = +5$. The project would be accepted, if we were to follow our usual tie-breaking convention. This would produce the social state

$$a = (1, 5, 5, 5) \tag{5.73}$$

and associated utilities -1, 10, 5 for the two agents and the government, respectively.

But considering the alternative social state

$$a' = (0, 0, 10, 5) \tag{5.74}$$

which has associated utilities 0, 10 and 5, we can see that a is not a Pareto optimum. Of course if the tie had been broken in the opposite way, a' would have been precisely the social state produced. Since a' is optimal, perhaps the mechanism is not so bad, after all.

This discussion prompts us to modify the definition of satisfactoriness for the present discussion in the obvious way.

Definition 5.6. A direct revelation mechanism will be called *essentially satisfactory* if either

(i) it is satisfactory

or

(ii) it fails to be satisfactory at some environments $(\overline{x}, v)$ where for the dominant admissible strategies, w^*, $d(w^*)$ is not the unique maximizer of $\sum_i w_i(\cdot)$ over $\mathscr{K}$.

Theorem 5.6. The pivotal mechanism is essentially satisfactory.

Proof. By virtue of theorem 5.5 it suffices to prove that $f(\hat{w}_1, \ldots, \hat{w}_N)$ is Pareto optimal.

Let

$$f(\hat{w}_1, \ldots, \hat{w}_N) = (K, x_1, \ldots, x_N, \sum_i (\overline{x}_i - x_i)) \tag{5.75}$$

Consider any feasible social state $a' = (K', x'_1, \ldots, x'_N, \sum_i(\overline{x}_i - x'_i))$. If $K' = K$ then it is clear that we cannot have $x'_i \geqq x_i$ and $u'_0 > u_0$ without having $x'_i = x_i$ for all i. Thus if $(\hat{w}_1, \ldots, \hat{w}_N)$ can be dominated, it must be by a social state such that $K' \neq K$.

Consider first the case of $K = 0$:

Let

$$\begin{aligned} I_+ &= \{i \mid 0 \leqq \hat{w}_i = v_i \leqq \overline{x}_i\} \\ I_- &= \{i \mid 0 \geqq \hat{w}_i = v_i \text{ and } |v_i| \leqq \overline{x}_i\} \\ J_+ &= \{i \mid 0 \leqq \hat{w}_i = \overline{x}_i < v_i\} \\ J_- &= \{i \mid 0 \geqq \hat{w}_i = -\overline{x}_i > v_i\} \end{aligned} \tag{5.76}$$

Let

$$t_i = \overline{x}_i - x_i, \qquad i = 1, \ldots, N \tag{5.77}$$

be the transfer under the pivotal mechanism.

In order that a' Pareto dominates $f(\hat{w}_1, \ldots, \hat{w}_N)$, we require that for all $i = 1, \ldots, N$,

$$v_i + x'_i \geqq x_i \tag{5.78}$$

or

$$x'_i \geqq x_i - v_i = \overline{x}_i - t_i - v_i \tag{5.79}$$

For the four groups of individuals above this means

$$\begin{aligned} i \in I_+ &: \; x_i' \geqq \bar{x}_i - t_i - \hat{w}_i \\ i \in I_- &: \; x_i' \geqq \bar{x}_i - t_i - \hat{w}_i \\ i \in J_+ &: \; x_i' \geqq \bar{x}_i - t_i - v_i \\ i \in J_- &: \; x_i' \geqq \bar{x}_i - t_i - v_i \geqq \bar{x}_i - t_i - \hat{w}_i \end{aligned} \tag{5.80}$$

Note that for $i \in J_+$, the relevant constraint is really $x_i' \geqq 0$ since $\bar{x}_i < v_i$ and $-t_i < 0$, so that the right hand side of the inequality above is negative.

Therefore the minimal amount of the private good necessary to sustain a Pareto superior point is the sum of the terms in (5.80) plus the government's revenue, $-\sum_i t_i$. This quantity, denoted R, can be written as

$$\begin{aligned} R &= \sum_{I_+}(\bar{x}_i - t_i - \hat{w}_i) + \sum_{I_-}(\bar{x}_i - t_i - \hat{w}_i) + \sum_{J_+} 0 + \sum_{J_-}(\bar{x}_i - t_i - \hat{w}_i) - \sum_i t_i \\ &= \sum_{I_+}\bar{x}_i + \sum_{I_-}\bar{x}_i + \sum_{J_-}\bar{x}_i - \sum_{J_+} t_i - \sum_{I_+ \cup I_- \cup J_-} \hat{w}_i \end{aligned} \tag{5.81}$$

Now, since $K = 0$ was the decision with strategies $\hat{w}_i$, $i = 1, \dots, N$, we have

$$0 > \sum_i \hat{w}_i = \sum_{I_+ \cup I_- \cup J_-} \hat{w}_i + \sum_{J_+} \hat{w}_i \tag{5.82}$$

or

$$-\sum_{I_+ \cup I_- \cup J_-} \hat{w}_i > \sum_{J_+} \hat{w}_i = \sum_{J_+} \bar{x}_i \tag{5.83}$$

Substituting (5.83) into (5.81), we have

$$R > \sum_i \bar{x}_i - \sum_{J_+} t_i \geqq \sum_i \bar{x}_i \tag{5.84}$$

since $t_i \leqq 0$ for all i, under the pivotal mechanism.

This inequality establishes the infeasibility of a Pareto superior allocation with $K = 0$.

In the case of $K = 1$ with $\sum_i \hat{w}_i > 0$ the proof is analogous. When $\sum_i \hat{w}_i = 0$, we may not have satisfactoriness, but by definition, the mechanism is essentially satisfactory. Q.E.D.

In order to discuss the results above in the context of other mechanisms, we proceed in two stages. We ascertain the class of mechanisms for which dominant strategies exist. Then, we study those which produce Pareto optimal results.

For simplicity we concentrate on the symmetric case, where $h_i(\cdot)$ is the same function for all i and is invariant with respect to permutations of the components of w_{-i}.[6] We denote this common function by $h(\cdot)$. In order

[6] None of our results depend on this restriction, which is used only for simplicity of notation.

that dominant strategies exist for the mechanism defined by h, it is first necessary that it have a non-empty set of admissible strategies for all possible levels of endowment. We can define

$$t_{-i}(h, w_i) = \min_{w_{-i}} t_i(w_{-i}, w_i) \tag{5.85}$$

where $t_i(w_{-i}, w_i)$ is the transfer received by the agent, under the mechanism defined by h. Since $\bar{x}_i > 0$, it is required that

$$\max_{w_i} t_{-i}(h, w_i) \geqq 0 \tag{5.86}$$

or that $\min_{w_{-i}} t_i(w_{-i}, w_i) \geqq 0$ for some choice of w_i.

Theorem 5.7. If for some w_{-i}, $h(w_{-i}) < \min(-\sum_{j \neq i} w_j, 0)$ then the mechanism defined by $h(\cdot)$ has no admissible strategies for $\bar{x}_i$ sufficiently small.

Proof. The proof is immediate for if $h(\bar{w}_i) < \min(-\sum_{j \neq i} \bar{w}_j, 0)$ then $t_i(w_i, \bar{w}_i)$ takes on one of the two values $\sum_{j \neq i} w_j + h(\bar{w}_{-i})$ or $h(\bar{w}_{-i})$ according to the sign of $\sum_i w_i$, and each of these is strictly negative. Q.E.D.

For $h(\cdot)$ satisfying $h(w_{-i}) \geqq \min(-\sum_{j \neq i} w_j, 0)$ admissible strategies exist. Before we analyze the dominance of one strategy within this set, we characterize the admissible set according to the following:

Theorem 5.8. The set of admissible strategies for each $\bar{x}_i > 0$ is an interval. Moreover,

(i) it is a closed interval if $h(\cdot)$ is continuous
(ii) it contains zero if $h(\cdot) > \min(-\sum_{j \neq i} w_j, 0)$.

Proof. Suppose w_i and w_i' are admissible strategies and $w_i > w_i'$. The following inequalities must then hold:

$$\begin{aligned} &\text{(i)} && h(w_{-i}) \geqq -\bar{x}_i && \text{for } \sum_{j \neq i} w_j < -w_i \\ &\text{(ii)} && h(w_{-i}) \geqq -\sum_{j \neq i} w_j - \bar{x}_i && \text{for } \sum_{j \neq i} w_j \geqq -w_i \\ &\text{(iii)} && h(w_{-i}) \geqq -\bar{x}_i && \text{for } \sum_{j \neq i} w_j < -w_i' \\ &\text{(iv)} && h(w_{-i}) \geqq -\sum_{j \neq i} w_j - \bar{x}_i && \text{for } \sum_{j \neq i} w_j \geqq -w_i' \end{aligned} \tag{5.87}$$

Since (iii) implies (i) and (ii) implies (iv), then for w_i'' such that $w_i' < w_i'' < w_i$ we will have that $h(w_{-i}) \geqq \sum_{j \neq i} w_j - \bar{x}_i$ for $\sum_{j \neq i} w_j \geqq -w_i$ and therefore for $\sum_{j \neq i} w_j \geqq -w_i''$. When $\sum_{j \neq i} w_j < -w_i'$ and therefore $\sum_{j \neq i} w_j < -w_i''$, we have $h(w_{-i}) \geqq -\bar{x}_i$.

The closedness of the interval of admissible strategies follows from the closedness of the requirement $x_i \geqq 0$ and the continuity of $h(\cdot)$.

The admissibility of zero for the function $h(\cdot) = \min(-\sum_{j\neq i} w_j, 0)$ follows from the definition. Q.E.D.

This characterizes the admissible strategies for every endowment. Because of their simple form, one can find the dominant admissible strategy directly. It will be the closest point to the true willingness-to-pay within the admissible set. From now on, we will assume that $h(\cdot)$ is continuous.

Theorem 5.9. Let $[-\alpha, \beta]$ be the interval of admissible strategies for an individual with endowment $\overline{x}_i$ and preferences v_i, when the mechanism is defined by $h(\cdot)$.

Let w_i^* solve $\min_{w_i \in [-\alpha, \beta]} |v_i - w_i|$. Then w_i^* is the dominant strategy.

Proof. Take the case $v_i < -\alpha < \beta$ and suppose $-\alpha < w_i < \beta$. Consider

$$\begin{aligned} u_i(f(-\alpha, w_{-i})) - u_i(f(w_i, w_{-i})) &= 0 && \text{if } \sum_{j\neq i} w_j \notin [-w_i, \alpha] \\ &= -v_i - \sum_{j\neq i} w_j && \text{if } \sum_{j\neq i} w_j \in [-w_i, \alpha] \end{aligned} \tag{5.88}$$

but $v_i < -\alpha$ and $\sum_{j\neq i} w_j < \alpha$ implies $-v_i - \sum_{j\neq i} w_j > 0$ and hence setting $w_i = -\alpha$ dominates any other admissible strategy. Other cases can be treated analogously. Q.E.D.

We have seen that we must place some restrictions on the arbitrary h – it must be continuous and bounded below by the pivotal mechanism – in order to insure the existence of a dominant admissible strategy for all agents. It is nevertheless still not true that the class of all mechanisms defined by functions h with these characteristics are satisfactory. That is, there are some dominant strategy mechanisms which may, in some cases, produce non-Pareto optimal outcomes. More strikingly, these instances of non-optimality are not due to tie-breaking procedures as in the example presented above. Therefore, these mechanisms are not even essentially satisfactory. An instance of this is the following:

Example 5.3. There are two individuals. For each i, the function $h(w_j)$ is defined by

$$\begin{aligned} h(w_j) &= -2(w_j + 1) && w_j \leqq -1 \\ &= 0 && w_j > -1 \end{aligned} \tag{5.89}$$

This function clearly satisfies our requirements. One can verify that the set of admissible strategies for each agent i, S_i^a, as it depends on his endowment, $\overline{x}_i$, is given by

$$\begin{aligned} S_i^a(\overline{x}_i) &= (-\infty, +\overline{x}_i] && \text{if } \overline{x}_i < 1 \\ &\quad (-\infty, +\infty) && \text{if } \overline{x}_i \geqq 1 \end{aligned} \tag{5.90}$$

Let

$$\begin{aligned} \overline{x}_1 &= 0. & v_1 &= 5 \\ \overline{x}_2 &= 10, & v_2 &= -3 \end{aligned} \tag{5.91}$$

Thus, $w_1 = 0$ and $w_2 = -3$ will be the dominant admissible strategies of the agents. The project will be rejected and a transfer of $+4$ will be given to agent 1. The social state attained is therefore $a = (0, 4, 10, -4)$ and the utilities attained are just the indicated consumption levels. However, consider the alternative social state $a = (1, 0, 14, -4)$. The utilities are now 5, 11, and -4 for the two agents and the government respectively. This clearly dominates the equilibrium attained by the mechanism.

It is therefore clear that only some of the satisfactory mechanisms for economies with unrestricted transfers of the private good continue to have this property when the non-negativity of consumption is required. We can then ask, naturally, for the class of mechanisms that are satisfactory or essentially satisfactory in this case.

Conditions sufficient to insure the successfulness of a dominant-strategy inducing mechanism are not hard to derive. It does not seem possible, however, to state a necessary and sufficient condition on the function h, in a readily interpretable form.

By virtue of the theorems above, the end-points of the interval defining the admissible strategies are both at least equal to the level of endowment in absolute value. If h is above the pivotal h-function, then strategies that exceed $\overline{x}_i$ in absolute value are allowable. Summarizing the analysis above, we can conclude:

Theorem 5.10. A sufficient condition that a mechanism be essentially satisfactory when consumption is bounded below by zero is that it never provide any individual a subsidy that is greater than the absolute value of the difference between his strategy and his endowment.

It may be seen that the pivotal mechanism, which never provides any subsidies, satisfies this criterion.

Proof. Using the notation paralleling that of theorem 5.6, let

$$\begin{aligned} I_+ &= \{i \mid 0 \leqq w_i = v_i\} \\ I_- &= \{i \mid 0 \geqq w_i = v_i\} \\ J_+ &= \{i \mid 0 \leqq w_i < v_i\} \\ J_- &= \{i \mid 0 \geqq w_i > v_i\} \end{aligned} \tag{5.92}$$

Suppose the strategies $(w_1^*, \ldots, w_N^*)$ are played and the mechanism rejects the project, resulting in the social state

$$a = (0, x_1^*, \ldots, x_N^*, \sum_i \overline{x}_i - \sum_i x_i^*) \tag{5.93}$$

and that

$$a' = (1, x_1', \ldots, x_N', \sum_i \overline{x}_i - \sum_i x_i') \tag{5.94}$$

were Pareto superior. This would require:

$$\begin{aligned} &\sum_i x_i' \leqq \sum_i x_i^* \\ &x_i' \geqq x_i^* - v_i \\ &x_i' \geqq 0 \end{aligned} \tag{5.95}$$

for each i. Hence the amount of the private good, R, necessary to sustain a' would have to satisfy

$$\begin{aligned} R &= \sum_i x_i' + \sum_i \overline{x}_i - \sum_i x_i^* \\ &= \sum_{I_+ \cup I_- \cup J_-} (x_i^* - w_i^*) + \sum_i \overline{x}_i - \sum_i x_i^* + \sum_{J_+} x_i' \\ &\geqq -\sum_{J_+} x_i^* - \sum_i w_i^* + \sum_{J_+} w_i^* + \sum_i \overline{x}_i \\ &\geqq \sum_i \overline{x}_i + \sum_{J_+} [(w_i^* - \overline{x}_i) - (x_i^* - \overline{x}_i)] \end{aligned} \tag{5.96}$$

since $\sum_i w_i^* < 0$.

Thus, if for every individual whose strategy is constrained from above, and who therefore responds with the upper endpoint of the set of admissible strategies, the maximum possible subsidy is less than the difference between this strategy and the endowment, the last sum will be positive and a contradiction to the non-optimality of a is derived.

By a parallel argument for the case of $\sum_i w_i^* > 0$, we arrive at the condition

$$R > \sum_i \overline{x}_i + \sum_{J_+} [(-w_i^* - \overline{x}_i) - (x_i^* - \overline{x}_i)] \tag{5.97}$$

which has the interpretation given above. Q.E.D.

Since this is a sufficient condition, one can extend the class of essentially satisfactory mechanisms by checking that they satisfy it. For example, the pivotal mechanism plus any constant transfer α will be essentially satisfactory since the maximum subsidy is α and the admissible strategies are $[-\overline{x}_i - \alpha, \overline{x}_i + \alpha]$, so that members of J_+ will choose $\overline{x}_i + \alpha$ and members of J_- will choose $-\overline{x}_i - \alpha$.

It is, however, not the case that any mechanism admitting the possibility of a subsidy greater than $|w_i| - \overline{x}_i$, will fail to be a member of this class. To see this, we consider an example closely related to that used above.

Example 5.4. Let

$$\begin{aligned} h(w_j) &= -w_j - 1 & w_j &< 1 \\ &= 0 & -1 \leqq w_j &< 0 \\ &= -w_j & 0 \leqq w_j &< 1 \\ &= -1 & 1 &\leqq w_j \end{aligned} \tag{5.98}$$

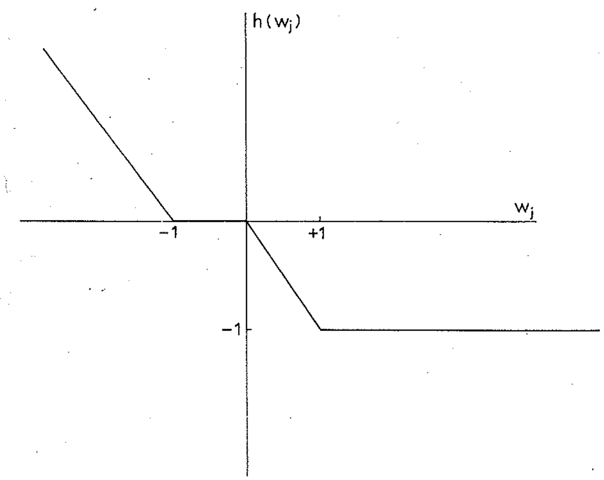

Figure 5.1.

The mechanism will have dominant admissible strategies by virtue of theorem 5.9. The admissible strategy correspondence is given by,

$$S_i^a(\overline{x}_i) = \begin{cases} [-\overline{x}_i, \overline{x}_i] & \text{if } \overline{x}_i < 1 \\ (-\infty, +\infty) & \text{if } \overline{x} \geqq 1. \end{cases} \tag{5.99}$$

We will now show that this mechanism is essentially satisfactory for all two-person economies even though it may give a transfer in excess of the absolute value of the difference between the strategy played and the endowment, for some agent. To see this consider the environment,

$$\begin{aligned} \bar{x}_1 &= 0, \quad & v_1 &= 5 \\ \bar{x}_2 &= 10, \quad & v_2 &= -3 \end{aligned} \tag{5.100}$$

The project is rejected, since $w_1 = 0$, $w_2 = -3$, and agent 1 is in J_+ with $h(w_2) = 2 > 0 = w_1 - \bar{x}_1$. This demonstrates that the non-existence of such situations is not a necessary condition for essential satisfactoriness and that the converse of theorem 5.10 is not valid.

In the proof of satisfactoriness to follow we use two conditions which are each sufficient to insure that a particular social state is Pareto optimal.

Let $a = (K, x_1, \ldots, x_N, \sum_i \bar{x}_i - \sum_i x_i)$. If

(*) $K = 1$ and $\sum_i v_i > 0$, or $K = 0$ and $\sum_i v_i < 0$, then a is Pareto optimal.

(**) If a is such that for each i in $J_+ \cup J_-$, $x_i - \bar{x}_i \leqq |w_i| - \bar{x}_i$, then a is Pareto optimal.

Condition (*) follows from the definition of Pareto optimality; (**) follows from the proof of theorem 5.10.

We consider three cases, which, by symmetry, are exhaustive:

$$\begin{aligned} &\text{(i)} \quad \bar{x}_1 \geqq 1, \quad & \bar{x}_2 &\geqq 1 \\ &\text{(ii)} \quad \bar{x}_1 < 1, \quad & \bar{x}_2 &\geqq 1 \\ &\text{(iii)} \quad \bar{x}_1 < 1, \quad & \bar{x}_2 &< 1. \end{aligned} \tag{5.101}$$

Case (i)

Here, $w_1 = v_1$ and $w_2 = v_2$ are both admissible. Thus the resulting social state will satisfy the hypothesis in condition (*) and will be Pareto optimal.

Case (ii)

We will have $|w_1| \leqq \bar{x}_1$, and $w_2 = v_2$. Only individual 1 can be in J_+ or J_-. Take the case in which he is in J_+, so that $v_1 > \bar{x}_1$, and $w_1 = \bar{x}_1$ is his dominant admissible strategy. If $w_2 > -\bar{x}_1$, condition (*) applies and the result is Pareto optimal. If $w_2 < -\bar{x}_1$, the project will be rejected even though $v_1 + v_2$ may be positive. However $w_2 < 0$ and rejection of the project implies that the subsidy received by agent 1 is exactly $h(w_2)$. If there exists a feasible Pareto superior situation, it means that the quantity of the private good consumed by agent 1, $h(w_2) + \bar{x}_1$ exceeds the amount of compensation it is necessary to give agent 2, $-v_2$, for the change in the decision towards acceptance of the project. Since in this case $w_2 = v_2$, the non-optimality of the social state selected means

$$h(w_2) + \bar{x}_1 \geqq -w_2. \tag{5.102}$$

But the left hand side of this inequality is either $-w_2 - 1$ or zero, according to whether w_2 is above or below -1. Consider first the situation where $w_2 < -1$. Then (5.102) is

$$-w_2 - 1 + \bar{x}_1 \geqq -w_2 \tag{5.103}$$

or, $\bar{x}_1 \geqq 1$ as the condition for non-optimality – but this contradicts the hypothesis of case (ii).

If $w_2 \geqq -1$, then we need $\bar{x}_1 \geqq -w_2$ for non-optimality. But since $w_1 = \bar{x}_1$ and $w_1 + w_2 < 0$, this condition cannot be satisfied.

Therefore a Pareto superior social state cannot exist in this case. The same reasoning applies if individual 1 is in J_-.

Case (iii)

Here $|w_1| \leqq \bar{x}_1 < 1$ and $|w_2| < \bar{x}_2 < 1$.

But for these strategies, the h function coincides with the pivotal mechanism. Therefore no positive subsidies can ever arise and condition (**) applies to insure optimality of the resulting social state.

In summary, the non-existence of potential situations, in which subsidies in excess of $|w_i| - \bar{x}_i$ are paid, is a sufficient, but not a necessary condition for the satisfactoriness of social decision mechanisms with bounded consumption sets. A more precise delineation of the class of satisfactory mechanisms is at present an open problem.

Finally we return to the question of neutrality in the case of bounded consumption possibilities as mentioned at the end of the last chapter. Our results are somewhat mixed. Some, but not all, Pareto optima can be attained as the dominant strategy equilibria of satisfactory Groves mechanisms for these environments. Those optima that cannot be attained in this way can nevertheless be attained through *some* Groves mechanism – but it will be one of those that we had to eliminate due to either their failure to attain optima or their failure to admit dominant strategies in all other specifications of the environment. We proceed through a series of simple two-person examples.

Our first example is one in which Pareto optima can be attained through the appropriate specification of the mechanism.

Example 5.5. Consider the environment

$$\begin{aligned} \bar{x}_1 &= 0 \quad & v_1 &= 5 \\ \bar{x}_2 &= 10 \quad & v_2 &= -1 \end{aligned} \tag{5.104}$$

and it is desired to attain the optimum (1, 0, 10, 0) which produces the utilities (5, 9, 0) for the two agents and the government respectively. If for the first agent the transfer function used is the pivotal transfer function plus the constant, 6, and for the second agent we use just the pivotal transfer function, then the true tastes are elicited and the desired outcome is obtained. Further, the pair of transfer functions used remain satisfactory for all bounded consumption environments, by virtue of the remarks following equation (5.97).

A slight nuance in this procedure is necessary in some cases where the optimum involves rejecting the project in a situation in which the aggregate stated willingness to pay for the project is zero – which ordinarily would be associated with its acceptance. Consider the same example as above and the Pareto optimum (0, 1, 10, −1) which produces the utilities (1, 10, −1) for the two agents and the government respectively. In order to attain the rejection of the project, agent 1 must have a transfer function that exceeds the pivotal one by an amount strictly less than unity. If this were not the case, his response could not be as high as $w_1 = +1$, and under our usual convention of accepting the project when there is a tie, we would not reach the desired decision. If the transfer function were strictly below the pivotal mechanism plus 1, then when the project is rejected he cannot have a consumption of +1, and hence a utility of +1 would be unattainable. However, reversing our usual convention for acceptance and rejection, we could arrive at the desired optimum by giving him the pivotal transfer plus one, he would still not be pivotal, and his consumption would be as indicated. It is only necessary to observe that reversing our convention for breaking ties does not upset the satisfactoriness of a mechanism in any way.

Lest we be too optimistic about achieving neutrality, the following example brings out some of the inherent difficulties with bounded consumption environments:

Example 5.6. Let,

$$\begin{aligned} \bar{x}_1 &= 6 & v_1 &= 6 \\ \bar{x}_2 &= 10 & v_2 &= -4 \end{aligned} \tag{5.105}$$

We try to reach the optimum which has the utilities (6, 12, 0) associated with it. (It should be noted that this utility distribution is individually rational taking the project $K = 0$ as the initial position). To give agent 1 a consumption of zero when the project is accepted, it is necessary to use a transfer function for him such as the pivotal transfer *minus* two. For agent 2 we use the pivotal mechanism plus 6. We have previously observed in

theorem 5.7 that such a mechanism may leave individual 1 without any admissible strategies. Since the process is then ill-defined, it is not a satisfactory mechanism. It works for the case above where the environment is known, but it must be excluded from consideration.

Ignoring this difficulty for the moment, it can still be the case that if we use this mechanism in order to reach the optimum (6, 12, 0) with parameters as in (5.105) but the true parameters were

$$\begin{aligned} \bar{x}_1 &= 7 & v_1 &= -6 \\ \bar{x}_2 &= 10 & v_2 &= +5 \end{aligned} \tag{5.106}$$

a feasible decision would be taken, but it would be non-optimal. The strategies would be $w_1 = -5$, $w_2 = +5$ and the outcome induced would be (1, 5, 11, 1). Utilities produced would be (−1, 16, 1) which is dominated by (0, 16, 1) attained in the Pareto optimal social state (0, 0, 16, 1).

Thus we see that in the course of trying to reach a particular optimum for (5.101) we may be forced to use indecisive mechanisms or mechanisms producing non-optimal outcomes. In summary, therefore, the satisfactory mechanisms for bounded consumption environments do not form a neutral class of decision making processes.

Finally, we should observe that the satisfactoriness of the pivotal mechanism as found in theorem 5.6 does not generalize to the case of more than two possible social decisions.

Example 5.7. Let $\mathscr{K} = \{0, 1, 2\}$ and let $v_i(\cdot) = (0, 4, 6)$ describe the true preference relation of agent i for whom $\bar{x}_i = 5$. Clearly $w_i(\cdot) = v_i(\cdot)$ is not an admissible strategy for this agent, since he would risk being pivotal and having to pay as much as 6 units of the transferable resource.

It is easily verified that the admissible strategy space for the pivotal mechanism is

$$\{w_i(\cdot) \mid \max_K w_i(K) - \min_K w_i(K) \leqq \bar{x}_i\} \tag{5.107}$$

Since strategies are equivalent up to additive shifts, we will analyze potentially dominant strategies of the form

$$w_i(\cdot) = (0, w, 5) \tag{5.108}$$

which are obviously the relevant candidates. To see that there are no dominant strategies of this form, let

$$\sum_{j \neq i} w_j(\cdot) = (0, z(1), z(2)) \tag{5.109}$$

represent the aggregate announcements of the other agents. There are three cases

$$\begin{aligned} &\text{I: } w < 3 \\ &\text{II: } 4 \geqq w \geqq 3 \\ &\text{III: } w > 4 \end{aligned} \tag{5.110}$$

In case I, $w_i(\cdot)$ is seen not to be dominant when, for instance,

$$\begin{aligned} z(1) &= -3 \\ z(2) &= -10 \end{aligned} \tag{5.111}$$

Here, $K = 0$ would be adopted for a payoff of zero whereas by playing

$$\tilde{w}_i(\cdot) = (0, 3 + \varepsilon, 5) \tag{5.112}$$

the agent would have a payoff of one. Case III is similarly eliminated. In case II the matter is more delicate.

When $w < 4$ a counterexample similar to that in case I can be arranged by taking $-4 < z(1) < -w$, yielding a positive payoff to $\tilde{w}_i(\cdot)$ in excess of that obtained by $w_i(\cdot)$.

When $w = 4$, consider

$$\begin{aligned} z(1) &= -3\tfrac{1}{2} \\ z(2) &= -4\tfrac{3}{4} \end{aligned} \tag{5.113}$$

Since

$$w_i(\cdot) + \sum_{j \neq i} w_j(\cdot) = (0, \tfrac{1}{2}, \tfrac{1}{4}) \tag{5.114}$$

the project $K = 1$ is adopted, yielding a net payoff of $+\frac{1}{2}$. But by playing

$$\tilde{w}_i(\cdot) = (0, 3, 5) \tag{5.115}$$

the project would be $K = 2$ and a utility of $+1\frac{1}{4}$ would be obtained.

This lack of a dominant strategy within the admissible space precludes the attainment of a satisfactory mechanism.

Because only the pivotal mechanism guarantees $\sum_i t_i \leqq 0$ within the satisfactory class, there is no mechanism that is both satisfactory for these environments and such that a positive surplus is generated for the government.[7]

[7] The existence of satisfactory mechanisms for bounded consumption environments with more than two projects and without requiring that a surplus be generated is an open question.

5.5. Coalition incentive compatibility

All of the considerations introduced above concerning the incentive compatibility of social decision processes applies to potential distortions by individuals. They are the most likely difficulties that might be encountered, and early writers on the free-rider problem certainly had this case in mind. But in the attempt to overcome the lack of appropriate individualized incentives, some adverse incentives for coalitions may have been inadvertently created. Consider, for example, the pivotal mechanism applied to the three person economy in which $\mathscr{K} = \{0, 1\}$ and $v_1 = -6, v_2 = +2, v_3 = +2$. Although no single agent can successfully distort his preferences, the coalition of individuals 2 and 3, if they could collude and agree to say $+7$ each, would be able to force the acceptance of the project and avoid all payments under the pivotal rules. Note also that individual 1 cannot retaliate without hurting himself, although as a counterthreat strategy he could play $v_1 = -13$ which would have the effect of increasing the pivotal payments for agents 2 and 3 to 6 each, which is beyond the acceptable level. Therefore, it is not certain that coalitions will form and distort the outcome of the pivotal mechanism, but this argument shows that the pivotal mechanism is no longer a successful procedure when such cooperative behavior is relevant.

It is therefore natural to ask whether any dominant strategy mechanism that produces Pareto optimal outcomes can avoid this problem. Cheating when the structure of coalitions is fixed and known to the government can be overcome by treating each coalition as an individual and applying any Groves mechanism to the strategies consisting of the aggregated statements over the members of each coalition. We will deal with the more realistic, and potentially troublesome, case in which the structure of coalitions is unknown and the mechanism is required to preclude the successful formation of any coalition. This requirement cannot be met in general.

Let C be a coalition of agents. The vector v_C will denote the projection of $v \in \mathbf{R}^N$, the true willingnesses to pay, into the coordinate subspace defined by the members of C. Let $w \in \mathbf{R}^N$ be a vector of professed willingnesses to pay. We use $(w \backslash w_C)$ to denote the components of w with the exception of w_i, $i \in C$ and the notation $w = (w_{N \backslash C}, w_C)$.

The payoff function of a coalition C is

$$\begin{aligned} U_C(w) &= \sum_{i \in C} t_i(w) + \sum_{i \in C} v_i & \text{if } d = 1 \\ &= \sum_{i \in C} t_i(w) & \text{if } d = 0 \end{aligned} \tag{5.116}$$

Definition 5.7. A revelation mechanism is *strongly coalition incentive compatible* (*SCIC*) if for all C, v_C is a dominant strategy for the coalition C, that is

$$U_C(w_{N\backslash C}, v_C) \geqq U_C(w_{N\backslash C}, w_C), \text{ for all } w_{N\backslash C}, \text{ and all } w_C.$$

Theorem 5.11. There exists no *SCIC* successful revelation mechanism.

Proof. Clearly, a *SCIC* revelation mechanism must be a *SIIC* revelation mechanism. From the characterization theorem (theorem 4.3), it must then be a Groves mechanism. Thus we can write,

$$\begin{aligned} t_i(w) &= \sum_{j\neq i} w_j + h_i(w\backslash w_i) && \text{if } \sum_i w_i \geqq 0 \\ &= h_i(w\backslash w_i) && \text{if } \sum_i w_i < 0 \\ & && i = 1, \ldots, N \end{aligned} \tag{5.117}$$

Consider now a coalition of two agents, say agents j and l, and the economy formed by the resulting $(N-1)$ agents. The *SCIC* revelation mechanism for the economy of N agents must in particular be a *SIIC* revelation mechanism for this economy of $(N-1)$ agents.

Let $t_{j,l}(w)$ be the transfer function relevant for the coalition (j, l). By the characterization theorem,

$$\begin{aligned} t_{j,l}(w) &= \sum_{\substack{i\neq j\\ i\neq l}} w_i + h_{j,l}(w\backslash w_j, w_l) && \text{if } \sum_i w_i \geqq 0 \\ &= h_{j,l}(w\backslash w_j, w_l) && \text{if } \sum_i w_i < 0 \end{aligned} \tag{5.118}$$

From (5.117)

$$\begin{aligned} t_j(w) + t_l(w) &= \sum_{i\neq j} w_i + \sum_{i\neq l} w_i + h_j(w\backslash w_j) + h_l(w\backslash w_l) && \text{if } \sum_i w_i \geqq 0. \\ &= h_j(w\backslash w_j) + h_l(w\backslash w_l) && \text{if } \sum_i w_i < 0. \end{aligned} \tag{5.119}$$

Equalizing (5.118) and (5.119) we have:

$$\begin{aligned} \sum_{\substack{i\neq j\\ i\neq l}} w_i + h_{j,l}(w_j, w_l) &= \sum_{i\neq j} w_i + \sum_{i\neq l} w_i + h_j(w\backslash w_j) + h_l(w\backslash w_l) && \text{if } \sum_i w_i \geqq 0. \\ h_{j,l}(w\backslash w_j, w_l) &= h_j(w\backslash w_j) + h_l(w\backslash w_l) && \text{if } \sum_i w_i < 0. \end{aligned} \tag{5.120}$$

or,

$$\sum_i w_i + h_j(w\backslash w_j) + h_l(w\backslash w_l) = h_{j,l}(w\backslash w_j, w_l) \quad \text{if } \sum_i w_i \geqq 0. \tag{5.121}$$

$$h_j(w\backslash w_j) + h_l(w\backslash w_l) = h_{j,l}(w\backslash w_j, w_l) \quad \text{if } \sum_i w_i < 0. \tag{5.122}$$

Consider now $w^0 = (w_1^0, \ldots, w_N^0)$, w_j', and $w' = (w_1^0, \ldots, w_j', \ldots, w_N^0)$ such that

$$w_j^0 + w_l^0 + \sum_{\substack{i\neq j\\ i\neq l}} w_i^0 > 0 \tag{5.123}$$

$$w_j' + w_l^0 + \sum_{\substack{i\neq j\\ i\neq l}} w_i^0 < 0 \tag{5.124}$$

Substituting w^0 in (5.121) and w' in (5.122) and subtracting yields

$$\sum_i w_i^0 + h_l(w^0\backslash w_l^0) = h_l(w'\backslash w_l^0) \tag{5.125}$$

Consider $\tilde{w}_l^0 = w_l^0 + \varepsilon$, such that both (5.123) and (5.124) are preserved. From (5.125) we obtain a contradiction. Q.E.D.

Therefore, if we allow any type of coalition and in particular size 2 coalitions, there exists no *SCIC* successful mechanism. Is it possible to obtain a positive result by allowing for a restricted class of coalitions? We show next that it is not possible.

Let us assume that we require only that the mechanism be *SIIC* and robust with respect to a coalition of a fixed size, say n, but not necessarily so with respect to others. In such a case we will say that the mechanism is *size n – SCIC*.

If $n = 2$, the proof of theorem 5.19 can be used to show that there exists no size 2 *SCIC* successful revelation mechanism. For $n > 2$, the proof is more delicate, and in fact parallels the method used in section 3, theorem 5.3.

Theorem 5.12. There exists no size-n *SCIC* successful revelation mechanism.

Proof. As in theorem 5.11, the mechanism has to be a Groves mechanism so that (5.117) applies. Consider now a coalition C, of size n. We must have, using an argument parallel to the one of theorem 5.11

$$(n-1)\sum_i w_i + \sum_{i\in C} h_i(w\backslash w_i) = h_C(w\backslash w_C) \quad \text{if } \sum_i w_i \geqq 0. \tag{5.126}$$

$$\sum_{i\in C} h_i(w\backslash w_i) = h_C(w\backslash w_C) \quad \text{if } \sum_i w_i < 0. \tag{5.127}$$

We are going to fix w_i, $i \notin C$ at w_i^0 so that $h_C(w\backslash w_C)$ will be a constant.

For convenience we choose $w_i^0 = 0$, $i \notin C$. Consider now w_i^0, $i \in C$ and $\beta > 0$, so that

$$-(n-2)\beta > \sum_{i \in C} w_i^0 > -(n-1)\beta \tag{5.128}$$

We can now follow the method of theorem 5.3 to conclude that an equality can be obtained between terms involving only the function h_N, evaluated at various points, and an expression that varies with w_N^0. Since w_N^0 is not an argument of h_N, the proof is completed as in theorem 5.3. Q.E.D.

Using the method of theorem 4.5, we give now a proof that the family of Groves mechanisms are not *SCIC* even when preferences are restricted to the one parameter class[8] of functions V defined by:

(i) $v_i(K, \theta)$ is strictly concave in K and three times differentiable in the interior of its domain $\mathscr{K} \times \Theta$, where $\mathscr{K}$ is a closed interval of $\mathbf{R}$ and Θ an open set of $\mathbf{R}$.

(ii) for each $\theta \in \Theta$, $v_i(K, \theta)$ has a maximum in int $\mathscr{K}$.

(iii) $v_i(K, \theta)$ is not separable in K and θ.[9]

Theorem 5.13. The Groves mechanisms defined over V are not *SCIC*.

Proof. Let $\mathscr{N}$ be the class of all coalitions, that is, the class of all subsets of $\{1, \ldots, N\}$.

A ***DRM*** defined by the transfer functions $t_i(\cdot)$ is *SCIC* only if it satisfies the following system of partial differential equations:[10]

for any C in $\mathscr{N}$

$$\sum_{i \in C} \frac{\partial t_i}{\partial \theta_j} \equiv -\sum_{i \in C} \frac{\partial v_i}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_j}, \text{ for any } j \text{ in } C. \tag{5.129}$$

In particular by choosing $C = \{1\}, \{2\}, \ldots, \{N\}$, one obtains that the ***DRM*** must be a Groves mechanism (see section 4.4).

Simple manipulations show that (5.129) reduces to

$$\frac{\partial t_i}{\partial \theta_j} \equiv -\frac{\partial v_i}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_j} \qquad \begin{array}{l} j = 1, \ldots, N \\ i = 1, \ldots, N \end{array} \tag{5.130}$$

[8] Generalizations to multi-parameter classes and multi-dimensional projects are straightforward.

[9] If such separability holds, the optimal size of the project does not depend on tastes within the allowable class.

[10] The argument is the same as in (4.23). One writes that the truth is the best strategy for each coalition. Since this must hold for any value of the truth, the first order conditions of each coalition are transformed in identities.

From Poincaré's theorem, the system (5.130) is integrable, if and only if,

$$\frac{\partial^2 t_i}{\partial \theta_j \partial \theta_l} = \frac{\partial^2 t_i}{\partial \theta_l \partial \theta_j}, \qquad j = 1, \ldots, N, \qquad l = 1, \ldots, N \tag{5.131}$$

For $j \neq i$, $l \neq i$, an equation (5.131) reduces to an identity.

For $j = i$, $l \neq i$, equations (5.131) reduce to

$$\frac{\partial^2 v_i}{\partial K \partial \theta_i} \cdot \frac{\partial K^*}{\partial \theta_i} \equiv 0, \qquad i = 1, \ldots, N \tag{5.132}$$

Using the fact that K^* defines a maximum of $\sum_i v_i$, (5.132) can be rewritten as:

$$\frac{\left[\dfrac{\partial^2 v_i}{\partial K \partial \theta_i}\right]^2}{\sum_i \dfrac{\partial^2 v_i}{\partial K^2}} \equiv 0 \tag{5.133}$$

which requires:

$$v_i(K, \theta_i) = \phi_1(K) + \phi_2(\theta_i) \tag{5.134}$$

contradicting (iii). Q.E.D.

To obtain a contradiction in the proof of the above theorem it is enough to derive (5.130) for a single pair of agents. This implies that any family of coalitions $\mathscr{C}$ containing one two-agent coalition will have the property that there is no mechanism which is *SCIC* for $\mathscr{C}$.

One can use the method of theorem 5.13 to explore the families of coalitions which admit *SCIC* Groves mechanisms. For a given family $\mathscr{C}$, the question amounts to the integrability of the following system of partial differential equations.

for all C in $\mathscr{C}$

$$\text{for all } j \text{ in } C, \sum_{i \in C} \frac{\partial t_i}{\partial \theta_j} = -\sum_{i \in C} \frac{\partial v_i}{\partial K} \cdot \frac{\partial K^*}{\partial \theta_j} \tag{5.135}$$

The complete characterization of these families of coalitions is still an open question. As an illustration we give below an example of a class of coalitions acceptable in the one parameter quadratic class.

Example 5.7.

$$\begin{array}{l} v_i(K, \theta_i) = \theta_i K - \dfrac{K^2}{2}, i = 1, 2, 3, 4 \\ \mathscr{C} = \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2, 3\} \end{array} \tag{5.136}$$

We know from (4.25), that the coalitions $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$ impose the restriction:

$$t_i(\theta) = -\frac{\theta_i^2}{2}\frac{(N-1)}{N^2} + \frac{\theta_i}{N^2}\sum_{j\neq i}\theta_j + h_i(\theta_{-i}) \tag{5.137}$$

where $N = 4$ $\qquad i = 1, 2, 3, 4$

Incentive compatibility for coalition $\{1, 2, 3\}$ requires

$$\begin{aligned}
\frac{\partial t_1}{\partial\theta_1} + \frac{\partial t_2}{\partial\theta_1} + \frac{\partial t_3}{\partial\theta_1} &= -\frac{\partial v_1}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_1} - \frac{\partial v_2}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_1} - \frac{\partial v_3}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_1}\\
\frac{\partial t_1}{\partial\theta_2} + \frac{\partial t_2}{\partial\theta_2} + \frac{\partial t_3}{\partial\theta_2} &= -\frac{\partial v_1}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_2} - \frac{\partial v_2}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_2} - \frac{\partial v_3}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_2}\\
\frac{\partial t_1}{\partial\theta_3} + \frac{\partial t_2}{\partial\theta_3} + \frac{\partial t_3}{\partial\theta_3} &= -\frac{\partial v_1}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_3} - \frac{\partial v_2}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_3} - \frac{\partial v_3}{\partial K}\cdot\frac{\partial K^*}{\partial\theta_3}
\end{aligned} \tag{5.138}$$

which reduces to:

$$\begin{aligned}
\frac{\partial h_2}{\partial\theta_1} + \frac{\partial h_3}{\partial\theta_1} &= -\frac{(N-1)(\theta_2+\theta_3)}{N^2} + \frac{2\theta_1}{N^2}\\
\frac{\partial h_1}{\partial\theta_2} + \frac{\partial h_3}{\partial\theta_2} &= -\frac{(N-1)(\theta_1+\theta_3)}{N^2} + \frac{2\theta_2}{N^2}\\
\frac{\partial h_1}{\partial\theta_3} + \frac{\partial h_2}{\partial\theta_3} &= -\frac{(N-1)(\theta_1+\theta_2)}{N^2} + \frac{2\theta_3}{N^2}
\end{aligned} \tag{5.139}$$

The system (5.139) is satisfied by

$$\begin{aligned}
h_1(\theta_2, \theta_3) &= \frac{\theta_2^2+\theta_3^2}{2N^2} - \theta_2\theta_3\frac{(N-1)}{N^2}\\
h_2(\theta_1, \theta_3) &= \frac{\theta_1^2+\theta_2^2}{2N^2} - \theta_1\theta_3\frac{(N-1)}{N^2}\\
h_3(\theta_1, \theta_2) &= \frac{\theta_1^2+\theta_3^2}{2N^2} - \theta_1\theta_2\frac{(N-1)}{N^2}
\end{aligned} \tag{5.140}$$

The second order conditions for coalition $\{1, 2, 3\}$'s maximization problem are then easily checked.

Chapter 6

INDIVIDUAL RATIONALITY

6.1. Individual rationality and anonymity

In many decision-making procedures a *desiratum*, often imposed as a constraint, is that no individual can experience a decrease in welfare as a result of its operation – vis à vis some benchmark level. This base for utility measurement is often taken to be the initial endowment. Lying behind this requirement is the ethical precept that no one should have to be forced to participate in the mechanism, and each has a right to withdraw from the system, abstain from consuming any public goods, and live independently with his endowment intact. It is a type of unanimity principle; but rather than being concerned with unanimous approval of the final outcome selected as opposed to any other outcome, it relates to the fact that whatever the final outcome is, it will be preferable to the initial situation, or status quo. It is usually called *individual rationality*, and we will continue to use that term here. We apply this criterion to the mechanisms presented thus far in this book.

Consider the situation of an individual, i, whose willingness to pay for the public project is negative. Such an agent will be suffering a decrease in utility, whenever the project is accepted against his will and he is not otherwise compensated. Therefore, in order to insure individual rationality, we must have

$$t_i(w) \geqq -v_i(= -w_i) \tag{6.1}$$

whenever the project is accepted. Since $t_i(w) = \sum_{j\neq i}w_j + h_i(w_{-i})$ whenever $\sum_{j\neq i}w_j + w_i \geqq 0$, we have that

$$h_i(w_{-i}) \geqq 0 \tag{6.2}$$

is required to insure individual rationality of the mechanism. Clearly, individual i, if he favors the project, will never be hurt by this choice of h_i, since his utility in the event of acceptance is

$$v_i + \sum_{j \neq i} w_j + h_i(w_{-i}) \geqq 0. \tag{6.3}$$

Similarly, the non-negativity of $h_i(\cdot)$ guarantees that he will gain when the project is rejected.

The converse can be argued directly from the definitions. Suppose that $h_i(\overline{w}_{-i}) < 0$ for some $\overline{w}_{-i}$. The individual for whom $v_i = -\sum_{j \neq i} \overline{w}_j - \varepsilon$ for any $\varepsilon > 0$ faces the prospect of telling the truth and obtaining a negative transfer of $h_i(\overline{w}_{-i})$ since the project is rejected. Such a mechanism cannot be individually rational. Thus we have

Theorem 6.1. The set of all strongly individually incentive compatible and successful mechanisms that are individually rational for all agents is characterized by $h_i(w_{-i}) \geqq 0$ for all i and all w_{-i}.

This requirement is an unfortunate one for the practicality of employing incentive compatible mechanisms, because when h is everywhere non-negative, the cost of the transfers may become exceedingly high. For example, the "cheapest" mechanism in this class is the $h \equiv 0$ mechanism which generates the total required transfer of

$$\begin{aligned} \sum_i t_i(w) &= \sum_i \sum_{j \neq i} w_j \\ &= (N-1) \sum_i w_i \end{aligned} \tag{6.4}$$

in the event of acceptance. The payments of the private good will be growing like the *square* of the size of the population, at given level of the average willingness to pay. This is clearly infeasible in any realistic setting and it calls into question some of the basic assumptions necessary for the implementation of incentive-compatible schemes. We deal with these precisely in later chapters.

For the present, let us note that one possible way around the problem is to use only some of the individuals for the purpose of making the decision. Whether or not there is reason to believe that their tastes are representative of the entire population – a problem whose precise specification is given in part III – it is natural to pose the individual rationality criterion for the unsampled agents as well. Let the size of the sample be denoted n. The spirit of this consideration is that everyone should agree in principle to the use of the mechanism, before the identity of the members of the sampled group is known. The welfare improvement of the unsampled group, however, is impossible to guarantee under any incentive-compatible mechanism. The transfer they receive (or pay) is determined entirely by the statements of the sampled individuals, as is the selection of the project. Therefore, if the willingness-to-pay of an unsampled agent is sufficiently large and negative, he will be hurt and the mechanism will have no way of "knowing" this and compensating him for it.

We might therefore think of weakening our requirements to the statement that the unsampled agents should not be forced to pay anything to the sampled group, in terms of private commodities. This requires that

$$\sum_i h_i(w_{-i}) \leqq \min(0, -\sum_i \sum_{j \neq i} w_j)$$

and hence, in the case where anonymity is respected,

$$h_i(w_{-i}) \leqq \min(-\sum_{j \neq i} w_j, 0) \tag{6.5}$$

for all i, so that the members of the sampled group would never be directly subsidized for their participation. Equation (6.5) means that the mechanism must be bounded above by the pivotal transfers. Unfortunately, as can easily be seen, (6.2) and (6.5) are incompatible. Thus, even with this weakened notion of individual rationality for unsampled agents, there are no mechanisms having these desirable properties. Apparently, one must either give up the feasibility of the process – the fact that the unsampled group's transfers exactly balance the net transfers of sampled agents – or further weaken the concept of individual rationality required. The former route is pursued in chapter 9. The basis of the second idea is that the mechanism is to be thought of as being used repeatedly and hence an expected value criterion should be used to evaluate whether it is worthwhile to have the mechanism in force on average.

6.2. Individual rationality on average

Without requiring that every individual always be made better off by the mechanism, we can still retain the essential idea of individual rationality by asking that the process results in welfare improvement "on average" for all individuals, over many repeated applications.[1] Let us suppose that

[1] Here we are assuming that the individual knows his own tastes and is subjectively uncertain about the others. Later in this section we treat the situation where the mechanism is to be institutionalized in anticipation of being used repeatedly, and agents do not know what their preferences will be for the decisions to be made later.

every sampled individual believes that the joint distribution of the tastes of the $n - 1$ others, w_{-i}, is μ, and that this is also the common subjective belief of the unsampled agents about the joint strategies of each subset of $n - 1$ sampled agents. Let x be the sum of these $n - 1$ strategies.

The expected welfare change of a sampled agent, i, with preferences v_i is therefore

$$v_i \int_{\{w_{-i} \mid x \geqslant -v_i\}} \mathrm{d}\mu + \int_{\{w_{-i} \mid x \geqslant -v_i\}} x \,\mathrm{d}\mu + \int h_i(w_{-i})\mathrm{d}\mu \tag{6.6}$$

Letting P be the distribution of x we can write this as

$$v_i \int_{-v_i}^{\infty} \mathrm{d}P(x) + \int_{-v_i}^{\infty} x\,\mathrm{d}P(x) + \int h_i(w_{-i})\mathrm{d}\mu. \tag{6.7}$$

If this is to be positive for all v_i, then we are requiring that

$$\int h_i(w_{-i})\mathrm{d}\mu \geqslant \sup_{v_i} \int_{-v_i}^{\infty} -(v_i + x)\mathrm{d}P(x). \tag{6.8}$$

The integrand on the right-hand side is non-positive throughout the range of integration. Note that the supremum may not be attained but the value of the right-hand side approaches zero, when $v_i \to -\infty$. Thus

$$\int h_i(w_{-i})\mathrm{d}\mu \geqslant 0 \tag{6.9}$$

is required for the individual rationality of the sampled group in this sense.

For the unsampled group, their expected payments depend on the distribution of v_i, so we will assume that v_i and x are independent, for simplicity, and that v_i are independently and identically distributed within the sampled group. Let ν be the density of v_i. Then,

$$n\left[-\int_{-\infty}^{\infty}\int_{-v_i}^{\infty} x\,\mathrm{d}P(x)\nu(v_i)\mathrm{d}v_i\right] - \sum_i \int h_i(w_{-i})\mathrm{d}\mu \tag{6.10}$$

is the expected payment received by the unsampled group as a whole. While little can be said about this in general, the interesting and important case in which both x and v_i have mean zero has the property that the first term is negative. Therefore, a strictly negative value of $\int h_i \mathrm{d}\mu$, for at least one i, would be required for individual incentive compatibility of the unsampled group in this sense.

On the whole the results are negative for this situation as well, unless some assumptions can be placed on the joint distribution of x and v_i guaranteeing that

$$-\int_{-\infty}^{\infty}\int_{-v_i}^{\infty} x\,\mathrm{d}P(x)\nu(v_i)\mathrm{d}v_i \geqslant 0. \tag{6.11}$$

It is impossible to guarantee that every agent's welfare will improve on average, given common expectations, under any incentive compatible mechanism.

There is, however, one sense in which there may be successful mechanisms that are superior to the status quo decision. We can again appeal to the idea that the mechanisms are to be used repeatedly. But unlike the above analysis, where we require a superior average outcome for each v_i, we can presume that the individuals themselves believe that their own willingnesses to pay for the projects to be considered will vary over time.

A mechanism would seem to be acceptable if the agents would agree to its adoption as a procedure to be followed for decision-making before they actually find out what projects are going to be considered in this way.[2] For symmetry and ease of treatment we might suppose that each individual believes that his own tastes in the future will be distributed according to some probability law. Let $\nu(\cdot)$ denote the density function believed by the agent regarding his own v_i, ex ante.

Let us assume that h_i is the same for all i and is function only of x, a convention that will be maintained throughout this section.

Therefore, for a sampled individual to prefer to have the mechanism used rather than to have the status quo it is necessary that

$$\int_{-\infty}^{\infty} v_i \int_{-v_i}^{\infty}\mathrm{d}P\nu\,\mathrm{d}v_i + \int_{-\infty}^{\infty}\int_{-v_i}^{\infty} x\,\mathrm{d}P\nu\,\mathrm{d}v_i + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_i(x)\mathrm{d}P\nu\,\mathrm{d}v_i \geqslant 0. \tag{6.12}$$

Relation (6.12) can be converted to

$$-\int_{-\infty}^{\infty}\int_{-v_i}^{\infty}(v_i + x)\mathrm{d}P\nu\,\mathrm{d}v_i \leqslant \int_{-\infty}^{\infty} h_i(x)\mathrm{d}P. \tag{6.13}$$

For the unsampled agents, our previous condition of the expected non-positivity of transfers remains the same. That is,

$$\int_{-\infty}^{\infty}\int_{-v_i}^{\infty} x\,\mathrm{d}P\nu\,\mathrm{d}v_i + \int_{-\infty}^{\infty} h_i(x)\mathrm{d}P \leqslant 0. \tag{6.14}$$

This is the expected transfer to a typical sampled agent. Because of our symmetry assumptions (6.14) is equivalent to the non-positivity of expected transfers to the sampled group as a whole.

Thus, a mechanism that is individually rational in this expected sense satisfies

$$-\int_{-\infty}^{\infty}\int_{-v_i}^{\infty}(v_i + x)\mathrm{d}P\nu\,\mathrm{d}v_i \leqslant \int_{-\infty}^{\infty} h_i(x)\mathrm{d}P \leqslant -\int_{-\infty}^{\infty}\int_{-v_i}^{\infty} x\,\mathrm{d}P\nu\,\mathrm{d}v_i. \tag{6.15}$$

[2] This is in contrast to the view of Rawls which requires that the utility of the worst-off member of society be maximized. The difference stems from the fact that Rawls is considering a social constitution of much broader impact which will remain in force and will cover all situations in the future. Utilities refer to lifetime levels of welfare. In contrast to this, the present formulation has decisions covering very specific aspects of the social state, and "utilities" really refer just to the change in welfare induced by the particular decision at hand. See below, section 3, on this point as well.

Whether or not there is any $h_i(\cdot)$ that can satisfy this depends on the sign of $\int_{-\infty}^{\infty} v_i \int_{-v_i}^{\infty} \mathrm{d}P\, v\, \mathrm{d}v_i$.

If v and P are symmetric around zero,[3] then (6.15) can hold since

$$\int_{-\infty}^{\infty} v_i \int_{-v_i}^{\infty} \mathrm{d}P\, v\, \mathrm{d}v_i = \int_0^{\infty} |v_i| \left[\int_{-|v_i|}^{\infty} \mathrm{d}P - \int_{|v_i|}^{\infty} \mathrm{d}P\right] v\, \mathrm{d}v_i \geqslant 0 \tag{6.16}$$

by the fact that v is a density function.

Thus, there are in fact a wide variety of h-functions that are compatible with individual rationality in this sense. We would like to know which h-functions are members of this class. The answer to this question depends in general on P and v. We will not attempt to give a complete characterization of this here. If v is normal and P is the $(N-1)$ fold convolution of this distribution, however (an assumption that will be used repeatedly in part III), then the pivotal mechanism will be one member of the set in question.

To see this, note first that the right-hand inequality in (6.15) is always assured when the pivotal mechanism is used, because the sampled group can only be taxed. The left-hand inequality in (6.15) can be proven as follows:

$$-\int_{-\infty}^{\infty}\int_{-v_i}^{\infty} (v_i + x)\mathrm{d}P\, v\, \mathrm{d}v_i =$$

$$= -\int_{-\infty}^{\infty}\left[v_i \int_{-v_i}^{\infty} \frac{1}{\sqrt{(2\pi(N-1))}} \mathrm{e}^{-x^2/2(N-1)}\mathrm{d}x + \frac{1}{\sqrt{(2\pi(N-1))}} \int_{-v_i}^{\infty} x\mathrm{e}^{-x^2/2(N-1)}\mathrm{d}x\right] \frac{1}{\sqrt{(2\pi)}}\mathrm{e}^{-v_i^2/2}\mathrm{d}v_i$$

$$= -\int_{-\infty}^{\infty} v_i \rho\left(\frac{v_i}{\sqrt{(N-1)}}\right)\frac{1}{\sqrt{(2\pi)}}\mathrm{e}^{-v_i^2/2}\mathrm{d}v_i - \frac{N-1}{\sqrt{(2\pi N)}} \tag{6.17}$$

where $\rho(\xi) = \int_{-\infty}^{\xi} v\, \mathrm{d}v_i$. And

$$\int_{-\infty}^{\infty} h_i \mathrm{d}P = \frac{-1}{\sqrt{(2\pi(N-1))}} \int_0^{\infty} x\mathrm{e}^{-x^2/2(N-1)}\mathrm{d}x = -\frac{(N-1)^{1/2}}{\sqrt{(2\pi)}}. \tag{6.18}$$

[3] This is a sufficient condition, but (6.15) can be shown to hold more generally.

Taking the difference between these expressions, we define

$$\phi(N) = -\frac{N-1}{\sqrt{(2\pi(N-1))}} + \frac{N-1}{\sqrt{(2\pi N)}} + \int_{-\infty}^{\infty} v_i \rho(v_i/\sqrt{(N-1)}) v\, \mathrm{d}v_i. \tag{6.19}$$

We prove that $\phi(N)$ is positive for $N \geqslant 2$ (and thus in any meaningful case) by showing

(i) $\lim_{N\to\infty} \phi(N) = 0$

(ii) $\phi'(N) < 0$.

Statement (i) follows from the fact that $\lim \rho(v_i/\sqrt{(N-1)}) = \frac{1}{2}$ for all v_i and that $\int_{-\infty}^{+\infty} v_i v\, \mathrm{d}v_i = 0$. To see the validity of (ii), note that

$$\phi'(N) = \frac{1}{\sqrt{(2\pi)}}\left(\frac{1}{\sqrt{N}} - \frac{1}{\sqrt{N-1}}\right) + \frac{N-1}{\sqrt{(2\pi)}}\cdot\frac{1}{2}\left(\frac{1}{(N-1)^{3/2}} - \frac{1}{N^{3/2}}\right) - \frac{1}{2\sqrt{(2\pi)}}\frac{1}{N\sqrt{(N)}} \tag{6.20}$$

which can be seen to be negative for $N \geqslant 2$ because it is negative at 2, has a positive derivative for all $N \geqslant 2$, and $\lim_{N\to\infty} \phi(N) = 0$. These results can be summarized in:

Theorem 6.2. In the family of Groves mechanisms:

(1) If we require utility improvement on average for each value of the willingness-to-pay of a sampled individual, then there are no individually rational mechanisms.

(2) If we require only utility improvement on average for members of the sampled group, there will be individually rational mechanisms when tastes are distributed symmetrically around zero.

(3) If tastes are normally distributed with mean zero, then the pivotal mechanism displays individual rationality on average.

6.3. Willingness to participate in the sample

For many purposes, the concept of individual rationality in the context of this type of decision-making process is irrelevant because the status quo may either be undefined or may be clearly worse than any alternative. In the general case where one public decision out of an arbitrary set $\mathscr{K}$ must be selected, the status quo may not correspond to any of these decisions. In other instances, even though $\mathscr{K} = \{0, 1\}$ the status quo may happen to be the project and it is its potential rejection that is being considered.

However, there is still a concept, related to but distinct from individual rationality, that is relevant to the question of how the mechanism should operate. Our view has been that the use of the mechanism should be agreed to by the participants in some ex ante sense. The spirit of this is related to the original position or contractarian ideal of moral philosophy. When no status quo is a clear benchmark, the mechanism still should be arranged in such a way that every agent will prefer to participate in the decision, through the mechanism, rather than withdrawing from society and leaving the decision up to the others. Instead of a fixed status quo, the result of the mechanism is thus compared to an alternative that would be optimal for a society in which the individual in question is indifferent between these outcomes.

There are two features that a successful mechanism must have if abstentions like this are to be considered. First, the decision taken must be Pareto optimal for the tastes expressed; and second, the respondents must be accurately representative of the entire population. The simplest way to achieve the latter criterion is that the mechanism must induce everyone to reply, rather than abstain. This is called the property of *universal participation.*

The decision regarding participation may depend on the individuals' beliefs regarding the statements made by other members of the sample. An interesting feature of the dominant strategy mechanisms in this regard is that the requirement of universal participation, independently of expectations, leads to a class of mechanisms defined by a set of functions $h_i(\cdot)$ which are bounded below by the pivotal mechanism. Thus, if the government wants to minimize the total subsidies required by these procedures (or maximize the revenues they collect) then the pivotal mechanism should be selected within the class inducing universal participation. The pivotal mechanism would therefore have a very strong claim on our attention.

Theorem 6.3. Let $h_i(w_{-i}) = \min(-\sum_{j\neq i} w_j, 0)$ define the pivotal mechanism. Let $h_i'(\overline{w}_{-i}) < h_i(\overline{w}_{-i})$ for some $\overline{w}_{-i}$. Then there exists $v_i \in \mathbf{R}$ and expectations concerning $\overline{w}_{-i}$ such that an individual with these tastes and beliefs would refuse to participate if the mechanism were defined by $h_i'(\cdot)$.

Proof. It suffices to consider $v_i = 0$, and expectations concentrated on $(\overline{w}_{-i})$. The expected payoff to such an individual will be

$$\sum_{j\neq i} \overline{w}_j + h'(\overline{w}_{-i}) \qquad \text{if } \sum_{j\neq i} \overline{w}_j \geqslant 0$$

$$h_i'(\overline{w}_{-i}) \qquad \text{if } \sum_{j\neq i} \overline{w}_j < 0$$

Since $h_i'(\overline{w}_{-i}) < h_i(\overline{w}_{-i})$, we will have a negative quantity in either case, from the definition of $h_i(\cdot)$. Hence the individual would be unambiguously worse off if he were to participate, as he would sustain a sure loss. Q.E.D.

Thus, any mechanism that gives lower transfers at any point than the pivotal mechanism will not necessarily induce universal participation.

We assume throughout this section that individuals are all risk-neutral. A mechanism induces participation by an individual, i, whose beliefs about $x = \sum_{j\neq i} w_j$ are given by P, only if the expected value of his payoff is non-negative. The subjective joint distribution of w_{-i} is denoted μ.

Theorem 6.4. An individual with $v_i = 0$ will choose to participate in the sampled group if and only if the mechanism would induce participation for any individual holding the same expectations.

Proof. The expected utility, given participation, can be written as:

$$\int_{-v_i}^{\infty} (v_i + x)\mathrm{d}P + \int_{-\infty}^{\infty} h_i(w_{-i})\mathrm{d}\mu \tag{6.21}$$

Non-participation would induce the expected utility,

$$\int_0^{\infty} v_i \mathrm{d}P \tag{6.22}$$

The difference between them is, therefore,

$$\int_{-v_i}^{0} (v + x)\mathrm{d}P + \int_0^{\infty} x\mathrm{d}P + \int_{-\infty}^{\infty} h_i(w_{-i})\mathrm{d}\mu \tag{6.23}$$

Note that the first integral is non-negative whereas the second two are independent of v_i. Further, as $v_i \to 0$, the first integral converges to zero.

Therefore, if for some v_i the mechanism does not induce participation, then it will not induce participation by an individual whose $v_i = 0$. Moreover, if the second two integrals are non-negative, then an individual for whom $v_i = 0$ will participate, and hence so will all individuals. Q.E.D.

Theorem 6.5. The set of all mechanisms that induce universal participation by risk neutral individuals independently of their expectations is characterized by the set of all $h_i(\cdot)$ such that

$$h_i(w_{-i}) \geqslant \min(-\sum_{j\neq i} w_j, 0).$$

Proof. For given expectations μ, theorem 6.4 implies that the criterion for universal participation can be established by setting $v_i = 0$ in (6.21). Thus,

$$\int_0^{\infty} x\mathrm{d}P + \int_{-\infty}^{\infty} h_i(w_{-i})\mathrm{d}\mu \geqslant 0.$$

ince it must hold for any expectations μ, it implies

$$h_i(w_{-i}) \geqslant \min(\sum_{j \neq i} w_j, 0). \tag{6.25}$$

).E.D.

There is one final concept relevant to the individual rationality question vhen abstention is not a relevant possibility (as for example in the comnittee chairman choice problem where membership in the committee arries with it a responsibility to participate in the election of a chairman).)ne might require that each individual not be taxed in the resulting outome unless he is willing to pay at least this amount to avoid what would ave been the social choice had he not expressed his preferences. This can e called the *non-confiscatory* requirement. It is an ex post rationality ather than ex ante. However, it is easily seen to be formally equivalent to ie universal participation requirement for all expectations. In particular, ie expectations may be beliefs held with certainty, concentrated on the rategies that were actually played. Therefore, these concepts are equivent, and, by virtue of theorem 6.5, the necessary and sufficient condition r a mechanism to be non-confiscatory is just that it exceeds or equals ie pivotal mechanism's transfers in all situations.

Chapter 7

APPROACHES TO INCENTIVE COMPATIBILITY NOT REQUIRING DOMINANCE

7.1. Introduction

Thus far we have pursued the possibility of devising mechanisms that always have dominant strategies and which always give rise, at the dominant strategy equilibrium, to Pareto optima. Because of the non-closedness of the system, a necessary drawback as shown in chapter 5, we cannot devise mechanisms that achieve efficient allocations for the participants alone. An outside source, or sink, for revenue must be included in the system if strict optimality is to be maintained. An alternative approach is to weaken either the optimality or dominant strategy requirements – but hopefully not too much – in order to close the system.

In the next section we consider the possibility of assuming that individuals play their maximin strategies in a game in which there are no dominant strategies. This corresponds to an implicit assumption that their degree of risk aversion is very high. Without this condition, there is no reason to suppose that the maximin solution will be achieved at the equilibrium of the game. Nevertheless, a very strong result can be achieved in this way. There exists a closed mechanism whose maximin equilibria are always Pareto optimal.

In sections 4 and 5 we will deal with the case in which the expectations of each agent about the statements to be made by the others are known in advance by the planner, though he remains ignorant of their willingnesses-to-pay. Part III of the book uses the largeness of the economy, a valid assumption in many public decision-making problems, to obtain successful results without adding assumptions as strong as those of the present chapter. To be precise about this will require an explicit consideration of statistical and expectational issues.

In section 6 we explore the possibility of obtaining successful results when one assumes that Nash equilibria are realized in the game created by the mechanism.

7.2. The Dubins mechanism

The following mechanism, due to Dubins [1974], has the property that each individual's true willingness-to-pay is his maximin strategy, and that an optimal result is achieved when these strategies are played. All of the assumptions of the previous chapters are maintained; in particular, we reemphasize the assumption of separable preferences and its restrictive character.

Let π be any probability distribution over the space of projects $\mathscr{K}$.

The Dubins mechanism is defined as follows:

$$d(w) \text{ maximizes } \sum_i w_i(K) \text{ over } \mathscr{K}. \tag{7.1}$$

Let $K^* \in \mathscr{K}$ be the maximizing project.

$$t_i(w) = -w_i(K^*) + \int_{\mathscr{K}} w_i(K)\,\mathrm{d}\pi(K) + \frac{1}{N}\sum_j w_j(K^*) - \frac{1}{N}\int_{\mathscr{K}} \sum_j w_j(K)\,\mathrm{d}\pi(K). \tag{7.2}$$

Theorem 7.1. For the Dubins mechanism, the set of maximin strategies for each agent i is

$$\{w_i(\cdot)\,|\,w_i(\cdot) = v_i(\cdot) + \alpha,\ \alpha \in \mathbf{R}\}. \tag{7.3}$$

But, as each of these strategies produces precisely the same outcome, we are justified in saying that $w_i(\cdot) = v_i(\cdot)$ is *the* maximin strategy.)

Proof. If truthful preferences are revealed by agent i, his payoff is:

$$v_i(\mathrm{d}(w)) + t_i(w) \geqslant \int_{\mathscr{K}} v_i(K)\,\mathrm{d}\pi \tag{7.4}$$

Consider a strategy $w_i(\cdot)$ which is not in (7.3). Let $\tilde{w}_i(\cdot)$ be defined by

$$\tilde{w}_i(K) = w_i(K) + \int_{\mathscr{K}} v_i(K)\,\mathrm{d}\pi - \int_{\mathscr{K}} w_i(K)\,\mathrm{d}\pi \tag{7.5}$$

because

$$\int_{\mathscr{K}} \tilde{w}_i(K)\,\mathrm{d}\pi = \int_{\mathscr{K}} v_i(K)\,\mathrm{d}\pi \tag{7.6}$$

the strategy $\tilde{w}_i(\cdot)$ is equivalent to $w_i(\cdot)$. Thus, there exists $K' \in \mathscr{K}$ such that

$$\tilde{w}_i(K') - v_i(K') = \lambda > 0 \tag{7.7}$$

Let us choose the strategies of the other agents such that

$$\begin{aligned} &\sum_{j\neq i} w_j(K) = -\tilde{w}_i(K) \qquad &&\text{for } K \in \mathscr{K},\ K \neq K' \\ &\sum_{j\neq i} w_j(K') = -\tilde{w}_i(K') + \varepsilon \qquad &&\text{for } \varepsilon > 0. \end{aligned} \tag{7.8}$$

When the strategies played are $\tilde{w}_i$ and w_j, the project K' is chosen and the payoff to agent i is

$$v_i(K') - \tilde{w}_i(K') + \int_{\mathscr{K}} \tilde{w}_i(K)\,\mathrm{d}\pi + \frac{1}{N}\left[\sum_{j\neq i} w_j(K') + \tilde{w}_i(K')\right] - \frac{1}{N}\int_{\mathscr{K}}\left[\sum_{j\neq i} w_j(K) + \tilde{w}_i(K)\right]\mathrm{d}\pi \leqslant -\lambda + \int_{\mathscr{K}} v_i(K)\,\mathrm{d}\pi + \frac{\varepsilon}{N} \tag{7.9}$$

By choosing ε sufficiently small, we have shown that any strategy that is not equivalent to the truthful one, risks an outcome inferior to $\int_{\mathscr{K}} v_i(K)\,\mathrm{d}\pi$, which is assured by truthful revelation. Q.E.D.

Theorem 7.2. For the Dubins mechanism, the sum of transfers is identically zero.

Proof. Obvious, on summing (7.2) over all i. Q.E.D.

An interpretation of the Dubins mechanism is that it provides for each of the individuals to share in the social surplus generated by the mechanism in the following way: Truthful strategies yield a maximin utility level of

$$\int_{\mathscr{K}} v_i(K)\,\mathrm{d}\pi. \tag{7.10}$$

If project K^* is selected when the truthful strategies are played at the maximin equilibrium, total utility in the population is exactly $\sum_i v_i(K^*)$, since $\sum_i t_i(v) = 0$ by theorem 7.2. Thus, the social surplus provided by the mechanism beyond the sum of the security levels is just

$$\sum_i v_i(K^*) - \int_{\mathscr{K}} \sum_i v_i(K)\,\mathrm{d}\pi. \tag{7.11}$$

The final utility level achieved by each agent can therefore be thought of as his security level plus an equal $(1/N)$ share of this social surplus. Since the project is worth $v_i(K^*)$ to him, the transfer he must be given at the truthful equilibrium is

$$-v_i(K^*) + \int_{\mathscr{K}} v_i(K)\mathrm{d}\pi + \frac{1}{N}\left[\sum_j v_j(K^*) - \int_{\mathscr{K}} \sum_j v_j(K)\mathrm{d}\pi\right] \tag{7.12}$$

which is identical with (7.2) at the truthful strategies.

As long as each individual's share in the redistribution of the surplus is non-negative, it will never disturb the security level that the mechanism provides for him. Thus, it is easy to see that the Dubins mechanism can be generalized slightly by using weights δ_i such that

$$\delta_i \geqslant 0, \qquad \sum_i \delta_i = 1 \tag{7.13}$$

and writing the transfer function as

$$t_i(w) = -w_i(K^*) + \int_{\mathscr{K}} w_i(K)\mathrm{d}\pi + \delta_i(\sum_j (w_j(K^*) - \int_{\mathscr{K}} w_j(K)\mathrm{d}\pi)). \tag{7.14}$$

Indeed the distribution of the surplus can be done in even more complex ways, without destroying the closedness of the system, the maximin property of truthful strategies or Pareto optimality. Using the normalization $\int_{\mathscr{K}} w_i(K)\mathrm{d}\pi = 0$ for all strategies, one can write the transfer function in general as

$$t_i(w) = -w_i(K^*) + \phi_i(w(\cdot)) \tag{7.15}$$

where the functions ϕ_i must satisfy only the following two conditions:

(i) $\inf_{\{w(\cdot) | \sum_j w_j(K) \text{ is maximized at } K^*\}} \phi_i(w(\cdot)) = 0$ for each $K \in \mathscr{K}$, $i = 1, \ldots, N$

(ii) $\sum_j \phi_j(w(\cdot)) \equiv \sum_j w_j(K^*)$ where K^* is the maximizing project.

These two properties do not restrict the ϕ_i to any simple parametric class. For example, ϕ_i can be an arbitrary polynomial in $\sum_j w_j(K^*)$ with non-negative coefficients that can depend in an arbitrary way on $w(\cdot)$ provided they remain bounded. In this way, still more complex maximin mechanisms can be generated. As these do not acquire any favorable properties not already possessed by the Dubins mechanism, we will concentrate on the equal-rebate procedure for the remainder of this chapter.

We turn now to the question of individual rationality in the Dubins mechanism, paralleling the discussion of chapter 6. Recall that a mechanism is said to be individually rational with respect to the status quo project $\overline{K} \in \mathscr{K}$ if, for all agents i and all preferences $v_i(\cdot)$, the outcome of the mechanism is not inferior to $v_i(\overline{K})$. It is easy to see that the Dubins mechanism is not individually rational in this sense. Consider the following example:

Example 7.1.

$$\begin{array}{lll} \overline{K} = 0, & \mathscr{K} = \{0, 1\}, & \pi \text{ uniform} \\ v_1(0) = 0 & v_1(1) = 4 & \\ v_2(0) = 0 & v_2(1) = -2 & \end{array} \tag{7.16}$$

The Dubins mechanism would choose $K = 1$ and would effect the transfers,

$$\begin{array}{l} t_1(v) = -1\tfrac{1}{2} \\ t_2(v) = 1\tfrac{1}{2} \end{array} \tag{7.17}$$

Clearly, individual 2 would prefer $K = 0$ and no transfers to the outcome of this process.

Perhaps more interestingly, there may be no project in $\mathscr{K}$ such that, if this were regarded as the status quo, then the Dubins mechanism would have the individual rationality property. To see this, just take a case in which there are as many agents as projects, and for each project there is someone for whom this is his least preferred outcome.

There is one sense, however, in which the Dubins mechanism can be regarded as individually rational. Indeed it is a natural one in our context. Suppose that the status quo is that the project will be chosen randomly according to the distribution π. This provides precisely the security level (7.8) for each agent; and the mechanism is obviously individually rational with respect to it in light of the above discussion.

Without further restrictions on the preferences, there is no other concept of status quo, with respect to which the Dubins mechanism would maintain its individual rationality. But there may be some circumstances in which there is a natural status quo project, $\overline{K}$, – one for which

$$v_i(\overline{K}) = \int_{\mathscr{K}} v_i(K)\mathrm{d}\pi \tag{7.18}$$

for every agent i. We will encounter such a system in chapter 15, where $\mathscr{K} = \mathbf{R}$, $\overline{K} = 0$, π is symmetric around $\overline{K}$ and $v_i(K)$ is linear in K. In this case, $\overline{K} = 0$ serves as a natural benchmark.

7.3. Problems with bounded consumption in maximin procedures

Although it is a natural and potentially useful way of avoiding the non-closedness of successful mechanisms while still retaining some favorable features of incentive compatibility, the Dubins mechanism fails to retain its efficiency properties when the potential for bankruptcy is introduced. In chapter 5, it was shown that some of the Groves mechanisms, but not all, are also successful in environments with consumption bounded below. It is natural to assume in such contexts, that the individual treats the bound on his consumption of the private good as a constraint and then uses the same objective function that we had chosen in the unconstrained case, in the present instance maximizing the minimal possible utility. It is therefore necessary to analyze the behavior of the maximin strategy when consumption is constrained. Let us treat the two-alternative case first. It is simpler, qualitatively different, and will also have an important application in part IV.[1]

Since social states are invariant under shifts in the strategies by constants, it is allowable for us to normalize them, without loss of generality. A strategy w_i will be an element of $\{w_i(\cdot) \mid \int_{\mathscr{K}} w_i(K)\mathrm{d}\pi = 0\}$.

The transfers can then be written in the simplified form,

$$t_i(w) = -w_i(K^*) + \frac{1}{N}\sum_j w_j(K^*) \tag{7.19}$$

where K^* is the maximizing project.

In the two-alternative case a strategy is just

$$w_i(\cdot) = (-w_i, +w_i) \tag{7.20}$$

using the assumption that π is the uniform distribution. As in the proof of the maximin property of the strategy $w_i(\cdot) = v_i(\cdot)$, we can easily see that $1/N \sum_j w_j(K^*)$ can be made arbitrarily small. Therefore, the requirement of feasibility is just

$$w_i(K^*) \leqslant \overline{x}_i \tag{7.21}$$

where $\overline{x}_i$ is the initial endowment of individual i. For the two-alternative case it is clear that if the truthful strategy $v_i(\cdot) = (-v_i, +v_i)$ is such that $|v_i| > \overline{x}_i$, then the constrained maximin strategy is just either $w_i(\cdot) = (-\overline{x}_i, +\overline{x}_i)$ or $(+\overline{x}_i, -\overline{x}_i)$, according to the sign of v_i.

In the Dubins mechanism we have not discussed explicitly the manner in which a tie in the stated willingness-to-pay across several projects is to be broken. Paralleling theorem 5.6, we can show that a Pareto optimum will be selected, except perhaps in the case of such ties.

[1] See chapter 15 for a maximin mechanism in non-separable environments.

Theorem 7.3. If there are two alternatives and one has a strictly higher sum of stated willingnesses-to-pay than the other, then when a Dubins mechanism is used, the result is Pareto optimal.

Proof. Let the two projects be denoted $K = 0$ and $K = 1$ and a strategy will be written as $w_i(\cdot) = (-w_i, +w_i)$ where the first entry applies to $K = 0$, and the second to $K = 1$. Clearly, if the Dubins mechanism selects a project for which the sum of the true willingnesses-to-pay is maximized, then the result will be Pareto optimal.

Suppose therefore that $K = 1$ is selected with a strict (apparent) superiority: $\sum_i w_i > 0$, but that $\sum_i v_i < 0$, because of the constraints introduced by potential bankruptcies. Let us divide the individuals into three groups as follows:

$$\begin{aligned} N_1 &= \{i \mid |v_i| \leqslant \overline{x}_i\} \\ N_2 &= \{i \mid v_i > \overline{x}_i \geqslant 0\} \\ N_3 &= \{i \mid -v_i > \overline{x}_i \geqslant 0\} \end{aligned} \tag{7.22}$$

This is clearly a mutually exclusive and exhaustive classification.

The constrained maximin strategies of these three types of agents are:

$$\begin{aligned} &i \in N_1 \colon w_i = v_i \\ &i \in N_2 \colon w_i = \overline{x}_i \\ &i \in N_3 \colon w_i = -\overline{x}_i \end{aligned} \tag{7.23}$$

Under the decision $K = 1$, their utilities attained are:

$$\begin{aligned} &i \in N_1 \colon u_i = v_i + \overline{x}_i - v_i + \gamma = \overline{x}_i + \gamma \\ &i \in N_2 \colon u_i = v_i + \overline{x}_i - \overline{x}_i + \gamma = v_i + \gamma \\ &i \in N_3 \colon u_i = v_i + \overline{x}_i + \overline{x}_i + \gamma = v_i + 2\overline{x}_i + \gamma \end{aligned} \tag{7.24}$$

where $\gamma = (\sum_{i \in N_1} v_i + \sum_{i \in N_2} \overline{x}_i - \sum_{i \in N_3} \overline{x}_i)/N > 0$ by hypothesis.

A Pareto superior point must involve the adoption of the other project. Let x_i be the consumption of individual i such that, with the other project as the social outcome, he would attain at least as high a utility as in the Dubins equilibrium. Thus we must have

$$\begin{aligned} &i \in N_1 \colon -v_i + x_i \geqslant \overline{x}_i + \gamma \\ &i \in N_2 \colon -v_i + x_i \geqslant v_i + \gamma \\ &i \in N_3 \colon -v_i + x_i \geqslant v_i + 2\overline{x}_i + \gamma \end{aligned} \tag{7.25}$$

Since $x_i \geqslant 0$ is required for feasibility, we can assert

$$\begin{aligned} &i \in N_1 \colon x_i \geqslant v_i + \overline{x}_i + \gamma > v_i + \overline{x}_i \\ &i \in N_2 \colon x_i \geqslant 2v_i + \gamma \quad > 2v_i \\ &i \in N_3 \colon x_i \geqslant 0 \end{aligned} \tag{7.26}$$

Adding we have

$$\sum x_i \geqslant \sum_{i\in N_1} v_i + \sum_{i\in N_1} \overline{x}_i + 2\sum_{i\in N_2} v_i$$
$$> \sum_{i\in N_1} \overline{x}_i + 2\sum_{i\in N_2} v_i - \sum_{i\in N_2} \overline{x}_i + \sum_{i\in N_3} \overline{x}_i \quad (7.27)$$

The latter inequality follows on subtracting $N\gamma$ from the right-hand side which is a positive quantity.

Using $v_i > \overline{x}_i$ for $i \in N_2$ we have

$$\sum_i x_i > \sum_i \overline{x}_i \quad (7.28)$$

which demonstrates the infeasibility of the hypothetically superior alternative resource allocation. Q.E.D.

However, if $N_2 = \phi$ and $\gamma = 0$, a strict inequality in the above proof cannot be demonstrated. Therefore, if there is a tie among the two projects, the potential for a non-optimal situation is suggested. The following example shows that a non-optimal Dubins equilibrium will exist if the "wrong" project were to be selected in such a case. Since the planner cannot, under our assumptions, know anything about whose statement is made under the constraint of bankruptcy and whose is not, but can only know the actual strategies played, this "mistake" of choosing the wrong project cannot be avoided.

Example 7.2. Consider two individuals whose characteristics are as follows:

$$v_1(\cdot) = (+10, -10) \qquad \overline{x}_1 = 8$$
$$v_2(\cdot) = (-8, +8) \qquad \overline{x}_2 = 12 \quad (7.29)$$

The constrained maximin strategies are

$$w_1(\cdot) = (+8, -8)$$
$$w_2(\cdot) = (-8, +8) \quad (7.30)$$

so a tie results. Assume that it is broken in favor of the project $K = 1$, the second entry in the strategy vector. Thus the transfers will be $t_1 = +8$, $t_2 = -8$, and the utilities will be

$$u_1 = +6$$
$$u_2 = +12 \quad (7.31)$$

But if the project $K = 0$ were adopted, and if the 20 units of aggregate endowment were allocated exclusively to the second individual, the utilities attained would be

$$u_1 = +10$$
$$u_2 = +12 \quad (7.32)$$

which is a Pareto superior situation since one individual's welfare has been strictly increased, although the other (a member of N_1 in the notation of theorem 7.3) is kept at the original welfare level.

Since the probability of such a tie is rather low in some sense, one might not view the inefficiency noted above as a serious drawback of the Dubins mechanism. However, it can be seen that the efficiency properties of the constrained Dubins procedure breakdown in a more fundamental sense when the number of alternatives exceeds two.

Let us investigate first the behavior of constrained maximin strategies in such cases. According to (7.12) the basic constraint is

$$\max_K w_i(K) \leqslant \overline{x}_i \quad (7.33)$$

together with the normalization $\int_{\mathscr{K}} w_i(K)\mathrm{d}\pi = 0$. Suppose that $v_i(\cdot)$ is such that $v_i(K) > \overline{x}_i$ for some K. When $v_i(K)$ is reduced to the $\overline{x}_i$ level, compensating variations in the v_i function for other projects must be made in order to restore the normalization of the strategy. When there was a single alternative project, the necessary change was clear: if $v_i(K)$ were reduced to $\overline{x}_i$ or increased to $-\overline{x}_i$, the other project would have to receive exactly the negative evaluation, suitably weighted in case π were not uniform. With several other projects, however, there are generally many variations compatible with restoring the normalization and with observing the bankruptcy constraint.

For example, if there were three projects, $K = 0, 1, 2$, and

$$v_i(\cdot) = (-6, 2, 4)$$
$$\overline{x}_i \ \ = 3 \quad (7.34)$$

then $v_i(2)$ would have to be reduced to 3. The values of $v_i(0)$ and $v_i(1)$ would have to rise to maintain $\int_{\mathscr{K}} w_i \mathrm{d}\pi = 0$. The maximin objective is

$$\max_{w_i} \min_K [v_i(K) - w_i(K)] \quad (7.35)$$

since the surplus can be arbitrarily small. Therefore, with π uniform, the one extra unit $(v_i(2) - \overline{x}_i)$ would be distributed evenly between the other two projects: The optimal $w_i(\cdot)$ would be $(-5\frac{1}{2}, 2\frac{1}{2}, 3)$.

However, if in the same example we had $\overline{x}_i = 2$, then the even distribution of $(v_i(2) - \overline{x}_i)$ among the other alternatives is itself incompatible with the bankruptcy constraint. Therefore, the optimum is

$$w_i(\cdot) = (-4, 2, 2) \quad (7.36)$$

All of the weight of the constraint falls on only one of the two alternatives, in this case $K = 0$.

This phenomenon can cause the Dubins equilibrium to be Pareto inefficient in a more robust and serious way than in the two-alternative case.

Example 7.3. Consider the two individuals whose characteristics are

$$v_1 = (-6, 2, 4) \qquad \bar{x}_1 = 2$$
$$v_2 = (6, -1, -5) \qquad \bar{x}_2 = 10 \tag{7.37}$$

As discussed above, the maximin strategies will be

$$w_1(\cdot) = (-4, 2, 2)$$
$$w_2(\cdot) = (6, -1, -5) \tag{7.38}$$

nd the alternative $K = 0$ will be adopted. Transfers will be

$$t_1 = 4 + 1 = 5$$
$$t_2 = -6 + 1 = -5 \tag{7.39}$$

nd hence utilities at the Dubins equilibrium are given by

$$u_1 = -6 + \bar{x}_1 + t_1 = +1$$
$$u_2 = 6 + \bar{x}_2 + t_2 = +11 \tag{7.40}$$

ut consider the alternative of accepting the project $K = 1$ and instituting he transfers

$$t'_1 = -2$$
$$t'_2 = +2 \tag{7.41}$$

his would produce utilities

$$u'_1 = 2 + \bar{x}_1 + t'_1 = +2$$
$$u'_2 = -1 + \bar{x}_1 + t'_2 = +11 \tag{7.42}$$

hich are strictly superior to the equilibrium levels.

Moreover, although one of the two individuals' utilities has not been nproved in this example, with small perturbations one can construct cases here both can be made strictly better off. The interested reader will easily e able to verify that if $\bar{x}_1 \in (2, 2\frac{1}{2})$ in this example, then an allocation rictly superior to the Dubins equilibrium will be attainable.

The failure of the Dubins mechanism to be successful in more-than-two-ternative environments is somewhat different than the related difficulty the pivotal mechanism that was explored in example 5.7. There, the

absence of dominant strategies within the admissible class meant that dominant strategies could not be defined. Here, however, maximin strategies exist, but the equilibrium is inefficient.

7.4. Mechanisms based on expected utility maximizing strategies

One parameter of the process which we did not investigate above is the probability distribution π that is used in the normalization of strategies. It would seem that this does not matter at all, at least in the model without consumption bounds, for the π selected will influence only the level of the values of $w_i(K)$, but will not differentiate among the $w_i(K)$ for various values of K. It cannot affect either the project selected or the transfers made. This omission is therefore not serious as long as we concentrate on maximin procedures.

It is interesting, however, to consider situations in which the measure π represents the agent's own beliefs concerning the probability of adoption of various projects. These will be derived from their subjective evaluations concerning the strategies to be played by the others, their knowledge of the workings of the mechanism and the knowledge of their own statements. The last point is crucial because it means that π cannot be regarded as a fixed distribution perhaps differing among the agents, but must be treated as a function of the strategy $w_i(\cdot)$ that is played.

Assume therefore that the government or planner knows the subjective distribution that each of the agents has regarding all of the others' strategies. With this information, the Dubins mechanism is instituted, but for each individual, i, the transfer is computed using the measures $\pi_j(w_j(\cdot); \cdot)$, $j = 1, \ldots, N$ that describe the subjective beliefs held concerning the likelihood of various projects being instituted. The individuals are assumed to be risk-neutral and to maximize their expected payoff using these distributions.

In such a mechanism, the truth is no longer an optimal response. However, the optimal strategies are simply a multiple of the true utilities; and the multiple is just the size of the population, N. In particular, it is the same for all agents and it is independent of the subjective beliefs held. These two properties of the expected utility-maximizing strategies allow us to conclude that the result of the mechanism will be unaffected by these distortions – a Pareto optimum will be selected.[2]

[2] A related mechanism due to Thompson [1967] explores the idea that agents beliefs about the probability of acceptance of various projects may be useful to the planner in constructing

We proceed now to analyse the two-alternative case. For environments with consumption unconstrained, the multi-project case is a direct analog. Denoting the two projects by $K = 0$ and $K = 1$, respectively, we can adopt the normalization $w_i(0) = 0$ for all i. (The reason for this change in the normalization procedure is that the way in which project weights change depends on the individuals' own statements. The methods of sections 1 and 2 of this chapter would require the normalization to be determined endogenously.)

Each individual is characterized by his preferences for the two projects, which we normalize in the same way as the strategies, and his beliefs about the strategies to be played by others. By virtue of this normalization we can denote preferences and strategies simply by v_i and w_i, respectively, indicating the tastes for the project, $K = 1$. The subjective beliefs of agent i are given by a probability measure $\mu_i(w_{-i})$ on $\mathbf{R}^{N-1}$.

The distribution μ_i induces two other measures of interest for this analysis. We can consider the distribution of the random variable, $x = \sum_{j \neq i} w_j$; this is denoted P_i. We will assume that P_i has a density function denoted $p_i(\cdot)$. We can calculate the probabilities over the projects induced by any particular statement of the individual. In this case there are only two projects, so without fear of confusion, we denote the probability that $K = 1$ will be accepted when the agent says w_i as:

$$\begin{aligned}\pi_i(w_i) &= P_i(\{x \geqslant -w_i\}) \\ &= \mu_i(\{w_{-i} \mid \textstyle\sum_j w_j \geqslant 0\}).\end{aligned} \tag{7.43}$$

The process to be defined depends on these $\pi_j(\cdot)$. One way to conceptualize it is that the central planner, who formulates the mechanism, knows the expectation μ_j for each j. From these he computes the $\pi_j(\cdot)$ and these functions are communicated to all agents along with the transfer functions which are analogous to those of the Dubins mechanism:

$$\begin{aligned} t_i(w) &= -w_i + w_i\pi_i(w_i) + \frac{1}{N}\sum_j w_j - \frac{1}{N}\sum_j w_j\pi_j(w_j) && \text{if } \sum_j w_j \geqslant 0 \\ &= w_i\pi_i(w_i) - \frac{1}{N}\sum_j w_j\pi_j(w_j) && \text{if } \sum_j w_j < 0 \end{aligned} \tag{7.44}$$

That is, the Dubins transfer is used except that the individuals' subjective beliefs are used as weights for the projects.

incentive compatible mechanisms. However these beliefs were not allowed to depend on the agent's own statement, which of course he should know. Recognizing this dependence is central to the approach taken here.

Since the individual is assumed to be risk neutral, his expected payoff can be decomposed into the expected payoff concerning the project itself, Ev_i, and the expected transfer, $E(t_i(w_i))$. These are

$$Ev_i = v_i\pi_i(w_i) \tag{7.45}$$

and

$$\begin{aligned} Et_i(w_i) = &\int_{\{w_{-i}\mid\sum_j w_j \geqslant 0\}} \left[-w_i + \frac{1}{N}\sum_j w_j + w_i\pi_i(w_i) - \frac{1}{N}\sum_j w_j\pi_j(w_j)\right]\mathrm{d}\mu_i \\ &+ \int_{\{w_{-i}\mid\sum_j w_j < 0\}} \left[w_i\pi_i(w_i) - \frac{1}{N}\sum_j w_j\pi_j(w_j)\right]\mathrm{d}\mu_i \end{aligned} \tag{7.46}$$

Consolidating terms somewhat, we obtain

$$Et_i(w_i) = \frac{1}{N}\left(\int_{\{w_i\mid\sum_j w_j \geqslant 0\}} \sum_{j\neq i} w_j\,\mathrm{d}\mu_i - \int_{\{w_{-i}\}} \sum_{j\neq i} w_j\pi_j(w_j)\,\mathrm{d}\mu_i\right).$$

Differentiating the objective function, $Ev_i + Et_i(w_i)$, with respect to w_i, we have,

$$v_i\frac{\mathrm{d}\pi_i(w_i)}{\mathrm{d}w_i} - \frac{1}{N}w_i p_i(-w_i) \tag{7.48}$$

The derivative $\mathrm{d}\pi_i(w_i)/\mathrm{d}w_i$ can be evaluated as follows:

$$\pi_i(w_i) = \int_{\{w_i\mid\sum_j w_j \geqslant 0\}} \mathrm{d}\mu_i = \int_{-w_i}^{\infty} p_i(x)\,\mathrm{d}x$$

$$\frac{\mathrm{d}\pi_i(w_i)}{\mathrm{d}w_i} = p_i(-w_i) \tag{7.49}$$

Hence the first-order condition requires $w_i = Nv_i$.

Since the second derivative of the objective function is $-(1/N)p_i(-w_i)$, which is negative, this strategy defines a true maximum.

The optimality of these strategies is independent of the distributions describing the beliefs of any of the agents. They depend only on the fact that the distributions used by the government to define the mechanism are in fact the same ones with respect to which the agents maximize their expected payoff. This is not a small requirement, for if these beliefs are privately held pieces of data similar to preferences, as it is natural to assume, then the individuals will have no reason to reveal them ex ante.

A "free-rider" problem of a greatly increased complexity would have to be solved.

There may be circumstances, however, in which the agents can view each other's strategies with uncertainty for objective, rather than subjective reasons. This will play a central role in chapters 12 and 13. For other purposes, the agents actually involved in the preference revelation game we have been describing may be those selected randomly as a sample to represent a larger population. If the statistical characteristics of this larger population are known to both the agents and the planner, and if the size of the sample is transmitted to the agents, as we have been assuming, then their beliefs about the joint distribution of the others can be inferred on a rational basis and used in the construction of the mechanism. In such cases the procedure studied here may be of relevance.

Finally, note that we have not been able to elicit the true preferences. However, since $\sum_i v_i = 1/N \sum_i w_i$, the project $K = 1$ will always be accepted or rejected according to whether it is optimal in the case at hand. No inefficiency can arise as a result of the distortions. Indeed, it is the case that the government, since it *knows* (by reading the demonstration above?) that individuals will respond with $w_i = Nv_i$, independently of these beliefs can simply divide by N in order to recover the true parameters.

This is therefore an example of a direct revelation mechanism that does not elicit the truth but can still be used to attain Pareto optimal results with probability one. Recently, d'Aspremont and Gérard-Varet [1977] have given a direct preference revelation mechanism that is successful and for which the truth is the expected utility maximizing strategy. In the notation above, the transfer functions for their mechanism are given by:

$$\begin{aligned} t_i(w) &= -w_i + \frac{1}{N}\sum_j w_j + \sum_l w_l \pi_i(w_i) \\ &\quad - \frac{1}{N}\sum_j \sum_l w_l \pi_j(w_j) \qquad \text{if } \sum_j w_j \geq 0 \\ &= \sum_j w_j \pi_i(w_i) - \frac{1}{N}\sum_l \sum_j w_j \pi_l(w_l) \qquad \text{if } \sum_j w_j < 0 \end{aligned} \tag{7.50}$$

By using $\sum_l w_l$ instead of w_i and w_j, $j = 1, \ldots, N$, respectively, in the two last terms of the transfer function, the maximizing strategy is shifted to $w_i = v_i$. All other qualitative properties and information requirements of this mechanism are the same as for that given above.

We conclude this section by showing that expected utility maximizing mechanisms are necessarily of the d'Aspremont-Gérard-Varet form (7.50) under certain conditions, and that this may be viewed as a modified version of Groves mechanisms, allowing for differing expectations among individuals. We adopt the continuously divisible project spaces $\mathscr{K}$ and the open set of parameters $\Theta_i \subset \mathbf{R}$ over which the valuation functions $v_i(K, \tilde{\theta}_i)$ are twice differentiable ($\tilde{\theta}_i \in \Theta_i$).

We restrict attention to the case of additively separable transfer functions $t_i(\theta)$ and we let $r_i(\theta_i)$ be the component of $t_i(\theta)$ corresponding to the agent's announced value of his parameter, θ_i.

The individual maximizes

$$\int_{\Theta_{-i}} (v_i(K^*(\theta), \tilde{\theta}_i) + r_i(\theta_i))\, d\mu_i(\theta_{-i}) \tag{7.51}$$

For incentive compatibility we require that the maximizer of (7.51) coincide identically with $\tilde{\theta}_i$ throughout Θ_i. Thus

$$\int_{\Theta_{-i}} \frac{\partial v_i}{\partial K}\frac{\partial K^*}{\partial \theta_i}\, d\mu_i(\theta_{-i}) \equiv -\frac{\partial r_i}{\partial \theta_i} \tag{7.52}$$

Integrating (7.52) in θ_i

$$r_i(\theta_i) = \int \int_{\Theta_{-i}} \frac{\partial v_i}{\partial K}\frac{\partial K^*}{\partial \theta_i}\, d\mu_i(\theta_{-i})\, d\theta_i + \text{constant} \tag{7.53}$$

From the theory of dominant strategy mechanisms (see (4.24) and (4.26)) we know that

$$\int \frac{\partial v_i}{\partial K}\frac{\partial K^*}{\partial \theta_i}\, d\theta_i = \sum_{j \neq i} v_j(K^*, \theta_j) + h_i(\theta_{-i}) \tag{7.54}$$

Reversing the order of integration in (7.53), (7.54) implies

$$r_i(\theta_i) = \int_{\Theta_{-i}} \sum_{j \neq i} v_j(K^*, \theta_j)\, d\mu_i(\theta_{-i}) + \text{constant} \tag{7.55}$$

The constraint that the mechanism is balanced can be used to determine the transfer functions for all agents i if we assume that the transfer function takes the symmetric form

$$t_i(\theta) = r_i(\theta_i) - \frac{1}{N-1}\sum_{j \neq i} r_j(\theta_j) \tag{7.56}$$

In this case, setting the constants on the right-hand side of (7.55) to zero, we obtain precisely the d'Aspremont-Gérard-Varet system,

$$t_i(\theta) = \int_{\Theta_{-i}} \sum_{j \neq i} v_j(K^*, \theta_j) \mathrm{d}\mu_i(\theta_{-i})$$
$$- \frac{1}{N-1} \sum_{l \neq i} \int_{\Theta_{-l}} \sum_{j \neq l} v_j(K^*, \theta_j) \mathrm{d}\mu_l(\theta_{-l}) \quad (7.57)$$

Choosing other constants clearly redistributes income but does not alter the basic character of these mechanisms. Therefore, subject to the restrictions of (7.56), the d'Aspremont-Gérard-Varet mechanism is the general form of an incentive compatible mechanism based on expected utility maximization in differentiable environments. Because the second-term on the right-hand side of (7.57) can be arbitrarily large and negative if other agents' beliefs are set appropriately, there are no expected utility maximizing methods that can avoid the problems of individual rationality.

7.5. Willingness to participate in processes based on expected utility maximizing strategies

In this section we address the problem of individuals' willingnesses to participate in the generalized Dubins mechanism outlined in section 4 above. We will treat only the case of a discrete, two-project set of possible decisions. Further, no restrictions on the transfer functions will be considered. Unlike in chapter 6, the closedness of the present system allows us to completely disregard the welfare of anyone whose preferences are not elicited. If it is the case that the mechanism operates only on a sample from the population it clearly cannot compensate non-sampled individuals for their decreased utility if a decision is made against their will. We will be presuming that the mechanism elicits the tastes of all agents. There are, therefore, only two concepts for participation paralleling those for the sampled agents in chapter 6.

We could require that for each agent, the expected utility from participation exceeds the utility of non-participation. We assume that each agent has the same expectations concerning the statements to be made by each of the others. Agent i's beliefs about each other agent j are denoted $\mu_{ij}(w_j)$. Thus, if he responds to the mechanism the individual's expected utility is:

$$v_i \pi_i(w_i) + E(t_i(w_i)) = v_i \int_{-w_i}^{\infty} \mathrm{d}P_i - w_i \int_{-w_i}^{\infty} \mathrm{d}P_i + \frac{1}{N} \int_{-w_i}^{\infty} (w_i + x) \mathrm{d}P_i$$

$$+ w_i \int_{-w_i}^{\infty} \mathrm{d}P_i - \frac{N-1}{N} \int_{-\infty}^{\infty} w_j \int_{-w_j}^{\infty} \mathrm{d}P_i \mathrm{d}\mu_{ij}$$
$$- \frac{1}{N} w_i \int_{-w_i}^{\infty} \mathrm{d}P_i \quad (7.58)$$

The first term is the expected utility from the project itself and the others are the expected value of the transfers from the generalized Dubins mechanism.

This mechanism does not induce participation for all agents and all patterns of expectations.

Example 7.4. Consider the case of an agent whose $v_i = 0$ in a three-individual system where the statements of the other two individuals are independently and uniformly distributed on $[-1, +1]$. His response will be $w_i = 0$, and the expression (7.58) is,

$$\tfrac{1}{3}\textstyle\int_0^1 x \mathrm{d}P_i - \tfrac{2}{3}\int_{-1}^{+1} w_j \int_{-w_i}^{+1} \mathrm{d}P_i \tfrac{1}{2} \mathrm{d}w_j \quad (7.59)$$

using the fact that the density of w_j is $\frac{1}{2}$ on $[-1, 1]$ and the density of x is

$$p_i(x) = 1 + x \qquad x \in [-1, 0]$$
$$= 1 - x \qquad x \in [0, +1]$$

as it is the sum of two independent uniform random variables. It can be shown that (7.59) equals $-\frac{11}{18}$. Since the expected utility of nonparticipation is zero for such an agent, he will not have the incentive to participate.

As discussed above in chapter 6, the failure of universal participation will lead to biased results whenever the actual distribution of willingnesses-to-pay around zero is asymmetric. Even if the anticipated distributions of willingnesses-to-pay is symmetric around zero, in any actual realization of the process, especially one in which the number of individuals is small, the empirical distribution of tastes will not be symmetric and significant biases may result.

There is, however, a weaker sense in which this mechanism is acceptable. The idea parallels the discussion of participation "on average" of chapter 6. We require that individuals be willing to go along with the institution of the procedure under the assumption that it will be used repeatedly to make public decisions and that their own tastes will vary from issue to issue, having the same distribution over issues as the distribution of other individuals' tastes do in a single given issue.

Participation is then worth, on average, the integral of (7.58) with respect to w_i, giving w_i the distribution μ_{ij}, which is actually independent of j.

$$\int_{-\infty}^{\infty} \frac{w_i}{N} \int_{-w_i}^{\infty} dP_i d\mu_{ij} - \int_{-\infty}^{\infty} w_i \int_{-w_i}^{\infty} dP_i d\mu_{ij} + \frac{1}{N} \int_{-\infty}^{\infty} \int_{-w_i}^{\infty} (w_i + x) dP_i d\mu_{ij}$$

$$+ \int_{-\infty}^{\infty} w_i \int_{-w_i}^{\infty} dP_i d\mu_{ij} - \frac{1}{N} \int_{-\infty}^{\infty} \int_{-w_j}^{\infty} (w_j + x) dP_i d\mu_{ij}$$

$$= \int_{-\infty}^{\infty} \frac{w_i}{N} \int_{-w_i}^{\infty} dP_i d\mu_{ij} \tag{7.60}$$

If the individual were never to participate, but instead the decisions would be made on the basis of the others' statement alone, his expected utility would clearly be

$$\int_{-\infty}^{\infty} v_i \int_0^{\infty} dP_i d\mu_{ij} \tag{7.61}$$

Using $v_i = w_i/N$, we find that the difference between (7.60) and (7.61) is

$$\frac{1}{N} \int_{-\infty}^{\infty} w_i \left[\int_{-w_i}^{\infty} dP_i - \int_0^{\infty} dP_i \right] d\mu_{ij} = \frac{1}{N} \int_{-\infty}^{\infty} w_i \int_{-w_i}^{0} dP_i d\mu_{ij} \tag{7.62}$$

which is clearly positive since $\int_{-w_i}^{0} dP_i \geqslant 0$ if and only if $w_i \geqslant 0$.

Thus we have established:

Theorem 7.4. The generalized[3] Dubins mechanism induces participation on average, independently of the distribution of preferences held by the individuals, provided the individuals' beliefs about each of the others are the same.

7.6. The Nash equilibrium approach

In this chapter, we have explored a number of mechanisms based on expected utility maximization. One may be particularly interested in those

[3] By analogous computations we find that the d'Aspremont–Gérard–Varet mechanism induces participation on average as well.

patterns of expectations that are self-fulfilling. If these self-fulfilling expectations are point expectations about the strategies of the other agents, the "equilibrium outcome" is a Nash equilibrium. An attractive mechanism would then be one for which a Nash equilibrium[4] always exists and every Nash equilibrium is Pareto optimal.[5]

The construction of outcome functions (or mechanisms) with such properties in exchange economies with or without public goods has been the object of much recent research (Hurwicz [1976], Hurwicz–Schmeidler [1976], Postlewaite–Schmeidler [1978], Groves–Ledyard [1975], Champsaur [1977], Wilson [1978] and Schmeidler [1976]).

This area is too unsettled to attempt a synthesis. In the next paragraphs, we discuss a few such constructions for economies with public goods.

7.6.1. The Groves–Ledyard government

The difficulties created by the need for a separability assumption and by the impossibility to balance exactly the budget in dominant strategy mechanisms (see chapter 5) led Groves and Ledyard [1975] to the construction of a new mechanism derived to some extent from the Groves mechanisms, with the property that Nash equilibria exist and are Pareto optimal. In this paragraph, we try to give an intuitive presentation of this "government" in the context of our previous analysis.

We noticed in chapter 5 that, when preferences are quadratic with only the linear parameter as an unknown, it is possible to design a balanced Groves mechanism by an appropriate choice of an h function. Groves and Ledyard, observing that the Pareto optimality of a Nash equilibrium in convex environments is a local property, utilize the concavity of the Groves–Loeb transfer functions to satisfy the first-order conditions for optimality. This allows them to obtain balancedness of the process by using only quadratic messages which can be defined by a single parameter.

Let us be somewhat more precise on this point. Consider agents having separable preferences with willingness-to-pay functions $v_i(K)$ and a convex cost function for the level of the public good in terms of the transferable resource, given by $\Gamma(K)$.

[4] These Nash equilibria are "naive" in the sense that an agent does not attempt to modify his behavior to reach a different Nash equilibrium (see Hurwicz [1977]).

[5] Peleg [1976a] and Dutta and Pattanaik [1975] explore mechanisms with the weaker requirement that there exists a Pareto optimal Nash equilibrium for all preference profiles. They also discuss the possibility of having Nash equilibria for coalitions.

The share of the cost is defined ex ante for each agent and is given by a fixed fraction, α_i, of the total. Thus the net willingness-to-pay of agent i is $v_i(K) - \alpha_i\Gamma(K)$. If the project level K is selected, this means that the individual will be taxed $\alpha_i\Gamma(K)$ in addition to any other taxes which he pays by virtue of the incentive aspects of the mechanism.

Agents' messages are restricted to be quadratic functions of the form

$$w_i(K) = qs_iK - \frac{qK^2}{2N} \tag{7.63}$$

where q is some fixed positive number not under the agents' control. A choice of a message amounts to a choice of s_i, which is therefore referred to as the "message", s_i, as well.

The mechanism is defined so that the K which is selected maximizes the sum of the net willingnesses-to-pay. This clearly implies

$$K = \sum_i s_i \tag{7.64}$$

The transfer to each agent is defined as follows:

Consider the Groves mechanism defined for environments in which the true willingnesses-to-pay were of the form (7.63), for which the sum of transfers was identically zero. Such a mechanism was derived in chapter 5, equation (5.47).

The transfer of each agent is this, minus his share of the cost as defined above. By virtue of this construction, the mechanism is sure to be closed for any combination of messages, s_i, $i = 1, \ldots, N$.

Unlike dominant strategy mechanisms, however, this mechanism is played by agents who maximize utility given the strategies played by everyone else. Using the definition of the mechanism, the utility of agent i is

$$v_i(\sum_l s_l) - \alpha_i\Gamma(\sum_l s_l) + t_i(s) \tag{7.65}$$

where t_i is the part of the transfer function defined through the particular Groves mechanism discussed above. Since it is a Groves mechanism, it can be written as

$$t_i(s) = \sum_{j \neq i} w_j(\sum_l s_l) + h_i(s_{-i}) \tag{7.66}$$

where $w_j(\cdot)$ are the net willingness-to-pay functions. Thus utility is maximized when s_i satisfies

$$\frac{\mathrm{d}v_i}{\mathrm{d}K}(\sum_l s_l) - \alpha_i\frac{\mathrm{d}\Gamma}{\mathrm{d}K}(\sum_l s_l) + \sum_{j \neq i}\frac{\mathrm{d}w_j}{\mathrm{d}K}(\sum_l s_l) = 0 \tag{7.67}$$

Because

$$\sum_i \frac{\mathrm{d}w_i}{\mathrm{d}K}(\sum_l s_l) = 0$$

(7.67) implies that

$$\frac{\mathrm{d}v_i}{\mathrm{d}K}(\sum_l s_l) - \alpha_i\frac{\mathrm{d}\Gamma}{\mathrm{d}K}(\sum_l s_l) = \frac{\mathrm{d}w_i}{\mathrm{d}K}(\sum_l s_l) \tag{7.68}$$

The interpretation of (7.68) is that, although the net willingness-to-pay functions are restricted to be quadratic, the announced net marginal valuation will be set equal to the true net marginal valuation when agents follow their Nash behavior. It is this property which accounts for the optimality of Nash equilibria in this mechanism.

To summarize, we have lost the dominant strategy property because the optimal strategy for the objective function (7.65) depends, in general, on $\sum_{j \neq i} s_j$. However, we have gained the closedness of the mechanism because the transfer functions (7.59) sum to zero identically. Finally, (7.66) optimality is preserved by virtue of the fact that since (7.58) is concave in s_i, (7.65) the true net marginal willingnesses-to-pay are elicited.

To be sure that this mechanism is a meaningful way of achieving optimal outcomes, we must examine the question of the existence of equilibria. In the separable case we have considered, this is easily done.

Let K^* satisfy

$$\sum_i \frac{\mathrm{d}v_i}{\mathrm{d}K}(K^*) = \frac{\mathrm{d}\Gamma}{\mathrm{d}K}(K^*) \tag{7.69}$$

and consider the messages s_i^* defined by

$$\frac{\mathrm{d}v_i}{\mathrm{d}K}(K^*) = \alpha_i\frac{\mathrm{d}\Gamma}{\mathrm{d}K}(K^*) + qs_i^* - \frac{q}{N}K^* \tag{7.70}$$

which, for each $i = 1, \ldots, N$, is a simple linear equation in s_i^*. The vector of messages $(s_i^*)_{i=1,\ldots,N}$ forms a Nash equilibrium, as can be seen directly by substitution in (7.60). Therefore the existence (and uniqueness, when (7.67) $v_i(\cdot)$ or $-\Gamma(\cdot)$ are strictly concave) of Nash equilibria are assured for this class of environments.

Two problems remain regarding the existence issue. First, when consumption is bounded below, we must insure that a feasible system of transfers is attained. Second, the argument above must be extended to non-separable utilities. There are problems in both of these directions.

In order to balance the mechanism, certain functions of the others' strategies had to be added to the transfers in (7.66). These are given in (5.56) as a function of the linear coefficient in the (hypothetical) quadratic utility. It is clear from this expression that for given values of s_j, $j \neq i$, the value of $h_i(s_{-i})$ may be negative. Therefore, even though negative values of s_i are allowed for each agent, the restriction $K \geqslant 0$ places a lower bound on s_i of $\sum_{j \neq i} s_j$. Therefore, if $h_i(s_{-i})$ is sufficiently large and negative, there may be no feasible strategy which could be used by an agent whose consumption is bounded below – and equilibria may fail to exist because of this implied emptiness of the feasible strategy set. This suggests that if endowments are known[6] to be sufficiently large, and if a priori bounds can be placed on willingnesses-to-pay, then the existence of equilibrium can be insured in bounded consumption environments, but not otherwise.

The non-separability problem is less likely to have a general solution. The method of proof used above, finding K^* and constructing the messages explicitly, will not work because the optimal project size is not known independently of the distribution of income. Therefore there may be no way of arranging transfers in a Pareto optimum so as to satisfy simultaneously the definition of the mechanism and the optimality of each agent's required Nash equilibrium response. A rigorous treatment of existence in these environments is an open problem.

In order to achieve existence and optimality of Nash equilibria more generally, it may be necessary to increase the dimensionality of the strategy space, which is one-dimensional in the Groves–Ledyard system.

7.6.2. The Hurwicz model

Hurwicz [1976] considers an economy with N agents with one private good, one public good and a two-dimensional message space. The public good is produced from the private good at a constant marginal cost equal to one.

Each agent transmits to the center a message (or strategy), $s_i = (K_i, q_i) \in \mathbf{R} \times \mathbf{R}_+$, which contains the amount K_i of public good desired by the agent and a price which will be used in the computation of agent $i + 1$'s personalized tax for the public good. Let $s = (s_1, \ldots, s_N)$.

To each N-tuple of messages, s, the outcome function (or mechanism) suggested by Hurwicz associates a level of the public good and a vector of personalized taxes as follows:

[6] This raises the issue of concealment of endowments (see chapter 4).

$$K(s) = \frac{1}{N}\sum_i K_i \tag{7.71}$$

$$t_i(s) = -K(s)\bar{q}_i - q_i(K_i - K_{i+1})^2 + q_{i+1}(K_{i+1} - K_{i+2})^2 \quad \text{for}^7\ i = 1, \ldots, N \tag{7.72}$$

where

$$\bar{q}_i = \frac{1}{N} - \sum_{j=1}^{N^*} (-1)^j q_{i+j} \tag{7.73}$$

with

$$N^* = N - 1 \quad \text{if } N \text{ is odd}$$
$$= N - 2 \quad \text{if } N \text{ is even.}$$

The chosen quantity of public good is the average of the desired quantities. The tax imputed to agent i is computed from a personalized price $\bar{q}_i$ and is corrected by some quadratic terms in order to balance the government's budget.

For $n \geqslant 3$, Hurwicz shows that the mechanism is balanced for any N-tuple of strategies and that, if s^* is a Nash equilibrium in messages, the resulting allocation is a Lindahl equilibrium. Reciprocally, to any Lindahl equilibrium one can associate a Nash equilibrium in the messages. Therefore, under classical assumptions, there exists a Nash equilibrium in messages and any Nash equilibrium is a Pareto optimum.

This model examplifies the type of very strong results which can be obtained through the Nash approach. But one must be cautious lest the behavior of the system away from equilibrium be ignored. For this mechanism, the optimal strategy of an agent outside a Nash equilibrium is obviously $(K_i^*, 0)$, where K_i^* is the optimal level of public good given the personalized price imposed by the others, and $q_i^* = 0$ is chosen to avoid any payment due to the quadratic term $q_i(K_i - K_{i+1})^2$. At a Nash equilibrium, since $K_i = K^*$ for all i, the agents are all indifferent with respect to their choice of q_i. As is common in general equilibrium theory, it is implicitly assumed that the appropriate q_i^* is chosen. Therefore, the correspondence which associates to a specification of the $(N - 1)$ other agents' strategies the optimal strategy of agent i exhibits a severe lack of lower hemi-continuity exactly at the Nash equilibria. It therefore appears impossible to use the agents' optimizing behavior to converge to a Nash equilibrium under any dynamic process requiring only local information.

[7] With the convention $N + 1 = 1$ for indices.

A prerequisite for proving the stability of the dynamic system associated with a mechanism seems to be the continuity and uniqueness of the optimal strategy at every Nash equilibrium. Is it possible to extend the strategy space to obtain such a result and still have Hurwicz's equivalence between Lindahl equilibria and Nash equilibria?

7.6.3. The Hurwicz attainability theorem [1977]

Although there may be many mechanisms having Pareto optimal Nash equilibria, there is one further requirement that might be added – individual rationality. This is closely related to our discussion in the last chapter. We still do not have a precise characterization of the mechanisms satisfying all these requirements, but, nevertheless, Hurwicz has recently provided a strong theorem characterizing the outcomes attainable through any such process. The power of this result derives from the fact that no restrictions, except for continuity and to some extent convexity, are placed on either the outcome functions or the strategy space.

The Hurwicz theorem states that any mechanism that is individually rational and all of whose Nash equilibria are Pareto optimal must essentially coincide with the Lindahl equilibrium correspondence.

The Nash equilibrium approach to the revelation of preferences brings together two difficult problems which must be solved to yield meaningful mechanisms.

Assuming myopic behavior along the Nash revelation game, it is necessary to show the stability of this game. But if one were to relax the myopic assumption, the incentives for manipulation along the dynamic process leading to a Nash equilibrium would then be in need of close scrutiny. Both issues are extremely difficult and await further work.

PART III

Chapter 8

LARGE NUMBER ECONOMIES

8.1. Public goods and large numbers

Although the fact that public goods are consumed by large numbers of economic agents was recognized early in the theoretical study of resource allocation problems, this observation was almost never employed to derive positive results.[1] Indeed the size of the economy was viewed primarily as a barrier to the cooperative behavior that might help overcome the free-rider problem in its extreme form. In many other areas of economic theory, however, an appeal to the number of agents concerned has been fruitful. For example, the role of large numbers in incentives questions has been pointed out by Roberts and Postlewaite [1976], who have shown in an exchange economy with private goods that the competitive mechanism becomes individually incentive compatible asymptotically, in large economies. This part of the book is devoted to the study of the extent to which the existence of a large number of decision-making units helps to overcome the problems with dominant strategy mechanisms discussed in chapter 5.

Bowen [1943] was the first one, and the only one to our knowledge, to make use of the large number assumption in designing an allocation mechanism in an economy with public goods. It is worth recalling in our language some elements of his contribution. Bowen assumes, on the basis of some empirical evidence on expenditures for private goods, that the true willingnesses to pay for a given public project are distributed normally.

[1] There exists a literature (see J. Kelly [1974] for a summary) which examines how the probability of social intransitivity, under a simple majority vote decision rule, depends on the number of alternatives and the number of votes. Very recently, Peleg [1976b] has shown that a large class of voting schemes was asymptotically non-manipulable, in the sense that the probability of misrepresentation of preferences by individuals being profitable goes to zero as the number of voters goes to infinity.

Then, he imputes to each agent an equal share of the cost of the project and obtains a normal distribution of net willingnesses to pay. Agents vote for or against the realization of the public project and the outcome is determined by majority rule. The implicit large number assumption enables Bowen to use the convergence of the empirical median towards the empirical mean for symmetric distributions. For that reason, when the majority rule results in acceptance of the project, it means that a majority of agents have a positive net willingness to pay and therefore that the aggregate net willingness to pay is positive.

In this part of the book, we restrict ourselves for simplicity to the $\{0, 1\}$ project space, which can also be thought of as an incremental step in a variable size project, an interpretation that will be developed in detail in part IV.

The nonbalancedness of the government's budget and the cost of the induced monetary transfers, the lack of coalition incentive compatibility, and the need for a separability assumption on preferences are the principal problems we attack in this way. The main thrust of this part is that most of these weaknesses disappear in a large economy, when appropriate use and modifications of Groves mechanisms are made. Chapter 9 shows the sense in which the nonbalancedness is a negligible phenomenon. Chapter 10 proves a weak form of coalition incentive compatibility in large economies. Chapter 11 is of a more negative character. Unlike the other problems exposed in chapter 5, the lack of separability is particularly unyielding. We show that although somewhat better information can be gathered in large numbers economies, the imperfectness of the information cannot be completely overcome. Chapter 12 deals with the various direct or distributional costs associated with the mechanisms. Chapter 13 explores problems in the strength of incentives when agents must incur real economic costs to discover their own true tastes. Sampling is used as a method for increasing the losses from non-truth telling behavior, and thereby including private investment in information gathering. In a sense, chapter 13 is a result of the large numbers assumption, since the strength of incentives falls with the number of individuals in the system.

8.2. Diversity of tastes

A crucial assumption in the next chapters is that of diversity of tastes. Technically, the diversity of tastes is obtained by considering the population as a random sample from a given non-degenerate probability distribution. The assumptions underlying this procedure are essentially two: We assume that agents' preferences are independent from one another and that human nature implies diversity.

The extreme character of this first hypothesis can be largely reduced. It is easy to accomodate a limited dependence of tastes by considering a sample that is not formed of independent drawings from the same distribution. The law of large numbers that we need on several occasions is known to be robust to such limited dependence conditions. Similarly, the results we use on order statistics can be extended to take into account some limited form of dependence. To carry out the analysis in those more general terms would only confuse the exposition of the main ideas; a precise delimitation of the validity of our results when dependence of tastes is allowed for is left for further work. It is clear however that particular situations may lead to a breakdown of the major principles we develop. The extent to which preferences are socially determined, particularly for public goods, suggests that our assumption is not completely innocuous. Also, the reliance on outside information to ascertain ex ante preferences may lead to great similarity in stated willingnesses to pay if not in true ones. On the other hand the assumption of non-degeneracy of the underlying distribution of willingnesses to pay is easily accepted in a world of unequal incomes, even if one believes that with equal incomes individuals have essentially the same tastes.

8.3. The validity of asymptotic results

Another assumption that we use extensively is the boundedness of the variance of the distribution of tastes – an assumption of thin tailed distributions. Of course, we could have used distributions with bounded supports, but there is no obvious natural bound in many problems. This assumption should be considered as a technical one which does not restrict the nature of our asymptotic results, but which should nevertheless be kept in mind. We find in our simulations that a number of agents such as 100 is enough to use the asymptotic results; however, this is under the proviso of relatively thin tailed distributions.[2]

[2] It has been recognized that the pure public good case, in which all agents consume the public good in equal quantity is extreme (see the early discussion between Samuelson and Margolis in the *Review of Economics and Statistics* [1955]). In particular, public goods are often local in nature; they may concern only a subset of the population.

It is important to know precisely what we mean by a large number situation. This is a difficult question because it bears on the speed of convergence towards asymptotic results. Economic and statistical theories usually fail to produce definitive answers in this regard. We tentatively explored this question by using simulations. Because, as mentioned above, it appears that a population size near 100 is generally enough for a sound application of our asymptotic results, most local public goods as supplied by local governments appear to be covered by the theory developed in this part of the book. On the other hand, applications to group decision making procedures in smaller organizations such as firms and committees will not be able to avail themselves of such asymptotic properties under most circumstances.

8.4. Large numbers, incentives and personal information

The large number assumption as described above has a number of implications that we can now discuss.

First of all, in most cases, each agent is small relative to the rest of society. The probability that his answer will influence the social decision is, by nature, low; therefore, we cannot expect to find mechanisms for which the strength of the incentive is strong for each agent. This difficulty must be shared by any decision-making process with a large number of participants. It is surely a very positive result if a mechanism gives a non-zero incentive to tell the truth and does not merely result in indifference between this and lying. Such a small incentive can then be used to foster other moral and social reasons to tell the truth as discussed in chapter 1.

The above reasoning is completely correct only when there is perfect personal information. Let us be precise about this very important point. The purpose of mechanisms is to induce the transmission of correct information from economic agents to a decision maker. Up to now, we have assumed that these agents had the correct information in their possession. But, very often, agents have only a vague idea of their true willingness to pay, either because they do not know completely the characteristics of the public good offered, or because it provides new services for which agents do not know their tastes. Another possibility, as discussed in chapter 3, is the fact that although preferences in the given situation are known, the individual cannot predict his ultimate willingness to pay in the new economic environment corresponding to a shift in private goods prices after the decision has been taken. In these circumstances agents must

spend some real resources to improve their own knowledge of their willingness to pay. The question of the strength of the incentive to tell the truth then becomes crucial. In order to be effective, the mechanism has to overcome the resource cost incurred in the individual search for the truth. This problem will be the topic of chapter 13.

Another feature of a large number situation is that each agent is personally very ignorant about most of the others. In some contexts it will be natural to suppose that there are objective reasons for this uncertainty that are discernable by the planner. In other cases, we will identify the subjective uncertainty of each of the individuals with the uncertainty faced by the government itself.

8.5. The role of sampling

In addition to the personal costs that agents incur to discover their own information, there are many other costs that the decision maker faces to acquire information about the agents. They may take the form of asking, processing, or compiling. On the other hand, in most cases, the decision maker needs to have only a statistical knowledge of the population's tastes, such as the mean or the variance of the willingnesses to pay.

The idea then comes naturally to ask only a sample of the population and to use statistical theory to infer the desired information about the whole population. Here again, we can go back to Bowen [1943] for a clear understanding of the route to follow:

"If a poll is based on a representative sample of the population, and if the questions are put in the same way as if the entire citizenry was voting, the results can of course be interpreted in exactly the same way."

Decision makers often have informal information which they want to use in the procedure. It is then very tempting to represent the decision maker as a Bayesian statistician endowed with an a priori probability distribution over the tastes of the population. To improve his decision process, the Bayesian statistician observes a random sample of the population. From his prior and this new information, he then computes a posterior probability distribution over the parameters of the population's tastes on which he bases his decision. This will be the basic point of view taken in chapters 12 and 13.

Of course sampling introduces problems of its own. Most importantly, it means that there will be some sampling variance and thus, even if the sample's preferences could be elicited without error, the decision taken might not be optimal for the whole population. Our results on successfulness will thus have to be modified to be of a stochastic form. That is, either we take the truly optimal decision with a high probability or in contexts with a large set of alternatives we take the decision that is close to optimal with a high probability.

Other costs of sampling result from the contrast between the usefulness of having a large set of agents for some purpose with the difficulties encountered when the set of agents is too large. An example of this occurs in the conflict between the strength of incentives, which is kept high in small-group situations, and the problem of preventing coalition formation, which is effectively eliminated when large numbers are present.

Finally, the issue of voluntary participation must be carefully dealt with to avoid a biased sample (see chapter 6).

8.6. Random social choice

It is important to notice that by the introduction of sampling in the revelation mechanisms we enter a new field of social choice theory, namely the field of random social choice mechanisms. We already know that the use of randomness solves Arrow's impossibility theorem. A random dictator fulfills all of Arrow's conditions, but this is clearly unattractive on other grounds.

Following the lead of Zeckhauser [1969] [1973], Fishburn [1972] and Gibbard [1977] have made important contributions in the study of random social choice mechanisms. Gibbard [1977] has recently produced a characterization of all the voting random social choice mechanisms which are not manipulable – that is, those such that it is in the interest of no agent to vote according to false preferences. This family consists of probability mixtures of random dictators with decisions taken between two randomly selected alternatives. This additional prospect is not very attractive either. In addition, it is only probability mixtures of random-dictator mechanisms that can produce Pareto optimal outcomes with certainty.

As Gibbard-Satterthwaite's theorem, this negative result is obtained with no restriction on individual preferences. Is it possible to obtain positive results beyond the ones known in the deterministic case, by restricting the space of preferences to separable utility functions as we have done in chapters 3 and 4? Clearly, yes. Consider the mechanism in which all the N agents are asked their willingnesses to pay for a public good, knowing that the decision will be taken according to a Groves mechanism on a

random sample of size n. It is clearly non-manipulable. There is a probability n/N to be in the sample. If the agent is in the sample it is a situation analogous to the deterministic case with a Groves mechanism and the truth is a dominant strategy. If he is not in the sample, he is indifferent between the truth and any other answer. Having elicited the true preferences of n agents in this way, it is clear that no benefit is provided by having the $N - n$ other statements. Therefore, instead of eliciting all preferences and then randomizing to select a subset, we could just as well randomize first and then elicit the preferences only of those chosen. Indeed this would economize on the direct costs of operating the mechanism, at no cost of lost information.

The mechanism described in this way will be called a *sampling Groves mechanism*, or *SGM*. Clearly an *SGM* will not take a Pareto optimal decision with probability one, as long as $n < N$. But if the agents' tastes are statistically independent and if N is large, then a relatively small sample can still keep the probability of a successful result high.

These remarks motivate our detailed study of the *SGM* and its modifications[3] which will be pursued in this part of the book.

[3] One might conjecture that the sequence of *SGM*'s with increasing sample size is the only sequence of revelation mechanisms such that truthful revelation is obtained from the agents who are asked and at the same time the probability of giving a Pareto outcome strictly increases as n goes from 1 to N. At present, a characterization of the "stochastically satisfactory" mechanisms is beyond reach.

Chapter 9

ON THE IMBALANCE OF THE BUDGET

9.1. Introduction

We know from chapter 5 that there is no Groves mechanism for which the sum of transfers needed to run the mechanism, referred to as the surplus, is identically zero, i.e., zero for any N-tuple of answers.[1] The lack of closedness is, of course, quite bothersome if we view the mechanism in the context of general equilibrium theory.[2]

In the first half of this chapter, we take the point of view that "almost closedness" may be enough on practical grounds and we provide justifications for such a partial equilibrium approach.

The spirit of our analysis is the same as that used in the general equilibrium analysis of economies with indivisibilities or other non-convexities in which an approximate equilibrium concept is employed.[3]

This defense is based on the presence in public economics of a large number of agents concerned. We use a bounded rationality argument; suppose it is possible to show that the non-zero surplus (positive or negative) is "small". Then, it appears reasonable to assume that if we redistribute this surplus in the population when it is positive and if we collect it from the population when it is negative, the agents will neglect this additional transfer in devising their optimal answer. Moreover, if, as will be done several times in this part, we ask only a random sample of the population, it seems that this surplus may be transferred to the unsampled population without affecting the incentives of the sampled group.

In a fully general equilibrium setting there is a second type of analysis

[1] See theorem 5.3.

[2] One direction to restore closedness is to weaken the incentive properties required of the mechanism, see chapter 7.

[3] See for example Arrow and Hahn [1971].

which recognizes the possibility that agents take into account the redistribution of the surplus, but shows that, "in general", the good quality of the mechanism, successfulness, is not really affected. In the second half of this chapter, we assume that individuals respond by maximizing their expected utility, given their stochastic beliefs about the statements to be made by the other agents, and the knowledge that the pivotal mechanism will be used and the surplus redistributed. We show that their optimal strategies are approximately truthful in large economies and that the decision taken as a result almost always coincides with the true optimal decision.

9.2. The surplus for the Clarke mechanism

The analysis in this chapter will be conducted with the Clarke mechanism on the $\{0, 1\}$ project space, for which we know that the surplus is always positive.

We consider the willingnesses-to-pay of agents as independent drawings from an absolutely continuous probability distribution $F(\cdot)$ (see chapter 8). We are then in a position to compute the expected sum of the surplus. If the mechanism is used to evaluate many independent public projects, we may be satisfied if the expected sum is negligible. In addition, this result would be of interest if the mechanism is used for a single project and if agents' expectations of the others' evaluations correspond to the distribution $F(\cdot)$. Since their expected share of the surplus would be negligible, in such a case, the agents could neglect the non-closedness of the system when they answer. The force of these remarks would be further strengthened if the variance of the per capita surplus were negligible.

We consider a population of N agents. The index i will be omitted in sections 1–7 of this chapter since we will assume that all agents' beliefs are identical. Since $F(\cdot)$ is the distribution function of an agent's willingness-to-pay, let us denote $F_{N-1}(\cdot)$ the distribution function[4] of the sum, x, of the $(N-1)$ other agents.

From the definition of the Clarke mechanism, an agent is a pivot if and only if

$$|x| < |v| \quad \text{and} \quad v < 0 < x \quad \text{or} \quad x < 0 < v \tag{9.1}$$

[4] In chapters 5 and 6, the distribution of x was denoted $P(\cdot)$. In this chapter we use the notation $F_{N-1}(\cdot)$ to emphasize the fact that x is a sum of independent random variables.

where v denotes the agent's stated willingness-to-pay; such a pivotal agent must pay $|x|$. His expected pivotal payment is

$$\int_{\{x \mid |x| < |v| \text{ and } xv < 0\}} |x| \, \mathrm{d}F_{N-1}(x) \tag{9.2}$$

Then, the expected total surplus in the population, given beliefs $F(\cdot)$, is,

$$E_N = N \int_{-\infty}^{+\infty} \Big[\int_{\{x \mid |x| < |v| \text{ and } xv < 0\}} |x| \, \mathrm{d}F_{N-1}(x) \Big] \mathrm{d}F(v) \tag{9.3}$$

The variance of the surplus is accordingly:

$$V_N = N \int_{-\infty}^{+\infty} \Big[\int_{\{x \mid |x| < |v| \text{ and } xv < 0\}} x^2 \, \mathrm{d}F_{N-1}(x) \Big] \mathrm{d}F(v) - E_N^2 \tag{9.4}$$

Assuming that the surplus is redistributed equally, the individual expected share of the surplus is $e_N = E_N/N$ and the variance of the individual share of the surplus is $v_N = V_N/N^2$.

9.3. The case of normally distributed statements

Following Bowen [1943], let us assume that $F(\cdot)$ is a normal distribution. Consider first the case of a zero mean and let us normalize the variance to 1. Then,

$$\begin{aligned} E_N &= 2N \int_{-\infty}^{0} \Big[\int_0^{-v} x \, \mathrm{d}F_{N-1}(x) \Big] \mathrm{d}F(v) \\ &= \frac{N}{\pi\sqrt{(N-1)}} \int_{-\infty}^{0} \Big[\int_0^{-v} x \mathrm{e}^{-\frac{x^2}{2(N-1)}} \mathrm{d}x \Big] \mathrm{e}^{-\frac{v^2}{2}} \mathrm{d}v \\ &= \frac{N\sqrt{(N-1)}}{\pi} \int_{-\infty}^{0} \Big(1 - \mathrm{e}^{-\frac{v^2}{2(N-1)}} \Big) \mathrm{e}^{-\frac{v^2}{2}} \mathrm{d}v \\ &= \frac{1}{\sqrt{(2\pi)}} [N\sqrt{(N-1)} - (N-1)\sqrt{(N)}] \end{aligned} \tag{9.5}$$

Using, for large N, the approximation

$$\sqrt{(N-1)} \approx \sqrt{(N)} - \frac{1}{2\sqrt{(N)}} - \frac{1}{8N\sqrt{(N)}} \tag{9.6}$$

we obtain:

$$E_N = \frac{1}{2\sqrt{(2\pi)}}\left[\sqrt{(N)} - \frac{1}{4\sqrt{(N)}} + \dots\right] \tag{9.7}$$

The expected surplus increases as the square root of the population size; therefore, the per capita expected surplus e_N goes to zero as $1/\sqrt{(N)}$.

The second moment M_N^2 of the surplus is

$$M_N^2 = N\int_{-\infty}^{+\infty}\Big[\int_{\{x\,|\,|x|<|v| \text{ and } xv<0\}} x^2\,\mathrm{d}F_{N-1}(x)\Big]\mathrm{d}F(v)$$

$$= \frac{N}{\pi\sqrt{(N-1)}}\int_{-\infty}^{0}\left[\int_{0}^{-v} x^2 \mathrm{e}^{-\frac{x^2}{2(N-1)}}\mathrm{d}x\right]\mathrm{e}^{-\frac{v^2}{2}}\mathrm{d}v$$

$$= \frac{N}{\pi\sqrt{(N-1)}}\left[\int_{-\infty}^{0}(N-1)v\,\mathrm{e}^{-\frac{v^2}{2}\left(\frac{N}{N-1}\right)}\mathrm{d}v\right.$$

$$\left. + (N-1)\sqrt{(N-1)}\int_{-\infty}^{0}\left(\int_{0}^{-\frac{v}{\sqrt{(N-1)}}}\mathrm{e}^{-\frac{w^2}{2}}\mathrm{d}w\right)\mathrm{e}^{-\frac{v^2}{2}}\mathrm{d}v\right]$$

$$\leqslant \frac{N(N-1)}{\sqrt{(N-1)}}\alpha + N(N-1)\beta(N) \tag{9.8}$$

where α is a constant independent of N and $\beta(N)$ goes to zero as N increases to infinity.

Consequently, the variance of the individual share of the surplus goes to zero as N increases to infinity; the individual share of the surplus converges in quadratic mean to zero.

Let us now give an example based on the above computations. Suppose that the distribution of tastes can be approximated by a normal distribution of mean zero and standard deviation \$100, i.e., that 95% of the population has a willingness-to-pay between −\$200 and +\$200.

The expected per capita share of the surplus according to (9.7) is:

$$E_N \approx \frac{100}{2\sqrt{(2\pi)}}\left[\frac{1}{\sqrt{(N)}} - \frac{1}{4N}\right] \tag{9.9}$$

For a population of 100, E_N is of the order of 2 dollars and of the order of 20 cents for a population size of 10,000.

When the mean of the distribution $F(\cdot)$ is non-zero, results are much stronger: the expected total surplus goes to zero exponentially as N increases. Simulations have been performed in the normal case for different values of μ and the results are reported in table 9.1 and in figure 9.1.

Table 9.1. Normal distribution.

	N						
μ	10	20	50	100	200	500	1000
0	0.61	0.89	$0.13\cdot10^{1}$	$0.18\cdot10^{1}$	$0.36\cdot10^{1}$	$0.75\cdot10^{1}$	$0.87\cdot10^{1}$
0.5	0.18	$0.77\cdot10^{-1}$	$0.27\cdot10^{-2}$	$0.75\cdot10^{-5}$	$0.53\cdot10^{-10}$	$0.57\cdot10^{-26}$	$0.48\cdot10^{-53}$
1	$0.53\cdot10^{-2}$	$0.52\cdot10^{-4}$	$0.24\cdot10^{-10}$	$0.48\cdot10^{-21}$	$0.17\cdot10^{-42}$	$0.25\cdot10^{-107}$	$0.82\cdot10^{-216}$
1.5	$0.14\cdot10^{-4}$	$0.27\cdot10^{-9}$	$0.94\cdot10^{-24}$	$0.52\cdot10^{-48}$	$0.12\cdot10^{-96}$	$0.65\cdot10^{-243}$	0
2	$0.41\cdot10^{-8}$	$0.10\cdot10^{-16}$	$0.16\cdot10^{-42}$	$0.95\cdot10^{-86}$	$0.22\cdot10^{-172}$	0	0

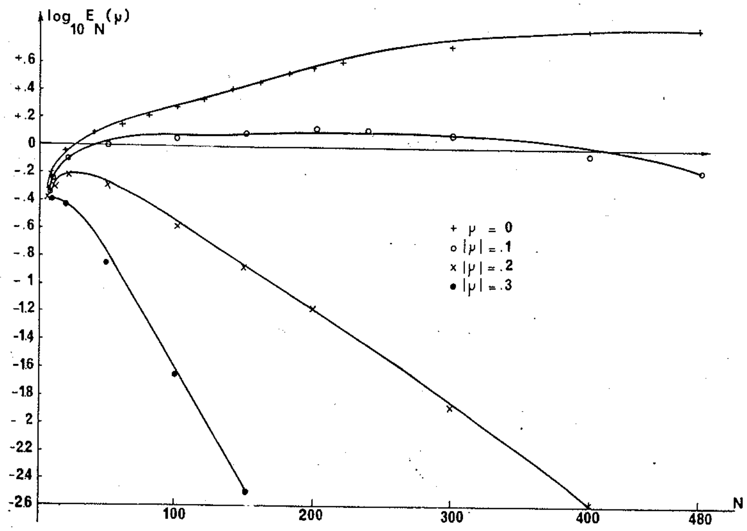

Figure 9.1. Normal distribution: $E_N(\mu)$ for various N and μ

Below, we give a brief proof of this result in the normal case (see theorem 9.1 below for a general proof).

$$E_N(\mu) = 2N \int_{-\infty}^{0} \left[\int_{-v}^{0} x \frac{1}{\sqrt{(2\pi(N-1))}} e^{-\frac{(x-(N-1)\mu)^2}{2(N-1)}} dx \right] \frac{e^{-\frac{v^2}{2}}}{\sqrt{(2\pi)}} dv \tag{9.10}$$

Let us use the change of variable: $\tilde{x} = (x - (N-1)\mu)/\sqrt{(N-1)}$; then,

$$E_N(\mu) = 2N \int_{-\infty}^{0} \left[\int_{-\sqrt{(N-1)}\mu}^{-\sqrt{(N-1)}\mu - \frac{v}{\sqrt{(N-1)}}} [\sqrt{(N-1)}\tilde{x} + (N-1)\mu] e^{-\frac{u^2}{2}} d\tilde{x} \right] \frac{e^{-\frac{v^2}{2}}}{\sqrt{(2\pi)}} dv \tag{9.11}$$

Then, using L'Hospital's rule and the Lebesgue theorem, it is easy to see that $N^{\alpha}E_N(\mu)$ goes to zero as N goes to infinity for any α.

9.4. The case of uniformly distributed statements

The case of normality may be criticized on the ground that it allows unbounded evaluations. For that reason we have dealt, as an example, with the case of a uniform distribution.

If $F(\cdot)$ is the uniform distribution on $[-1, +1]$, let $f_{N-1}(\cdot)$ denote the probability density function of the sum x of the $N-1$ other agents.

Then,

$$E_N = -2N \textstyle\int_{x=-1}^{x=0} \int_{z=-x}^{z=1} x \tfrac{1}{2} dz\, f_{N-1}(dx) \tag{9.11}$$

This expression may be computed using the distribution of the sum of $N-1$ independent uniform random variables (see Renyi [1966]):

$$E_N = \frac{-N}{2(N-2)!} \int_{x=-1}^{x=0} \left(\int_{z=-x}^{z=1} dz \right) \sum_{j=0}^{\left[\frac{x+N-1}{2}\right]} x(-1)^j C_{N-1}^j \left(\frac{x+N-1}{2} - j \right)^{N-2} dx \tag{9.12}$$

$$= \frac{-N}{2(N-2)!} \int_{x=-1}^{x=0} \sum_{j=0}^{\left[\frac{x+N-1}{2}\right]} x(1+x)(-1)^j C_{N-1}^j \left(\frac{x+N-1}{2} - j \right)^{N-2} dz$$

where

$$\left[\frac{x+N-1}{2} \right] = \frac{N}{2} - 1 \quad \text{if } N \text{ is even}$$
$$= \frac{N-1}{2} - 1 \quad \text{if } N \text{ is odd.}$$

and

$$C_{N-1}^j = \frac{(N-1)!}{j!(N-1-j)!}$$

For any x in $[-1, 0]$, we can permute the signs $\int$ and $\sum$ and we obtain:

$$E_N = \frac{1}{2^{N-1}(N-1)!} \sum_{j=0}^{N^*} (-1)^j C_{N-1}^j [(N-1-2j)^N + (N-2-2j)^N + \frac{2}{N+1}[(N-2-2j)^{N+1} - (N-1-2j)^{N+1}]] \tag{9.13}$$

with

$$N^* = \frac{N}{2} - 1 \quad \text{if } N \text{ is even}$$
$$= \frac{N-1}{2} - 1 \quad \text{if } N \text{ is odd.}$$

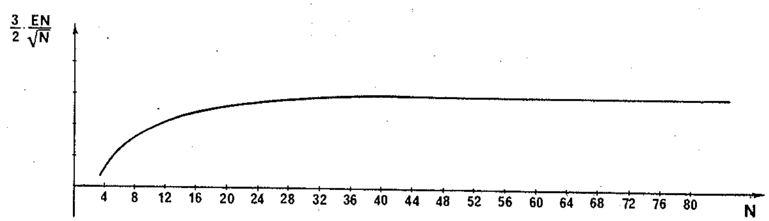

Figure 9.2. Uniform Distribution on [−1, +1]

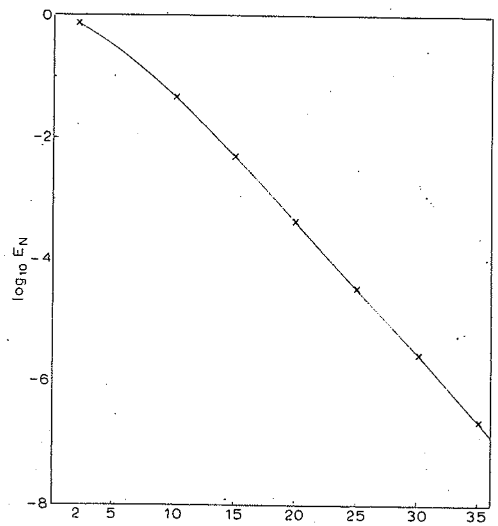

Figure 9.3. Uniform Distribution on [−1, +2].

A numerical evaluation of (9.13) is reported on figure 9.2. A simulation has also been performed for the case of a uniform distribution on [−1, +2], that is, with a mean different from zero. The exponential convergence to zero is obtained (see figure 9.3), as in the normal case.

9.5. General distributions of statements

Before interpreting these results, let us first show that they are quite general.

Theorem 9.1. If the distribution function $F(\cdot)$ is absolutely continuous with mean m and bounded variance (normalized to 1) and if, for $m = 0$,

F has a continuous density function $f(\cdot)$ with a unique mode equal to zero, then

$$\lim_{N\to\infty} \frac{E_N}{\sqrt{(N)}} = \frac{1}{2\sqrt{(2\pi)}}$$

If $m \neq 0$ and $N^\alpha f_N(x)$ converges uniformly towards zero for all $\alpha > 0$ when N goes to infinity, then

$$\lim_{N\to\infty} N^\alpha E_N = 0$$

for all $\alpha > 0$.

Proof. Consider first the case $m = 0$. We want to evaluate the limit as N goes to infinity of:

$$\frac{E_N}{\sqrt{(N)}} = \sqrt{(N)} \int_{-\infty}^{+\infty} \mathrm{d}F(v) \int_{\{x \mid |x| < |v| \text{ and } xv < 0\}} |x| \,\mathrm{d}F_{N-1}(x) \tag{9.14}$$

Let $y = x/\sqrt{(N-1)}$ and let $G_{N-1}(\cdot)$ be the distribution function of y. From the Lindeberg–Levy theorem, $G_{N-1}(\cdot)$ converges at each point towards the normal distribution function which is absolutely continuous. Consequently, the density function $g_{N-1}(\cdot)$ converges almost everywhere towards the normal density function. Observe that the maximum of $g_{N-1}(\cdot)$ is always at zero. Since at 0, $g_{N-1}(0)$ converges to $1/\sqrt{(2\pi)}$, the sequence of functions $g_{N-1}(\cdot)$ is uniformly bounded by a constant A.[5]

$$\begin{aligned}\frac{E_N}{\sqrt{(N)}} &= \sqrt{(N)} \int_{-\infty}^{0} f(v)\mathrm{d}v \int_{0}^{-v} x f_{N-1}(x)\mathrm{d}x - \sqrt{(N)} \int_{0}^{+\infty} f(v)\mathrm{d}v \int_{-v}^{0} x f_{N-1}(x)\mathrm{d}x \\ &= \sqrt{\left(\frac{N}{N-1}\right)} \int_{-\infty}^{0} f(v)\mathrm{d}v \int_{0}^{-v} x g_{N-1}\left(\frac{x}{\sqrt{(N-1)}}\right)\mathrm{d}x \\ &\quad + \sqrt{\left(\frac{N}{N-1}\right)} \int_{0}^{\infty} f(v)\mathrm{d}v \int_{-v}^{0} (-x) g_{N-1}\left(\frac{x}{\sqrt{(N-1)}}\right)\mathrm{d}x\end{aligned} \tag{9.15}$$

Since

$$\left| f(v)\, x\, g_{N-1}\left(\frac{x}{\sqrt{(N-1)}}\right)\right| < A\,|f(v)x| \tag{9.16}$$

[5] This is indeed the crucial condition which could be obtained with other assumptions.

and

$$\int_{-\infty}^{0} f(v)\,\mathrm{d}v \int_{0}^{-v} x\,\mathrm{d}x + \int_{0}^{\infty} f(v)\,\mathrm{d}v \int_{-v}^{0} (-x)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{v^2}{2} f(v)\,\mathrm{d}v = \tfrac{1}{2} \tag{9.17}$$

From the Lebesgue theorem, we have:

$$\begin{aligned}\lim_{N\to\infty} \frac{E_N}{\sqrt{(N)}} &= \int_{-\infty}^{0} f(v)\,\mathrm{d}v \int_{0}^{-v} x \lim_{N} g_{N-1}\left(\frac{x}{\sqrt{(N-1)}}\right)\mathrm{d}x \\ &\quad + \int_{0}^{\infty} f(v)\,\mathrm{d}v \int_{-v}^{0} (-x) \lim_{N} g_{N-1}\left(\frac{x}{\sqrt{(N-1)}}\right)\mathrm{d}x \\ &= \frac{1}{2\sqrt{(2\pi)}}\end{aligned} \tag{9.18}$$

since

$$\lim_{N} g_{N-1}\left(\frac{x}{\sqrt{(N-1)}}\right) = \frac{1}{\sqrt{(2\pi)}} \text{ for all } x.$$

Consider now the case of $m \neq 0$.

Since $N^{\alpha} f_{N-1}(x)$ is uniformly bounded by A, the function $|N^{\alpha} x f_{N-1}(x) f(v)|$ is bounded by $|A x f(v)|$ which is integrable over the domains $\{x \mid 0 > x > -v, 0 < v < \infty\}$ and $\{x \mid 0 < x < -v, 0 > v > -\infty\}$ since $F(\cdot)$ has a bounded variance. A direct application of the Lebesgue theorem gives the desired result. Q.E.D.

The results obtained in this section are indeed quite general. The assumption of a unique mode at zero is convenient and not very restrictive for our problem but could be replaced by more technical and more general conditions to yield the uniform boundedness of $g_{N-1}(\cdot)$.

Therefore, very generally, the per capita transfer is of the order of $(0.2/\sqrt{N})\sigma$, where σ is the standard deviation of the distribution for $m = 0$. When $m \neq 0$, the results are much stronger and the total transfers are negligible. However, it is clear that the most relevant problem corresponds to $m = 0$. In this case, there is a genuine uncertainty about what the tastes of the population are. If the expectations of the government are described by a distribution with mean $m \neq 0$, there is hardly the necessity of asking individuals, since with a large population the sign of the sum of willingness-to-pay will be almost surely the same as the sign of m.

Similarly, a general proof that the variance of the per capita share goes to zero as N goes to infinity could be given by assuming that the third moment of $F(\cdot)$ exists.

If, on the other hand, we interpret $F(\cdot)$ as an approximation to the empirical distribution in the population, these results tell us that in most cases the transfers are negligible and that, even in the exceptional cases corresponding to $m = 0$, the per capita transfer will be negligible.

The assumption that agents' willingnesses-to-pay are independent random drawings from a given distribution is of course important; but we know that we could admit a limited amount of dependence while retaining the validity of the central limit theorem.[6]

Extreme cases of dependence, however, cannot be accepted.

If we have a population for which

$$\begin{aligned} v &= -1 \quad \text{for} \quad N/2 \text{ agents} \\ &= +1 \quad \text{for} \quad N/2 \text{ agents} \end{aligned}$$

all agents with $v = +1$ are pivotal and must pay 1; the total transfers are then $N/2$, and the per capita transfer, 1/2, is important and non-decreasing with N.

9.6. Asymptotic dominance

The interpretation of the asymptotic feasibility results given above follows the spirit of the approximation results found in a variety of contexts in Walrasian general equilibrium theory.[7] That is, if an imbalance becomes negligible on a per capita basis in a large economy, then the system is approximately in equilibrium. All measurements, both physical and economic, being imprecise, we can therefore regard the system as being in a true, exact equilibrium.

In the current model, that is the case. The taxes collected are negligible on a per capita basis and the dominant strategy property insures that truthful responses are always being given. However, it can be argued that we should not be completely satisfied with such an approximation in the present context. Unlike the case of individuals demanding private goods at prices which depart slightly from a true equilibrium, in the present model the incentive to behave optimally grows progressively weaker with

[6] See Loève [1963, ch. 8] for a presentation of these more general conditions.
[7] For example Malinvaud [1972b] and Hildenbrand [1974].

the size of the economy. (This intuitive result is made explicit in chapter 13.) This raises the following query: Suppose that the taxes collected were rebated equally to all individuals. Would the slight distortion that this introduces outweigh the weak incentive for truthtelling and give rise to responses that depart markedly from the truth?

In order to provide a meaningful answer to this question we must specify the expectations each individual holds concerning the distribution of stated willingnesses-to-pay as expressed by the other people. The rebates, which insure the exact feasibility of the process, destroy the dominant strategy property. To keep the analysis parallel to that of the previous sections, we assume that each of the individuals in the society believes that all of the others are drawn independently from a normal population with zero mean. We will then calculate his optimal stated willingness-to-pay and show that this approaches the truth as the economy grows.

It should be pointed out that this is a much stronger result than that obtained in the earlier sections. It is not enough that per capita expected rebates are going to zero. The potential regret for distorting one's answers is also decreasing since it is progressively less likely that each individual is pivotal. One must weigh the expected *change* in the rebate to be received against this regret, and must show that the latter is the stronger effect. As both the rebate and the regret are approaching zero on the order of $N^{-1/2}$, a more refined analysis is necessary before the optimal response can be determined.

Let us first compute the expected utility of any response, w, given that decisions will be taken and taxes will be computed using the pivotal mechanism, and that equal rebates will then be made to all individuals. For simplicity, we will concentrate on the case in which $w \geqslant 0$, the case of $w \leqslant 0$ being symmetric.

Each agent assumes that the sum, x, of the $N-1$ other agents is distributed normally with mean zero and variance $N-1$ (without loss of generality). Then the expected utility derived from the project itself is

$$v\int_{-w}^{\infty} dF_{N-1}(x) \tag{9.19}$$

Since $1/N$ of one's own taxes are rebated, the expected payments net of this rebate are

$$\frac{N-1}{N}\int_{-w}^{0} x\, dF_{N-1}(x) \tag{9.20}$$

because the individual is pivotal (when $w > 0$) if and only if $-w < x < 0$.

Finally, we have to compute the expected rebates from the $N-1$ other pivotal payments. The expected payments by these individuals, if they are pivotal in the affirmative, are

$$(N-1)\int_{-\infty}^{-w} dF_{N-2}(x)\int_{-w-x}^{\infty}(w+x)\,dF(z) \tag{9.21}$$

where $F_{N-2}(x)$ is the distribution of the sum of $N-2$ random variables each of which is distributed normally with zero mean and unit variance. From the individual's point of view, his statement, w, is known, and only the other $N-1$ are random. This explains the asymmetry of this expression about zero. Similarly, expected total payments by pivotal individuals who have defeated the public project are

$$-(N-1)\int_{-w}^{\infty} dF_{N-2}(x)\int_{-\infty}^{-w-x}(w+x)\,dF(z) \tag{9.22}$$

Therefore, since the individual is entitled to $1/N$ of these as a subsidy, his expected utility from this source is

$$\frac{N-1}{N}\left[\int_{-w}^{\infty} dF_{N-2}(x)\int_{-\infty}^{-w-x}(w+x)\,dF(z) - \int_{-\infty}^{-w} dF_{N-2}(x)\int_{-w-x}^{\infty}(w+x)\,dF(z)\right] \tag{9.23}$$

The sum of (9.19) and (9.23) minus (9.20) is the expected utility of the response $w > 0$.

It is easy to see that because $1/N$ of one's own taxes are rebated, the optimal response considering only (9.19)–(9.20) is $[N/(N-1)]v$. This is true independent of one's beliefs about the distribution of responses by the others. However, considering the fact that one's share in the rebates of others' pivotal payments is partially determined by one's own statement, the optimal answer may be somewhat modified.

Consider the case of three individuals with v's equal to 1, -4, and 2, respectively. The first individual is not a pivot. He receives a subsidy of 1 which is his share of the second individual's pivotal payment. By saying $1\frac{1}{2}$, however, he can increase his subsidy to $3\frac{1}{6}$; any statement less than 2 will cause the second individual to be pivotal by a larger amount, and hence will increase his payment. However, if he follows this course, but is ignorant of the actual statements to be made by the others, then he may be hurt by this exaggeration. For example, if the third individual would say 3 instead of 2, then by saying $1\frac{1}{2}$ he has made himself pivotal and must pay a tax which would have been avoided by a truthful response. The larger the variability of the sum of everyone else's answers, the more risky it becomes to try and increase one's share in total taxes collected by distorting one's response. For uncertain beliefs about the others' tastes that are symmetric around zero, it can be shown that this consideration induces a response closer to zero than $[N/(N-1)]v$.

We will now argue that $w > 0$ is the relevant case to consider when $v > 0$ and by symmetry the optimum will be characterized by $w < 0$ when $v < 0$. The original mechanism, which has a payoff function of (9.19)–(9.20) $\times$ $[N/(N-1)]$ is optimized at $w = v$ by the dominant strategy property. The additional term of

$$-\frac{1}{N}\int_{-w}^{0} x\,\mathrm{d}F_{N-1}(x) \tag{9.24}$$

which enters into (9.20), represents the expected rebate on one's own pivotal payments. It is clearly a symmetric function around zero which achieves its minimum there and is monotone on either side of this point. Therefore, the maximum of (9.19) and (9.20) must be at a point w^* which has the same sign as v and is greater in absolute value.

The value of the expression (9.23), on the other hand, is maximized at zero and is a symmetric function which is monotone decreasing on its positive branch and increasing on its negative branch. Therefore, the maximal expected utility will be attained between w^* and zero. Therefore, when $v > 0$, the relevant maximand is the expected utility indicated above. Furthermore, the remarks above indicate that the maximum will be attained on the interior of the interval $[0, w^*]$, and will be characterized by a zero of the derivative at that point.

This derivative can be calculated to be

$$\begin{aligned} v f_{N-1}(-w) &- \frac{N-1}{N} w f_{N-1}(-w) \\ &+ \frac{N-1}{N}\left[\int_{w}^{\infty} \mathrm{d}F_{N-2}(x)\left[\int_{-\infty}^{-w-x} \mathrm{d}F(z) - (w+x)f(-w-x)\right]\right] \\ &- \frac{N-1}{N}\left[\int_{-\infty}^{-w} \mathrm{d}F_{N-2}(x)\left[(w+x)f(-w-x) + \int_{-w-x}^{\infty} \mathrm{d}F(z)\right]\right] \end{aligned} \tag{9.25}$$

that is

$$\begin{aligned} v f_{N-1}(-w) &- \frac{N-1}{N} w f_{N-1}(-w) + \frac{N-1}{N}\int_{-\infty}^{\infty} \mathrm{d}F_{N-2}(x)\int_{-\infty}^{-w-x} \mathrm{d}F(z) \\ &- \frac{N-1}{N}\int_{-\infty}^{-w} \mathrm{d}F_{N-2}(x) \\ &- \frac{N-1}{N}\int_{-\infty}^{\infty} \mathrm{d}F_{N-2}(x)(w+x)f(-w-x) \end{aligned} \tag{9.26}$$

We calculate (9.26) in the case of normally distributed individual willingness-to-pay, with zero mean and unit variance. In this case, let $[(N-1)/N]$ $g_N(w)$ denote the last term of (9.26). Then

$$g_N(w) = \frac{1}{\sqrt{(2\pi(N-2))}}\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{x^2}{2(N-2)}}(w+x)\mathrm{e}^{-\frac{(w+x)^2}{2}}\,\mathrm{d}x \tag{9.27}$$

But

$$\frac{x^2}{N-2} + (w+x)^2 = \frac{N-1}{N-2}\left(x + \frac{N-2}{N-1}w\right)^2 + \frac{w^2}{N-1} \tag{9.28}$$

so that

$$\begin{aligned} g_N(w) &= \frac{1}{\sqrt{(2\pi(N-2))}\sqrt{(2\pi)}}\mathrm{e}^{-w^2/(N-1)}\int_{-\infty}^{\infty} \mathrm{e}^{\frac{-\left(x+\frac{N-2}{(N-1)}w\right)^2}{(N-2)/(N-1)}} \\ &\quad \left(x + \frac{N-2}{N-1}w + \frac{w}{N-1}\right)\mathrm{d}x \\ &= \frac{1}{\sqrt{(2\pi(N-2))}\sqrt{(2\pi)}}\cdot \mathrm{e}^{-w^2/2(N-1)}\cdot\frac{w}{(N-1)}\sqrt{\left(2\pi\frac{N-2}{N-1}\right)} \end{aligned} \tag{9.29}$$

Clearly,

$$g_N(w)\sqrt{N} \text{ converges to zero uniformly in } [0, w^*]. \tag{9.30}$$

Denote now the sum of the third and fourth terms of (9.26) by $[(N-1)/N]$ $h_N(w)$. Then

$$\frac{d}{dw}h_N(w) = -\frac{1}{\sqrt{(2\pi(N-2))}\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2(N-2)}} e^{-\frac{(w+x)^2}{2}}\, dx$$

$$+\frac{1}{\sqrt{(2\pi(N-2))}}e^{-\frac{w^2}{2(N-2)}}$$

$$= -\frac{1}{\sqrt{(2\pi(N-1))}}e^{-\frac{w^2}{2(N-1)}} + \frac{1}{\sqrt{(2\pi(N-2))}}e^{-\frac{w^2}{2(N-2)}} \qquad (9.31)$$

It is clear that $\sqrt{(N)}(d/dw)h_N(w)$ converges uniformly to zero in $[0, w^*]$. Since $h_N(0) = \frac{1}{2} - \frac{1}{2} = 0$, it follows that

$$\sqrt{(N)}h_N(w) \text{ converges uniformly to zero in } [0, w^*]. \qquad (9.32)$$

Let w_N be the maximum point of the expected utility (9.19)–(9.20) + (9.23). Then, setting the derivative (9.26) equal to zero, we get

$$\left(v - \frac{(N-1)}{N}w_N\right)\frac{1}{\sqrt{(2\pi(N-1))}}e^{-\frac{w_N^2}{2(N-1)}}$$

$$+\frac{N-1}{N+1}h_N(w_N) - \frac{N-1}{N}g_N(w_N) = 0 \qquad (9.33)$$

Multiplying this equation by $\sqrt{(N)}$ and applying (9.30) and (9.32), we get that w_N converges to v, as was to be shown.

This discussion can be summarized by the following theorem.[8]

[8] These results do not subsist when m is different from zero. Intuitively the reason is the following. Suppose that m is positive and consider an agent with a negative true willingness-to-pay. Then, in a large population the agent is almost sure that he cannot change the decision (in particular for distributions with bounded supports) and that he will not be pivotal himself, whatever he says. However, he may distort his answer downwards in an attempt to increase the number of positive pivotal agents and hence his rebate. Of course, the expected gain of doing that is quite small because almost nobody will be pivotal, but it turns out that with the normal distribution the gain is larger than the loss, due to the increase in his own expected pivotal payments.

Lengthy computations, not reported here, show that the problem is asymptotically locally concave around the solution of the asymptotic first order condition which is:

$$v - \frac{\left(e^{\frac{m^2}{2}} - 1\right)}{m} = \frac{e^{mw}}{m} + w$$

Clearly, there is not convergence towards the truth. For given v, and for m small, the expression above can be approximated by

$$w = \frac{v}{2} - \frac{1}{2}\left[\frac{m}{2} + \frac{1}{m}\right].$$

If m is small enough and negative, w is large and positive; if m is small enough and positive, w is large and negative, as suggested by the discussion above. For example, if $v > 0$ and $m > 0$, a reasonable case, the agent underreports his willingness-to-pay by first halving his true willingness-to-pay and then by subtracting a function of m.

Theorem 9.2. If an agent believes that the stated willingnesses-to-pay of the $N - 1$ other members of the population are independent, normally distributed random variables with zero mean, then his optimal strategy $w_N(v)$, for the pivotal mechanism with equal rebates of the surplus, converges to v for large N.

9.7. Asymptotic truthfulness without symmetry assumptions

Among the hypotheses of the last section, perhaps the least attractive is the condition that the distribution of willingnesses-to-pay in the population be symmetric. Normality of the normalized sum of strategies of N agents is useful as an approximation for large N, but the symmetry of the distribution for a given individual cannot be defended on such grounds. Symmetry is the assumption that plays the role in the statement just preceding (9.32): $h_N(0) = 0$, for all N; the normality of $F(\cdot)$ is not necessary. Failure of the symmetry would require that the asymptotic properties and speed of convergence of $h_N(0)$ be investigated directly.

In this section we demonstrate that the results on asymptotic truthfulness do not require this condition in an essential way.

Let $F_N(\cdot)$ have compact support, B, and let $f_N'(\cdot)$ denote the derivative of the density $f_N(\cdot)$, which is assumed to exist.

Theorem 9.3. For a distribution $f(\cdot)$ with compact support if
(i) $\sqrt{(N)}f_N(z)$ is uniformly bounded;
(ii) $\sqrt{(N)}f_N(z)$ converges uniformly to a constant $c > 0$ on any compact set as N goes to infinity, the optimal answer converges to the truth as N goes to infinity.

Proof. The expected utility of an answer w when the truth is v is

$$\int_{-w}^{\infty} v\, dF_{N-1}(x) + \frac{N-1}{N}\int_{-w}^{0} x\, dF_{N-1}(x)$$

$$+\left(\frac{N-1}{N}\right)\left[\int_{-w}^{\infty} \mathrm{d}F_{N-2}(x) \int_{-\infty}^{-w-x} (w+x)\,\mathrm{d}F(z)\right.$$
$$\left.- \int_{-\infty}^{-w} \mathrm{d}F_{N-2}(x) \int_{-w-x}^{\infty} (w+x)\,\mathrm{d}F(z)\right] \tag{9.34}$$

We want to show first that for each N, chosen sufficiently large, the maximum of (9.34) exists.

Consider the first term. It is clearly bounded by $v_{\max}$ independently of N, where $v_{\max} = \max_B v$. The second term, for $|w|$ large, is negative and of the order of $-\sqrt{(N)}$. The third term is positive and goes to zero with $|w|$ by virtue of the condition that $x f_N(x)$ approaches zero for N large enough.

Combining these observations, we see that (9.34) will be made large and negative by the choice of $|w|$ large, when N is large. Since for $w = 0$ the first term is bounded below and the third is positive while the second is zero, the response of $w = 0$ will dominate that of $|w|$ large. Thus the range of responses can be restricted to a compact set and the maximizing answer satisfies the first-order condition.

$$\left[v - \frac{N-1}{N} w\right]\sqrt{(N)} f_{N-1}(-w) + \sqrt{(N)} h_N(w) + \sqrt{(N)} g_N(w) = 0 \tag{9.35}$$

where

$$h_N(w) = \frac{N-1}{N} \int_{-\infty}^{+\infty} \mathrm{d}F_{N-2}(x) \int_{-\infty}^{-w-x} \mathrm{d}F(z) - \int_{-\infty}^{-w} \mathrm{d}F_{N-2}(x) \tag{9.36}$$

$$g_N(w) = -\frac{N-1}{N} \int_{-\infty}^{+\infty} (w+x) f(-w-x) f_{N-2}(x)\,\mathrm{d}x \tag{9.37}$$

We show now that any solution of the first-order condition (9.35) converges to the truth as N grows.

$$h'_N(w) = -\left(\frac{N-1}{N}\right)\left[\int_{-\infty}^{+\infty} f(-w-x) f_{N-2}(x)\,\mathrm{d}x - f_{N-2}(-w)\right] \tag{9.38}$$

$$\sqrt{(N)} h'_N(w) = -\left(\frac{N-1}{N}\right) \int_{-\infty}^{+\infty} [\sqrt{(N)} f_{N-2}(x) -$$
$$- \sqrt{(N)} f_{N-2}(-w)]\,\mathrm{d}F(-w-x) \tag{9.39}$$

Since $\sqrt{(N)}[f_{N-2}(x) - f_{N-2}(-w)]$ is uniformly bounded from (*i*), we can apply Lebesgue's theorem and since $\sqrt{(N)} f_{N-2}(x)$ converges pointwise to a constant c for any x, from (ii) we have that $\sqrt{(N)} h'_N(w)$ converges to zero on any compact set of values for w.

Therefore, since $\lim_{N\to\infty} \sqrt{(N)} h_N(0) = 0$, we conclude that $\sqrt{(N)} h_N(w)$ converges to zero for any w.

Consider

$$\sqrt{(N)} g_N(w) = \frac{N-1}{N} \int_{-\infty}^{+\infty} (w+x) \sqrt{(N)} f_{N-2}(x)\,\mathrm{d}F(-w-x) \tag{9.40}$$

Let us make the change of variable, $\tilde{x} = w + x$. Equation (9.40) becomes

$$\sqrt{(N)} g_N(w) = \frac{N-1}{N} \int_{-\infty}^{+\infty} \tilde{x} f(\tilde{x}) \sqrt{(N)} f_{N-2}(\tilde{x} - w)\,\mathrm{d}\tilde{x} \tag{9.41}$$

We know that $\sqrt{(N)} f_{N-2}(\tilde{x} - w)$ is uniformly bounded and $|\tilde{x} f(\tilde{x})|$ is integrable because $f(\cdot)$ has a bounded variance; therefore, by the Lebesgue theorem

$$\lim_N \sqrt{(N)} g_N(w) = \int_{-\infty}^{+\infty} \tilde{x} f(\tilde{x}) \lim_N \sqrt{(N)} f_{N-2}(\tilde{x} - w)\,\mathrm{d}\tilde{x}$$
$$= c\int_{-\infty}^{+\infty} \tilde{x} f(\tilde{x})\,\mathrm{d}\tilde{x} = 0 \tag{9.42}$$

Finally, since $\sqrt{(N)} f_{N-1}(w)$ converges to a positive constant for any w, any solution of the first-order condition, and therefore the optimal answer converges to the truth v. Q.E.D.

Therefore, we have shown that, if the population is large enough, the optimal answer of an agent, even if he takes into account the effect of the rebate, in a strictly closed mechanism, is close to the truth. The mechanism yields asymptotically truthful answers.

This is promising but not enough for our purpose. It is possible that a small deviation by each agent does not remove the possibility of an error in the social decision in a large population. We take up this issue next.

9.8. Asymptotic successfulness

The primary goal of these mechanisms is to induce collective decisions that are Pareto optimal. Closing the system through a rebate of the pivotal payments collected induces a bias in individual responses that is small if the individual perceives himself to be only one among a large number of agents in the system. But, this does not guarantee that the social decision is "approximately optimal". Indeed, since we view individuals having stochastically generated tastes, the approximate optimality of such closed mechanisms, if it can be proven to hold at all, must be a statistical statement.

We will say that a mechanism is *asymptotically successful* if for large enough economies it makes the correct decision regarding acceptance or rejection of the project with arbitrarily high probability. If the rebate mechanism analyzed above did not have this property, the fact that individuals could be induced to give asymptotically correct answers would be of little relevance. A priori, it seems possible that if in order to induce answers of a given degree of accuracy we need a very large number of agents, then the errors made by these agents in responding to the mechanism might accumulate in such a way as to distort the aggregate.

In this section we assume that the v_i are independently and identically distributed with zero mean. Further, it is supposed that the distribution through which they are generated is symmetric about zero and has a density $f(\cdot)$.

In light of the preceding section, we know that the optimal response $w_N(v_i)$ of agent i believing that the population is of size N satisfies.

(W.1) $w_N(\cdot)$ is an integrable function of v_i.
(W.2) $w_N(v_i)$ approaches v_i as N goes to infinity, for every v_i.
(W.3) $|w_N(v_i) - v_i| < |v_i|$
(W.4) $w_N(v_i) = w_N(-v_i)$.

To prove that the mechanism is asymptotically successful, it suffices to show that

$$\Pr\left[\sum_i v_i \gtrless 0 \text{ and } \sum_i w_N(v_i) \lessgtr 0\right] \text{ goes to zero as } N \text{ approaches infinity.} \tag{9.43}$$

As a result, the probability of taking a non-optimal decision approaches zero with increasing N.

Theorem 9.4. Under the hypotheses above, the probability that the mechanism will take the correct decision approaches one as the number of agents goes to infinity.

Proof. Note first that (9.43) is equivalent to

$$\Pr\left[\frac{\sum_i v_i}{\sqrt{(N)}} \gtrless 0 \text{ and } \frac{\sum_i w_N(v_i)}{\sqrt{(N)}} \lessgtr 0\right] \to 0 \text{ when } N \to \infty. \tag{9.44}$$

Denote:

$$\hat{v}_{iN} = \frac{v_i}{\sqrt{(N)}} \qquad \hat{w}_N = \sum_i \hat{w}_{iN}$$

$$\hat{v}_N = \sum_i \hat{v}_{iN} \qquad y_N = [\hat{v}_N, \hat{w}_N] \tag{9.45}$$

$$\hat{w}_{iN} = \frac{w_N(v_i)}{\sqrt{(N)}}$$

Let $y = [y^1, y^2]$ be the vector of jointly normally distributed random variables with zero mean and variance-covariance matrix

$$\begin{bmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{bmatrix}$$

According to the theorem of Varadarajan,[9] it suffices to show that for all $(\lambda_1, \lambda_2) \in \mathbf{R} \times \mathbf{R}$,

$$\lambda_1 \hat{v}_N + \lambda_2 \hat{w}_N \to \lambda_1 y^1 + \lambda_2 y^2, \tag{9.46}$$

in order to conclude that $y_N \to y$. That is, $\lambda_1 \hat{v}_N + \lambda_2 \hat{w}_N$ approaches a univariate normal with zero mean and variance $(\lambda_1 + \lambda_2)^2 \sigma^2$.

Let

$$a_{iN} = \lambda_1 \hat{v}_{iN} + \lambda_2 \hat{w}_{iN} \tag{9.47}$$

and let F_{iN} be the distribution of a_{iN}. Finally, let

$$a_N = \lambda_1 \hat{v}_N + \lambda_2 \hat{w}_N \tag{9.48}$$

We will apply the Lindeberg-Feller Central Limit theorem to conclude that a_N converges to a normal distribution for every choice of λ_1 and λ_2, and that the variance of this distribution is the same as that of the random variable $\lambda_1 y^1 + \lambda_2 y^2$, namely, $(\lambda_1 + \lambda_2)^2 \sigma^2$.

Let,

$$\sigma_\infty^2 = \lim_N \operatorname{var} a_N$$

The Lindeberg-Feller theorem says that if

(i) $\sigma_\infty^2 > 0$

and

(ii) for all $\bar{\alpha} > 0$

$$\sum_{i=1}^N \int_{|\alpha| > \bar{\alpha}} \alpha^2 \, dF_{iN} \to 0$$

as $N \to \infty$

[9] See Rao [1973], p. 128.

then,

a_N approaches a normal random variable with mean zero and variance σ_∞^2.

First, we verify (i), proving in particular that $\sigma_\infty^2 = (\lambda_1 + \lambda_2)^2\sigma^2 > 0$. Then we verify (ii). By the remarks above, this will complete the proof of the theorem.

Let $w_N(v_i) - v_i = e_N(v_i)$. Note that, by (W.4) and symmetry of the distribution of v_i, $Ee_N(v_i) = 0$. Thus

$$a_{iN} = \frac{1}{\sqrt{(N)}}[(\lambda_1 + \lambda_2)v_i + \lambda_2 e_N(v_i)]$$

$$\sigma_{iN}^2 = \frac{1}{N}[(\lambda_1 + \lambda_2)\sigma^2 + 2\lambda_2(\lambda_1 + \lambda_2)Ev_i e_N(v_i) + \lambda_2^2 Ee_N(v_i)^2] \qquad (9.49)$$

Let

$$\sigma_N^2 = \sum_i \sigma_{iN}^2 = (\lambda_1 + \lambda_2)^2\sigma^2 + 2\lambda_2(\lambda_1 + \lambda_2)\frac{1}{N}\sum_i Ev_i e_N(v_i) + \lambda_2^2 \frac{1}{N}\sum_i Ee_N(v_i)^2$$

But

$$\frac{1}{N}\sum_i Ev_i e_N(v_i) = Ev_i e_N(v_i) = \int_{-\infty}^{+\infty} v_i e_N(v_i) f(v_i)\,\mathrm{d}v_i = \int_{-\infty}^{+\infty} \psi_N(v_i)\,\mathrm{d}v_i \qquad (9.50)$$

since the v_i are distributed identically, with $|\psi_N(v_i)| < v_i^2 f(v_i)$ by (W.3) and $\psi_N(v_i) \xrightarrow{N} 0$ for each v_i by (W.2).

Hence, by the Lebesgue Dominated Convergence theorem,

$$Ev_i e_N(v_i) \xrightarrow{N} 0$$

Similarly,

$$\frac{1}{N}\sum_i Ee_N(v_i)^2 = Ee_N(v_i)^2 \xrightarrow{N} 0$$

Hence,

$$\sigma_N^2 \to (\lambda_1 + \lambda_2)^2\sigma^2 = \sigma_\infty^2$$

To prove that (ii) holds, take $\bar{\alpha} > 0$ and consider

$$\int_{|\alpha|>\bar{\alpha}} \alpha^2\,\mathrm{d}F_{iN} = \int_{A_N} \frac{[\lambda_1 v_i + \lambda_2 w_N(v_i)]^2}{N}\,\mathrm{d}F(v_i)$$

where

$$A_N = \{v_i \,|\, |(\lambda_1 + \lambda_2)v_i + \lambda_2 e_N(v_i)| > \bar{\alpha}\sqrt{(N)}\}$$

Since the v_i are identically distributed, these integrals are independent of i. The following inequalities show when multiplied by N they are each still converging to zero. Hence the sum will go to zero as required by (ii).

$$|\lambda_1 v_i + \lambda_2 w_N(v_i)|^2 \leqslant \lambda_1^2 v_i^2 + \lambda_2^2 w_N(v_i)^2 + 2\lambda_1\lambda_2 v_i w_N(v_i) \leqslant (\lambda_1 + 2\lambda_2)^2 v_i^2 \quad \text{by (W.3).} \qquad (9.51)$$

Hence,

$$\int_{A_N} \alpha^2\,\mathrm{d}F_{iN} < \frac{(\lambda_1 + 2\lambda_2)^2}{N}\int_{A_N} v_i^2\,\mathrm{d}F(v_i). \qquad (9.52)$$

Since $\int v_i^2\,\mathrm{d}F(v_i)$ is finite, it suffices to show that $\bigcap_{N=1}^{\infty} A_N = \{0\}$ in order that $\int_{A_N} v_i^2\,\mathrm{d}F(v_i)$ will be made arbitrarily small with N.

But $|\lambda_1 v + \lambda_2 w_N(v)| \leqslant (\lambda_1 + 2\lambda_2)|v|$ and thus for every v, there exists N_o such that

$$|v| \leqslant \frac{\alpha\sqrt{(N)}}{\lambda_1 + 2\lambda_2} \qquad (9.53)$$

for any $N \geqslant N_o$.

Thus $\bigcap_{N=1}^{\infty} A_N = \{0\}$ and for any $\varepsilon > 0$, there exists N^* such that

$$\int_{A_N} v_i^2\,\mathrm{d}F(v_i) < \frac{\varepsilon}{(\lambda_1 + 2\lambda_2)^2}$$

or,

$$\left|\sum_i \int_{|\alpha|\geqslant\bar{\alpha}} \alpha^2\,\mathrm{d}F_{iN}\right| < \varepsilon \qquad (9.54)$$

for $N > N^*$. This verifies (ii).

Hence y_N converges in distribution to y, and for N sufficiently large,

$$|\Pr[\hat{v}_N \gtrless 0 \text{ and } \hat{w}_N \lessgtr 0] - \Pr[y^1 \gtrless 0 \text{ and } y^2 \lessgtr 0]| < \varepsilon \qquad (9.55)$$

Since

$$\Pr[y^1 \gtrless 0 \text{ and } y^2 \lessgtr 0] = 0$$

we have the desired result of asymptotic successfulness of the mechanism.

Q.E.D.

For simplicity, we have concentrated here on the case of symmetric distributions; however, it is easy to see, at least for distributions with compact supports, that the theorem generalizes to nonsymmetric distributions, using the results of section 7.

9.9. Conclusion

We hope to have convinced the reader that the problem of balancing the budget is not a very serious problem for the mechanisms we are studying. When the mean of the population is expected to be zero, i.e., the most interesting case for the decision maker (since he is really uncertain about what is the optimal decision), it is possible to close the system by redistributing the surplus without damaging the mechanism.

When expected valuations are zero, dominant strategies can be maintained by disposing of a surplus which is small relative to the population size. If the mean is other than zero, the expected surplus is even small in absolute terms. Therefore, for large populations at least, an approximately optimal system can be achieved either with or without rebating these pivotal payments.

Chapter 10

COALITION INCENTIVE COMPATIBILITY

10.1. Introduction

Although the main focus of this book is the elicitation of privately held information on the individual level, there is the danger that cooperative distortions by coalitions would upset the efficiency of the outcomes. In light of theorems 5.10 and 5.11 this possibility is clearly a relevant one. There is no way to design individually incentive compatible mechanisms so as to prevent cheating by any coalition of two or more agents, even if the potentially deviant coalition could be identified in advance. Some results for limited classes of coalitions can be obtained only for highly restricted domains of valuation functions (example 5.7).

The purpose of this chapter is to show that in large economies, coalitions can gain so little by cheating that they are unlikely to attempt to distort their preferences in any realistic situation in which there is a positive cost to collusion. Therefore, individual incentives, which favor truth-telling, will dominate.

In section 10.2 we study optimal cheating behavior for a coalition when the allocation is made using a Clarke mechanism on the individual level. The main result discussed above is proven in section 10.3, for this mechanism. Generalizations to a subclass of all Groves mechanisms are treated in section 10.4. A final section gathers some interpretations and more general remarks on the coalition problem.

10.2. Optimal cheating by a coalition

Let us recall the simple example of a three person economy, discussed at the beginning of section 5.5. Preferences are $v_1 = -6$, $v_2 + 2$, $v_3 = +2$.

Using any Groves mechanism each agent has an incentive to respond with $w_i = v_i$. The project would be rejected but if individuals 2 and 3 could agree to each say $+7$, then the pivotal mechanism would lead to acceptance with no transfers, and this is better than they would achieve by telling the truth. However, they cannot be assumed to know the statement that will be made by other individuals; or, more generally, they will even be ignorant of how many other individuals there are. Distortions will involve some risks. For example, if $w_1 = -13$ were the true preferences revealed by agent 1, then the distortion mentioned above would lead to both individuals in the coalition being taxed 6 units by the pivotal mechanism, which is more than their true willingness to pay for the project they have caused to be accepted. The larger their distortion, the surer they are of forcing their preferred social decision, but the more risk they must accept if their guess about others' statements proves erroneous.

To formalize this trade-off, let us consider a coalition C of size n. There are N other individuals. We assume that everybody in the coalition has the same expectations on the sum of the answers by the N others; these expectations are represented by a probability distribution function $F_N(x)$, which is assumed to be strictly monotone in x.

Let w_i be the answer given by the agent i for i in C. Let $V = \sum_{i\in C} v_i$ and $W = \sum_{i\in C} w_i$. We will call these the *global true value* and the *global answer* for coalition C.

For the coalition C, the expected utility of saying (w_i), $i \in C$, is:

$$U(N, n, (w_i)) = V \int_{-W}^{\infty} \mathrm{d}F_N(x) + \sum_{j\in C} \int_{-\sum_i w_i}^{-\sum_{j\neq i} w_j} [x + \sum_{j\neq i} w_j]\,\mathrm{d}F_N(x) \tag{10.1}$$

which is decomposed into its expected utility of getting the project minus its expected tax.

The answer of a coalition, C, is *symmetric* if everybody in the coalition gives the same personal evaluation, that is $w_i = w_j$ for i and j in C.

Theorem 10.1. For any global answer W by the coalition, the optimal answer of the coalition is symmetric.

Proof. Let us consider the program

$$\max_{(w_i)} U(N, n, (w_i)) \qquad \sum_{i\in C} w_i = W \tag{10.2}$$

(10.2) can be rewritten as

$$\max V \int_{-W}^{\infty} \mathrm{d}F_N(x) + \sum_{i\in C} \int_{-W}^{-W+w_i} [x + W - w_i]\,\mathrm{d}F_N(x) \tag{10.3}$$

$$\text{s.t. } \sum_{i\in C} w_i = W$$

The first order conditions are (λ being the Lagrange multiplier)

$$-\int_{-W}^{-W+w_i} \mathrm{d}F_N(x) = \lambda, \quad i \in C \tag{10.4}$$

For any $(i, i') \in C$, (10.4) gives

$$F_N(W + w_i) = F_N(W + w_{i'}) \quad \text{or } w_i = w_{i'} \tag{10.5}$$

It is easy to see that the second order conditions are fulfilled. Q.E.D.

Let w be the common individual answer of the members of the coalition C.

$$W = nw \tag{10.6}$$

The expected utility gain of saying W instead of the truth V is:

$$G(N, n) = V \int_{-W}^{-V} \mathrm{d}F_N(x) + n \int_{-W}^{-\frac{W(n-1)}{n}} \left[x + \left(\frac{n-1}{n}\right)W\right]\mathrm{d}F_N(x) - \sum_{i\in C} \int_{-V}^{-V+v_i} [x + \sum_{j\neq i} v_j]\,\mathrm{d}F_N(x) \tag{10.7}$$

The first-term represents the utility gained from the fact that the project will be accepted in some circumstances in which it would not otherwise be, or vice versa, because of this coalition's exaggeration of its beliefs. The second term is the expected pivotal payment that the coalition will make. Since they answer symmetrically they will all be taxed in a like manner, when $x \in (-nw, -(n-1)w)$. Finally, the last term is the pivotal payments they would have had to pay had they answered truthfully; these pivotal payments are now avoided.

Theorem 10.2. If $\sqrt{(N)}f_N(-nw)$ converges to a constant as N goes to infinity, and $\sqrt{(N)}f_N'(y)$ is continuous in y and for each y converges to zero, then w converges to V (and therefore W converges to nV) as N goes to infinity.

Proof. The optimal answer for each N is obtained by maximizing $G(N, n)$, i.e. from

$$Vf_N(-w) + (n-1)\int_{-W}^{-W+\frac{W}{n}} \mathrm{d}F_N(x) - Wf_N(-W) = 0 \tag{10.8}$$

which can be rewritten

$$(V - nw)f_N(-nw) + (n-1)wf_N(-nw + \delta w) = 0 \quad \text{for some } \delta \in [0,1] \tag{10.9}$$

or

$$(V - w)f_N(-nw) + (n-1)\delta w^2 f_N'(-nw + \tilde{\delta} w) = 0 \quad \text{for some } \tilde{\delta} \in [0,\delta] \tag{10.10}$$

Multiplying (10.10) by $\sqrt{(N)}$ we obtain the result directly from the assumptions. Q.E.D.

The somewhat technical assumptions of theorem 10.2 are indeed quite general and in particular are satisfied by the normal distribution with mean zero (used below). In the next section, we will assume that N is large enough to validate the approximation of W by nV, i.e. that the optimal way of cheating is for each member of a coalition of size n to say n times the average true willingness to pay in the coalition.

It is interesting to observe that the optimal cheating answers depend on the size of the coalition and that the answer should be the same for every member of the coalition.

Because of the second point they will either all be pivotal together, or no one in the coalition will be pivotal. This makes side-payments unnecessary within the group. The first point makes implicit coalition formation without communication much more difficult, because they have to guess the size of the group with which they are tacitly colluding, if they are to cheat optimally.

10.3. Value of optimal cheating

In this section we restrict ourselves to the case of a distribution function $F_N(\cdot)$, which is normal with mean zero and variance N. A more general

theory would be much more technical without throwing additional light on the basic result.

The expected gain (10.7) of cheating for a coalition of size n can then be bounded as follows.

The expected gain of cheating optimally is:

$$V\int_{-nV}^{-V} \mathrm{d}F_N(x) + n\int_{-nV}^{-(n-1)V} [x + (n-1)V]\,\mathrm{d}F_N(x) - n\int_{-V}^{-\frac{(n-1)}{n}V} \left[x + \left(\frac{n-1}{n}\right)V\right]\mathrm{d}F_N(x) \tag{10.11}$$

The integral (10.11) can be broken up in three parts, A, B, C:

$$A = V\int_{-nV}^{-V} \frac{1}{\sqrt{(2\pi N)}}\cdot \mathrm{e}^{-\frac{1}{2}\frac{x^2}{N}}\,\mathrm{d}x \leqslant \frac{(n-1)V^2}{\sqrt{(2\pi N)}}\cdot \mathrm{e}^{-\frac{1}{2}\frac{V^2}{N}} \tag{10.12}$$

$$B = n\int_{-nV}^{-(n-1)V} [x + (n-1)V]\frac{1}{\sqrt{(2\pi N)}}\mathrm{e}^{-\frac{1}{2}\frac{x^2}{N}}\,\mathrm{d}x$$

$$\leqslant n(n-1)V^2\frac{1}{\sqrt{(2\pi N)}}\mathrm{e}^{-\frac{1}{2}\frac{(n-1)^2V^2}{N}} + \frac{n\sqrt{(N)}}{\sqrt{(2\pi)}}[\mathrm{e}^{-\frac{n^2V^2}{2N}} - \mathrm{e}^{-\frac{(n-1)^2V^2}{2N}}] \tag{10.13}$$

$$C = -n\int_{-V}^{-\frac{(n-1)V}{n}} \left[x + \frac{(n-1)}{n}V\right]\mathrm{d}F_N(x)$$

$$\leqslant -\frac{(n-1)}{n}V^2\frac{\mathrm{e}^{-\frac{V^2}{2N}}}{\sqrt{(2\pi N)}} - \frac{n\sqrt{(N)}}{\sqrt{(2\pi)}}\left[\mathrm{e}^{-\frac{V^2}{2N}} - \mathrm{e}^{-\frac{(n-1)^2}{n^2}\frac{V^2}{2N}}\right] \tag{10.14}$$

Thus, (10.11) is bounded by:

$$\frac{(n-1)V^2}{\sqrt{(2\pi N)}}\cdot \mathrm{e}^{-\frac{1}{2}\frac{V^2}{N}} + \frac{n(n-1)}{\sqrt{(2\pi N)}}V^2\mathrm{e}^{-\frac{1}{2}\frac{(n-1)^2V^2}{N}}$$

$$-\frac{(n-1)V^2}{n\sqrt{(2\pi N)}}\cdot \mathrm{e}^{-\frac{V^2}{2N}} - \frac{n\sqrt{(N)}}{\sqrt{(2\pi)}}\left[\mathrm{e}^{-\frac{V^2}{2N}} - \mathrm{e}^{-\left(\frac{n-1}{n}\right)^2\frac{V^2}{2N}} - \mathrm{e}^{-\frac{n^2V^2}{2N}} + \mathrm{e}^{-\frac{(n-1)^2V^2}{2N}}\right]$$

Expanding the middle term one can see that for a fixed coalition (that is to say a fixed n and a fixed V), the expected gain of cheating can be made as small as we wish, for N large enough.

However, when we increase the size of the population, the probability that there exists a coalition of size n, with a very large V, will increase if we view the tastes in the population as being drawn independently from some distribution. Therefore, we will need to show a stronger result before the likelihood of cheating can be proven small. Namely, for any n, we must have the probability that there will exist a coalition of size n among the members of the population such that its gain from cheating is larger than any given number, can be made as small as we wish by increasing N.

Indeed we might be interested in an even further strengthening of this result to allow the size of the cheating coalition whose existence we want to preclude to grow with the population size, instead of being fixed at n. In the next theorem we allow coalitions to grow at the rate of $N^{1/6}$, where N is the population size. In particular this means that for any fixed n, a sufficiently large population will make it extremely unlikely that they can beneficially distort their preferences.[1]

Theorem 10.3. For any positive numbers ε and δ, the probability that there exists a coalition of any size n such that $n \leqq N^{1/6-\delta}$, capable of sustaining a per capita gain from cheating greater than ε, goes to zero as N goes to infinity.

Proof. Note first that the global willingness to pay, V, of a potential cheating coalition of size n is such that

$$V \leqq n\, v_{\max} \tag{10.16}$$

where $v_{\max}$ is the maximum willingness to pay in the entire population. We will now argue that for any fixed probability $\eta > 0$,

$$\Pr(v_{\max} > N^{\delta}) < \eta. \tag{10.17}$$

for N sufficiently large. We will show that whenever $v_{\max} < N^{\delta}$, then no coalition of any size n such that $n \leqq N^{1/6-\delta}$ can achieve a per capita gain of ε. We consider the population of size N to be a random sample in a normal population with zero mean and variance one in computing the probability in (10.17).

From David [1970] we know that the asymptotic distribution of the random variable

$$M_N = (2\log N)^{\frac{1}{2}}\,[v_{\max} - (2\log N)^{\frac{1}{2}}] \tag{10.18}$$

is such that

$$\Pr(M_N \leqq v) \cong e^{-e^{-v}} \tag{10.19}$$

Therefore, for N large we have that

$$\Pr\left(v_{\max} < \frac{v}{(2\log N)^{\frac{1}{2}}} + (2\log N)^{\frac{1}{2}}\right) \approx e^{-e^{-v}} \tag{10.20}$$

Letting $v = (2\log N)^{\frac{1}{2}}\,(N^{\delta} - (2\log N)^{\frac{1}{2}})$, (10.20) becomes

$$\Pr(v_{\max} < N^{\delta}) \approx \exp[-\exp[-(2\log N)^{\frac{1}{2}}N^{\delta} - (2\log N)^{\frac{1}{2}}]] \tag{10.21}$$

which clearly approaches 1. Thus (10.17) is valid, as desired, for large N.

Using $n \leqq N^{1/6-\delta}$ we have that

$$V < N^{1/6-\delta} \cdot v_{\max} \tag{10.22}$$

which implies from (10.21) that, with probability $1 - \eta$,

$$V < N^{1/6-\delta} \cdot N^{\delta} = N^{1/6} \tag{10.23}$$

Let us use (10.23) to evaluate the per capita gain given in (10.15). Using the Taylor expansion for the bracketed expression in (10.15), we find that it is of the order of $n^2V^2/\sqrt{(N)}$. One can see that this is the dominant order of all the terms on the right-hand side of (10.15), under the conditions $n \leqq N^{1/6-\delta}$, $V \leqq N^{1/6}$.

On a per capita basis, therefore, the gain is bounded by

$$\frac{nV^2}{\sqrt{(N)}} < \frac{N^{1/6-\delta}N^{1/3}}{\sqrt{(N)}} = N^{-\delta}$$

which clearly converges to zero as N goes to infinity. Thus, as long as $V < N^{1/6}$, which we know will happen with at least probability $1 - \eta$, the conclusion of the theorem[2] holds. Q.E.D.

[1] Since our result on the rate of growth of the minimal coalition size necessary for cheating advantageously is an asymptotic one, we do not know the *actual* size of coalitions that are allowed or precluded. It is therefore incorrect to say that $n < N^{1/6}$ implies that a coalition of size n cannot cheat. We only know that if α is fixed and if $n < \alpha N^{1/6}$ and N is large, then a coalition of size n will not be able to gain by cheating. The size required for N depends on α.

[2] In this proof we used the quite pessimistic upper bound

$V \leqq n\, v_{\max}$.

However, one can see from the argument that the result bounding the rate of growth of the coalition size cannot be improved upon by a more precise approximation of the asymptotic

10.4. Generalizations

In this section we extend the results to a subclass of Groves mechanisms.

We can express the transfer function of any Groves mechanism as that of the Clarke mechanism plus an arbitrary function, $\delta_i(w_{-i})$, depending only on the statements of other agents.

$$\begin{aligned} t_i(w) &= \sum_{j\neq i} w_j + \min(-\sum_{j\neq i} w_j, 0) + \delta_i(w_{-i}) \qquad \text{if } \sum_i w_i \geqq 0 \\ &= \min(-\sum_{j\neq i} w_j, 0) + \delta_i(w_{-i}) \qquad \text{if } \sum_i w_i < 0 \end{aligned} \tag{10.24}$$

The $\delta_i(\cdot)$ functions clearly form an equivalent characterization of the class of Groves mechanisms.

Definition 10.1. A Groves mechanism is said to be *impersonal*, if for every i, $\delta_i(w_{-i})$ is the same symmetric function of its argument.

Definition 10.2. A Groves mechanism is said to be *concave* (*differentiable*) if for every i, $\delta_i(w_{-i})$ is a concave (differentiable) function of w_{-i}.

Theorem 10.4. For any concave differentiable impersonal Groves mechanism and for any desired aggregate answer W by a coalition, the optimal answer of a coalition is symmetric.

Proof. The objective of the coalition is to choose $(w_i)\ i \in C$ to maximize

$$V\int_{-W}^{\infty} \mathrm{d}F_N(x) + \sum_{i\in C} \int_{-W}^{-W+w_i} (x + W - w_i)\mathrm{d}F_N(x) + \sum_{i\in C} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \delta(w_{-i})\mathrm{d}\mu((w_j)_{j\notin C}) \tag{10.25}$$

properties of V. Even if $v_{\max}$ were *bounded*, say by B, the dominant term in (10.15) would be

$$\frac{nV^2}{\sqrt{(N)}} \leqq \frac{n^2B^2}{\sqrt{(N)}}$$

and thus

$$n \leqq N^{1/6-\delta}$$

is still required to insure its convergence towards zero.

where μ is coalition C's common beliefs about the joint distribution of $(w_j)_{j\notin C}$.

The first order condition for maximization with respect to w_i for $i \in C$ is

$$VF_N(-W) + F_N(W) - F_N(-W + w_i) + \sum_{\substack{l\in C\\ l\neq i}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\partial}{\partial w_i}(\delta(w_{-l}))\mathrm{d}\mu((w_j)_{j\notin C}) = 0\,. \tag{10.26}$$

Equating (10.26) for two different agents i and i' we obtain

$$\begin{aligned} &F_N(-W + w_i) - \sum_{\substack{l\in C\\ l\neq i}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\partial}{\partial w_i}\delta(w_{-l})\,\mathrm{d}\mu((w_j)_{j\notin C}) \\ &= F_N(-W + w_{i'}) - \sum_{\substack{l\in C\\ l\neq i'}} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \frac{\partial}{\partial w_{i'}}\delta(w_{-l})\mathrm{d}\mu((w_j)_{j\notin C}) \end{aligned} \tag{10.27}$$

Because the mechanism is impersonal, it is possible to subtract from both sides of (10.27) all the terms in the summations where $l \neq i$ and $l \neq i'$. After these subtractions the left-hand side of (10.27) becomes a function of w_i and not $w_{i'}$, because the only term left in the summation is that associated with $l = i'$, and $w_{i'}$ is not an argument of that function.

Similarly, the right-hand side of (10.27) would now be a function of $w_{i'}$ and not w_i.

Since the mechanism is impersonal, the two sides are the same functions of their respective arguments. Moreover by the concavity of $\delta(\cdot)$ they are both decreasing and can be equal only if the two arguments are equal. It is easy to check that the second order conditions are fulfilled. Q.E.D.

For simplicity, we impose the further restriction that δ is a function of the sum of everybody else's answer, i.e.

$$\delta(x + \sum_{\substack{j\in C\\ j\neq i}} w_j)$$

The following proof can be extended to other cases but requires conditions on $\delta(\cdot)$ which are not easily interpreted.

Definition 10.3. A differentiable Groves mechanism is said to be *asymptotically pivotal* if $\delta_i(y)$ approaches a constant B as $|y|$ tends toward infinity and $|\delta_i'(y)|$ is integrable.

Theorem 10.5. If $\sqrt{(N)}f_N(x)$ is uniformly bounded and converges at each point to a constant B, and $\sqrt{(N)}f_N'(y)$ is continuous in y and converges to zero, then for an impersonal, concave, differentiable, asymptotically pivotal Groves mechanism, the optimal global answer of a coalition of size n approaches nV (where $V = \sum_{i\in C} v_i$) as N becomes large.

Proof. The additional term in the analog of (10.7) is

$$(n-1)\int_{-\infty}^{+\infty} \delta'\left(x + \frac{W(n-1)}{n}\right)\sqrt{(N)}f_N(x)\,\mathrm{d}x \tag{10.28}$$

since we can drop now the indices of the δ which can be rewritten as:

$$(n-1)\int_{-\infty}^{+\infty} \delta'(\tilde{x})\sqrt{(N)}f_N\left(\tilde{x} - \frac{W(n-1)}{n}\right)\mathrm{d}\tilde{x} \tag{10.29}$$

We know that $\sqrt{(N)}f_N(\tilde{x} - (W(n-1)/n))$ is uniformly bounded, hence $|\delta'(\tilde{x})\sqrt{(N)}f_N(\tilde{x} - W(n-1)/n)|$ is bounded by an integrable function since $|\delta'(\tilde{x})|$ is integrable.

By Lebesgue's theorem:

$$\lim_{N\to\infty}(n-1)\int_{-\infty}^{+\infty} \delta'(\tilde{x})\sqrt{(N)}f_N\left(\tilde{x} - \frac{W(n-1)}{n}\right)\mathrm{d}\tilde{x}$$

$$= (n-1)B\int_{-\infty}^{+\infty}\delta'(\tilde{x})\,\mathrm{d}\tilde{x} = 0 \tag{10.30}$$

Consequently (10.28) can be made as small as wished by increasing N. Therefore, the result of theorem 10.2 still holds. Q.E.D.

The only mechanisms $\delta(\cdot)$ which are simultaneously concave and satisfy the assumption of being asymptotically pivotal are the pivotal mechanisms plus a constant. However, the requirement of concavity is obviously a sufficient but not a necessary one for obtaining the result of theorem 10.4. Finally, we can state

Theorem 10.6. For any concave, impersonal, differentiable, asymptotically pivotal Groves mechanism, theorem 10.3 holds.

Proof. The proof is similar to the proof of section 2 but we have here an additional term in the gain attained by optimal cheating, namely:

$$S(n,N) = n\int_{-\infty}^{+\infty}\left[\delta(x + V(n-1)) - \delta\left(x + \frac{V(n-1)}{n}\right)\right]f_N(x)\,\mathrm{d}x \tag{10.31}$$

But

$$\left|\delta(x + V(n-1)) - \delta\left(x + \frac{V(n-1)}{n}\right)\right| < \frac{(n-1)^2}{n}V\delta'(x + \lambda V)$$

$$\text{for } \lambda \in \left[\frac{n-1}{n}, n\right] \tag{10.32}$$

In per capita terms

$$\frac{|S(n,N)|}{n} < \int_{-\infty}^{\infty}\frac{(n-1)^2}{n}V\delta'(\tilde{x})f_N(\tilde{x} - \lambda V)\,\mathrm{d}\tilde{x}$$

$$< \int_{-\infty}^{\infty} v_{\max}(n-1)^2\delta'(\tilde{x})f_N(\tilde{x} - \lambda V)\,\mathrm{d}\tilde{x} \tag{10.33}$$

As seen in the proof of theorem 10.4, we have, with arbitrarily high probability,

$$v_{\max} \leqq N^{\delta}$$

for any $\delta > 0$.

Taking $n \geqq N^{1/4-\delta}$ for $\delta > 0$, we find

$$\frac{|S(n,N)|}{n} < \int_{-\infty}^{\infty} N^{-\delta}\sqrt{(N)}\delta'(\tilde{x})f_N(\tilde{x} - \lambda V)\,\mathrm{d}\tilde{x} \tag{10.34}$$

$$< \alpha N^{-\delta}, \text{ for some constant } \alpha.$$

since $\delta'(\cdot)$ is integrable and $\sqrt{(N)}f_N(\cdot)$ converges uniformly to a constant.

Therefore the result of theorem 10.4 implies that the extension to more general mechanisms as in the hypothesis does not upset the result as long as $n \leqq N^{1/6-\delta}$ is required. Q.E.D.

10.5. Conclusion

The purpose of this section is to give an interpretation of the preceding results. We have shown that the probability that there exists a coalition of size n less than the sixth root of the population size with a per capita gain of more than ε, can be made as small as we wish with large N.

The requirement of a per capita strictly positive gain can be justified by the need for communication to form coalitions. Communication costs are often thought to be of the order of the square of the relevant *population* size. Here we have taken the point of view that they are only proportional to the size of the coalition. This neglects the costs of trial and error in the process of finding a viable coalition, and hence our results overstate the likelihood of deviant coalitions.

What about large coalitions? First, these coalitions will clearly have free rider problems of their own. Moreover, they will have very high enforcement costs. We could have interpreted the required per capita gain as a necessary enforcement cost for members of the coalition, to be sure that the binding agreement they made is respected. For large populations, this type of cost becomes unbearable especially in a society which punishes recognized deviant coalitions. Also, the large coalitions, because the optimal method of cheating is to grossly exaggerate, will be easily recognized. Of course, they could try to take this risk of discovery into account, but then the attainable gain from cheating would be even smaller. Finally, it would be very easy for a society to catch such coalitions by infiltration.

Consequently, for all these reasons, very large coalitions should not be a problem. For small coalitions, our result indicates that manipulation by coalition is not *likely* to create difficulties for social decision making based on an appropriate Groves-type mechanism.

Chapter 11

LARGE NUMBERS AND SEPARABILITY

11.1. Biased mechanisms with non-separable preferences

One of the most severe restrictions on our analysis has been the assumption of additively separable preferences. This was investigated in chapter 5 where it was shown that for certain sufficiently rich domains, containing the separable preferences as a subclass, it was impossible to assure the existence of a satisfactory mechanism. This result allows for mechanisms whose strategy space is the set of all possible valuation functions of the transfer, rather than the real numbers which would be only large enough to describe a separable relation. Because of the failure to attain successful results in this instance, we investigated the properties of a simpler mechanism using the real numbers as a strategy space. Particular attention was paid to the limiting behavior of the process as the extent of the non-separability becomes small. In this way it was shown that a nearly successful result could be attained if income effects could be almost neglected.

This analysis leaves us in a rather uncomfortable state for many important public decisions. When the public project under consideration has a positive income effect for all agents, as will typically be the case, a significant downward bias will be introduced in the stated net willingness to pay through understatement by agents for whom it is desirable. Many such projects would therefore fail to be adopted when the optimal action would be just the reverse.

In the spirit of the other chapters of this part, we inquire as to whether superior results could be attained in large economies. Our results are somewhat mixed. It will be shown that the bias in individuals' statements is actually independent of the number of other agents. This is in contrast to the results in chapters 9, 10 and 12, in which the extent to which individual agents distort their true preferences due to the various imperfections considered becomes negligible in large economies.

The bias can be overcome to some extent if the government or planner has exogenous information regarding the extent of the income effect as a function of the agent's net willingness to pay in the absence of transfers, but of course has only imperfect information regarding the empirical distribution of these willingnesses-to-pay in the population.

We will show that the probability of taking an incorrect decision can be reduced by using a mechanism which is not a direct revelation mechanism. Its strategy spaces are the real numbers, whereas a direct revelation mechanism would require the space of real-valued functions (of the transfer) to be fully described. More importantly, the government counteracts the biases in individuals' answers by accepting the project even if the sum of the responses is somewhat negative, whereas a direct revelation mechanism would have to respect announced tastes. However, in contrast to the other asymptotic results of this part, it will be shown that the persistence of a non-negligible bias on the individual level as the economy grows will preclude an asymptotic successfulness result of the form of section 9.8. Some positive risk of taking the wrong decision cannot be eliminated.

Let us begin by recalling the results of section 5.1. We consider an individual described by his willingness to pay as a function of the transfer received, $v_i(t)$. The mechanism under consideration is given by:

$$\begin{aligned}
&S_i = \mathbf{R} \qquad i = 1, \ldots, N\\
&d(w) = 1 \qquad \text{iff } \textstyle\sum_i w_i \geqslant 0\\
&t_i(w) = \sum_{j \neq i} w_j + \min\left(-\sum_{j \neq i} w_j, 0\right) \qquad \text{if } \sum_i w_i \geqslant 0\\
&\qquad\; = \min\left(-\sum_{j \neq i} w_j, 0\right) \qquad \text{if } \sum_i w_i < 0.
\end{aligned} \tag{11.1}$$

The idea of this mechanism, intuitively, is that the agent is asked to reveal what his willingness to pay would be in the absence of a transfer, $v_i(0)$. Of course the mechanism is just a function and the individual is not compelled to think of the "no transfer" situation, or any other particular aspect of his preferences. He chooses his response, w_i, in order to maximize his expected utility, the expectation being taken with respect to his subjective beliefs concerning the sum, x, of strategies played by the other agents. By virtue of theorems 5.1 and 5.2, no *SIIC* successful mechanisms exist for this environment, so that optimal strategies cannot be determined independently of expectations.

We denote the deviation from the response $v_i(0)$ by δ_i:

$$w_i = v_i(0) + \delta_i \tag{11.2}$$

and the choice of δ_i is the actual decision variable of the agent. It will be determined through the true valuation function $v_i(\cdot)$ and the subjective distribution on x.

The subjective distribution on x will in fact depend on the number of other agents in the system. There is, therefore, some hope that by increasing this number, δ_i can be made to tend to $v_i(0)$. Our first result, and a severe negative one, is that the expected utility maximizing choice of δ_i is in fact independent of the subjective distribution held, and therefore that no asymptotic dominance or asymptotic truthfulness results can be obtained via this method. We have, actually, already observed this in chapter 5 for the case in which the willingness-to-pay functions were approximated linearly. Now we proceed in more generality.

Paralleling chapter 5 we consider the case of an agent for whom $v_i(0) > 0$. His choice of δ can be broken down into the following 3 regions.

case I: $\delta > 0$
case II: $-v_i(0) < \delta < 0$
case III: $\delta < -v_i(0) < 0$.

Let us write $v = v_i(0)$ and $\delta = \delta_i$ to simplify notation. We showed in chapter 5 that the expected gain from saying $v + \delta$ instead of v can be written as

$$\begin{aligned}
&\text{case I:} \quad \int_{-v-\delta}^{-v} (x + v_i(x))\,\mathrm{d}P(x)\\
&\text{case II:} \quad \int_{-v}^{-v-\delta} (-x - v_i(x))\,\mathrm{d}P(x)\\
&\text{case III:} \quad \int_{-v}^{0} (-x - v_i(x))\,\mathrm{d}P(x) + \int_{0}^{-v-\delta} (-x - v_i(0))\,\mathrm{d}P(x)
\end{aligned} \tag{11.3}$$

where $P(\cdot)$ is the subjective distribution function. The important feature of (11.3) that we shall exploit is that the value of δ enters into the optimization only through the limits of integration in each case. Case III can be eliminated because the value of such a policy is negative compared with $w_i = v_i(0)$. In the other two cases, first-order conditions for optimization with respect to δ are

$$\begin{aligned}
&\text{case I:} \quad [-v - \delta + v_i(-v - \delta)]p(-v - \delta) = 0\\
&\text{case II:} \quad [-v - \delta + v_i(-v - \delta)]p(-v - \delta) = 0
\end{aligned} \tag{11.4}$$

where $p(\cdot)$ is the density function of the distribution P. (The second order conditions are fulfilled when $v_i'(\cdot) > -1$.) We assume that $p(\cdot)$ is positive, and hence that these first order conditions are actually in both cases I and II:

$$-v - \delta + v_i(-v - \delta) = 0 \tag{11.5}$$

The optimal δ in each case will depend on the non-separability of $v_i(\cdot)$ because each condition must be solved for δ. If $v > 0$ as we have assumed, and $v_i'(\cdot) > 0$ so that the public project is a normal good at all levels of income, then

$$v_i(-v - \delta) < v \tag{11.6}$$

and thus $\delta < 0$ would be required to satisfy the first order condition (11.5). By virtue of the analysis above, the value of δ solving (11.5) would satisfy

$$0 > \delta > -v. \tag{11.7}$$

This result had been derived in chapter 5 for the case in which $v_i(t)$ was approximated linearly around $v_i(0)$. For the case $v_i(0) < 0$, similar arguments imply that

$$\delta = 0. \tag{11.8}$$

Thus we have established:

Theorem 11.1. Under the mechanism (11.1), an individual with non-separable preferences $v_i(\cdot)$ satisfying $v_i'(t) > 0$ for all t will distort his response downward, and, moreover, his response is independent of his expectations concerning the other agents.

The problem is clearly asymmetric as may be seen intuitively as follows. When an agent has a negative evaluation $v_i(0)$ (with $v_i'(\cdot) > -1$), he can gain by becoming pivotal and receiving a payment only if he kills the project, but then there is no income effect. It is as if we were in the separable case ($v_i'(\cdot) > -1$ excludes the anomalous case where he would become pivotal by making the project done because at a lower level of income he would enjoy the public good enough to compensate for this lost income: people would see their welfare improve by throwing away money.)

11.2. Symmetry and the elimination of bias

The distinctly downward bias of the individual's statements is due to the absence of complete symmetry in this problem, as we have generally assumed previously. Even if $v_i(0)$ is symmetrically distributed around zero, it would be necessary to have a symmetric distribution of $v_i'(t)$ conditional on the value of $v_i(0)$ in order to eliminate the expected bias in statements by individuals whose willingness to pay is fixed at any level. This would be a singularly bad assumption, and we shall not adopt it. Public goods need not be normal, but it makes little sense to suppose that the income effect is equally likely to have either sign. Similar symmetry conditions of a more general variety should be rejected on the same grounds.

Let us suppose, therefore, that the members of the economic system are characterized by a distribution, μ, over the space of real valued functions $v_i(\cdot)$, with the property that the induced marginal distribution of the parameter $v_i(0)$ is symmetrically distributed around zero with unique mode at that point. Moreover the variance of $v_i(0)$ is assumed to be finite. This corresponds to the assumptions used in section 9.9, and although they can be generalized somewhat in that context we shall not do so here. The analysis becomes much more complex without symmetry, and moreover, since in the next section we obtain negative conclusions concerning the potential for achieving the asymptotic successfulness of this process, it suffices to do so under the symmetry condition imposed.

For each function $v_i(\cdot)$ we can write the optimal deviation from the statement $v_i(0)$ as:

$$\delta_i(v_i(\cdot)) = w_i - v_i(0). \tag{11.9}$$

Let the expected deviation be denoted e:

$$e = \int \delta_i(v_i(\cdot))\, \mathrm{d}\mu(v_i(\cdot)). \tag{11.10}$$

As noted above, e will be negative. Therefore, the expected sum of the strategies played by members of an economy of size N is:

$$NEw_i = NE[v_i(0) + \delta_i(v_i(\cdot))] = Ne. \tag{11.11}$$

Thus, in a large system, the decision taken will almost surely be the result of this bias and will not reflect the true sign of $\sum_i v_i(0)$, since, by the central limit theorem this is increasingly unlikely to exceed Ne in absolute value.

This proves that the mechanism defined by (11.1) will virtually always reject the project.

11.3. Direct correction of the bias

The considerations of the previous section lead straightforwardly to a natural way of attempting to eliminate the bias. Assume that the planner knows the true distribution μ from which the functions $v_i(\cdot)$, $i = 1, \ldots, N$ are selected in a statistically independent fashion. This is the parallel of our usual assumption in which the distribution of tastes, parameterized by a single real number, is known to the planner. From μ he can calculate $\delta(v_i(\cdot))$ and hence e, which, it is important to note, will be independent of N.

Since the expected bias is known, it can be corrected by subtracting Ne from the aggregate statements of agents, and taking the decision according to the sign of the result. It is as if the government were to participate in the process by voting so as to offset the expected bias. In doing so, it might affect the statements made by the individuals, and therefore alter the bias it was designed to nullify. However, it can be seen that this modification of the mechanism does not change the optimal choice of $\delta(v_i(\cdot))$.

The mechanism is now defined by

$$S_i = \mathbf{R} \qquad i = 1, \ldots, N$$

$$d(w) = 1 \qquad \text{iff } \sum_i w_i \geqq Ne$$

$$\begin{aligned} t_i(w) &= \sum_{j \neq i} w_j - Ne + \min\Big(-\sum_{j \neq i} w_j + Ne, 0\Big) \quad \text{if } \sum_i w_i - Ne \geqslant 0 \\ &= \min\Big(-\sum_{j \neq i} w_j + Ne, 0\Big) \quad \text{if } \sum_i w_i - Ne < 0,\ i = 1, \ldots, N \end{aligned} \tag{11.12}$$

Writing now:

$$x = \sum_{j \neq i} w_j - Ne \tag{11.13}$$

we see clearly that the individual's problem is precisely the same as that described in section 1, above, except that the relevant subjective distribution $P(\cdot)$ has been shifted to the extent of Ne. By theorem 11.1, $\delta(v_i(\cdot))$ is unaffected by changes in $P(\cdot)$, and thus in particular by shifts of this type.

Hence we have proven:

Theorem 11.2. The mechanism (11.2) has the property that $\delta(v_i(\cdot))$ is the same function of $v_i(\cdot)$ for every N, and, for every N,

$$E\left[\sum_{i=1}^{N} w_i - Ne\right] = 0$$

whenever the underlying distribution of $v_i(0)$ has mean zero.

In this way, the bias discussed in section 11.2 has been offset, on average. It remains to be studied, however, whether or not the process results in a Pareto optimum with arbitrarily high probability as the economy grows. In the light of definition 5.1 and the discussion that follows it, we can concentrate on the possibility of choosing the correct decision to accept or reject the project *given* the resulting pattern of transfers. Showing that this mechanism cannot satisfy this criterion with high probability necessarily implies that it cannot attain Pareto optimality with high probability, as the latter requires simultaneously a correct decision and an optimal pattern of transfers.

There is a significant difference between the optimality concept of chapter 5 and the definition of asymptotic successfulness in chapter 9; relating to the fact that the mechanism of chapter 9 is closed whereas the present mechanism may collect transfers from some agents. If we follow the procedure of identifying the welfare of the planner with the level of transfers collected, then it makes sense to define asymptotic successfulness for processes in non-separable environments by

Definition 11.1. A mechanism is said to be *asymptotically successful over V* (a set of allowable valuation functions) if

$$\Pr\left(\sum_{i=1}^{N} w(v_i(\cdot)) \cdot \sum_{i=1}^{N} v_i(t_i(w(v_j(\cdot)), j = 1, \ldots, N)) < 0\right)$$

converges to zero with N.

If income effects for every individual can be bounded below by a strictly positive constant, then there are always asymptotically successful mechanisms for environments in which $Ev_i(0) = 0$. Namely, we can tax every individual to a very large extent, and this will virtually assure the negativity of $\sum_i v_i(t_i)$ so that, under our definition of successfulness, rejection of the project would always be optimal. This mechanism is a poor choice for several reasons. Arbitrary taxation at high levels, independent of announced preferences, would be an irrelevant option for society. It would be exactly in such cases that the mechanism would be successful over V, under the definition of chapter 5, but would be choosing Pareto optima only because the government's revenues would be large, rather than because of a well-taken decision which uses the non-closedness of the process purely to establish proper incentives. Moreover, the possibility of continually extracting large quantities of the private good is mitigated by the lower boundedness of consumption sets, neglected explicitly herein but not entirely forgotten, and by the adverse effects on individual willingnesses to participate in the process.

To some extent, the mechanism (11.12) is doing exactly that. It collects transfers and thereby reduces willingnesses to pay for the project, at a rate proportional to $\sqrt{(N)}$ (see section 9.3). Thus, as N grows, the project is almost surely undesirable. There is therefore good reason to consider a related mechanism in which government revenues will not increase so rapidly.

Suppose that for each N, the government computes the expected transfers

$$E_N = E\sum_i t_i(w_j(v_j(\cdot)), j = 1, \ldots, N) \tag{11.14}$$

and adds $-E_N/N$ to each of the individual transfer functions. For large N, E_N/N is very small and thus the effect of this addition on the answer chosen by each individual can be neglected. On average, his transfer will be zero. Whenever he is pivotal he will pay virtually his ordinary pivotal payment, and when he is not, he will receive a small subsidy – but the likelihood of being pivotal is so small that the latter is the typical case.

The sum $\sum_i v_i(0)$ is approximately $\sum_i v_i(t_i)$, as discrepancies can only be due to non-linearities in $v_i(\cdot)$ or to differences between $\sum_i t_i(w_j(v_j), j = 1, \ldots, N)$ and E_N. Thus we study the concept of approximate asymptotic successfulness defined formally by:

Definition 11.2. A mechanism is said to *approximately asymptotically successful* if

$$\Pr\left[\left(\sum_{i=1}^{N} w_i(v_i(\cdot)) - Ne\right)\cdot\left(\sum_{i=1}^{N} v_i(0)\right) < 0\right]$$

converges to zero with N.

In principle, the mechanism we are discussing may make errors for two types of reasons. Either $(\sum_i w_i(v_i(\cdot)) - Ne)$ and $(\sum_i v_i(0))$ have opposite signs, or, even if they have the same sign, it may be different from that of

$$\sum_i v_i\left(t_i(w(v_j(\cdot)), j = 1, \ldots, N) - \frac{E_N}{N}\right) \tag{11.15}$$

which gives the true correct decision for the transfers actually induced. Although it is possible to have cases in which both errors are simultaneously present and nullify each other, there is no reason to suppose that they will always happen together. Thus, by proving that this mechanism is not approximately asymptotically successful we will have shown that the problem of significant income effects cannot be overcome through large numbers arguments.

Theorem 11.3. Whenever μ assigns positive probability to $\{v_i(\cdot) \mid v_i(0) < 0\}$ and $\{v_i(\cdot) \mid v_i(0) > 0$ and $v_i'(t) > 0$ for all $t\}$, the mechanism (11.12, 11.14) is not approximately asymptotically successful.

The condition of this theorem rules out situations in which either the value of the project is surely determined in the absence of transfers, so that no mechanism is really necessary or there are no income effects; there the results of chapter 4 yield satisfactory outcomes. For simplicity we consider only the case $v_i'(0) > 0$, for all i, that is where the project is always a normal good. The proof of theorem 11.3 will be broken down into several lemmas.

Lemma 1. For each i, $\delta(v_i(\cdot)) = 0$ if $v_i(0) < 0$

$\delta(v_i(\cdot)) < 0$ and $|\delta(v_i(\cdot))| < |v_i(0)|$ if $v_i(0) > 0$.

Proof. From section 1, we know that when $v_i(0) < 0$, the agent reveals the true $v_i(0)$ independently of his expectations. If $v_i(0) > 0$, $\delta(\cdot)$ satisfies the first order condition given in section 1, namely:

$$\delta + v_i(0) = v_i(-v_i(0) - \delta) \tag{11.16}$$

If δ is positive, the left hand side of (11.16) is larger than $v_i(0)$ and the right hand side is smaller than $v_i(0)$, a contradiction. Hence $\delta < 0$.

If $\delta < -v_i(0)$, the left hand side is negative and the right hand side is positive, a contradiction. Q.E.D.

Lemma 2. For each i, $\delta(v_i(\cdot))$ is integrable.

Proof. First observe that

$$\left|\int_{-\infty}^{+\infty} \delta(v_i(\cdot))\,\mathrm{d}\mu(v_i(\cdot))\right| \leqq \int_0^{\infty} |v_i(0)|\,\mathrm{d}\mu(v_i(\cdot)) \tag{11.17}$$

by virtue of lemma 1. The integral, $\int_0^\infty |v_i(0)|\,\mathrm{d}\mu(v_i(\cdot))$ exists since we assumed in section 2 that the distribution of $v_i(0)$ had a finite variance. Q.E.D.

We are interested in the joint distribution of

$$V_N = \frac{\sum_{i=1}^{N} v_i(0)}{\sqrt{(N)}} \quad \text{and} \quad W_N = \frac{\sum_{i=1}^{N} [v_i(0) + \delta(v_i(\cdot)) - e]}{\sqrt{(N)}} \tag{11.18}$$

Let us compute the asymptotic matrix of variances and covariances of (V_N, W_N).

$$EV_N = 0 \quad \text{since} \quad \int v_i(0)\,d\mu(v_i(\cdot)) = 0 \text{ by assumption.} \tag{11.19}$$

$$EW_N = 0 \qquad \text{from above and the definition of } e. \tag{11.20}$$

Let us denote $\int v_i(0)^2\,d\mu(v_i(\cdot)) = \sigma^2$. From the independence of the sampled preferences we have:

$$\operatorname{var} V_N = \sigma^2 \tag{11.21}$$

$$\operatorname{var} W_N = \operatorname{var}[v_i(0) + \delta(v_i(\cdot)) - e] \tag{11.22}$$

which is independent of N by theorem 11.2, and,

$$\operatorname{cov}(W_N, V_N) = \operatorname{cov}[v_i(0), v_i(0) + \delta(v_i(\cdot)) - e] \tag{11.23}$$

Lemma 3. The functions $\delta(v_i(\cdot))^2$ and $v_i(0)\delta(v_i(\cdot))$ are integrable.

Proof. We know that

$$E\delta(v_i(\cdot))^2 \leqslant Ev_i(0)^2 = \sigma^2. \tag{11.24}$$

Let $E\delta(v_i(\cdot))^2 = \alpha\sigma^2$. By virtue of lemma 1 we have $0 < \alpha \leqq 1$. Thus, $E|v_i(0)\delta(v_i(\cdot))| \leqslant Ev_i(0)^2 = \sigma^2$. Q.E.D.
Thus, $Ev_i(0)\delta(v_i(\cdot)) < 0$ from lemma 1, and we can let $Ev_i(0)\delta(v_i(\cdot)) = \beta\sigma^2$ where $-1 \leqq \beta < 0$. From these relations we have immediately that:

$$\operatorname{var}[v_i(0) + \delta(v_i(\cdot)) - e] = \sigma^2(1 + \alpha + 2\beta) - e^2 \tag{11.25}$$

$$\operatorname{cov}[v_i(0), v_i(0) + \delta(v_i(\cdot)) - e] = \sigma^2(1 + \beta) \tag{11.26}$$

By using Varadarajan's theorem and the Lindeberg–Feller theorem as in section 9.8, it is then easy to prove that the vector (V_N, W_N) converges to a normal distribution with mean $(0, 0)$ and matrix of variances and co-variances

$$\Delta = \begin{bmatrix} \sigma^2 & \sigma^2(1 + \beta) \\ \sigma^2(1 + \beta) & \sigma^2(1 + \alpha + 2\beta) - e^2 \end{bmatrix} \tag{11.27}$$

Proof of theorem 11.3. We want to show that

$$\lim_{N\to\infty} \Pr[W_N V_N < 0] > 0 \tag{11.28}$$

Let $W = \lim_{N\to\infty} W_N$

$V = \lim_{N\to\infty} V_N$

(11.28) can be approximated by $\Pr[WV < 0]$ for N large enough. In order for this probability to be zero, and in view of the normality of W and V, it is required that

$$W = kV \qquad \text{for some } k > 0 \tag{11.29}$$

or a matrix of variances and covariances equal to:

$$\begin{bmatrix} \sigma^2 & k\sigma^2 \\ k\sigma^2 & k^2\sigma^2 \end{bmatrix} \tag{11.30}$$

This would require

$$(1 + \alpha + 2\beta) - \frac{e^2}{\sigma^2} = (1 + \beta)^2 \tag{11.31}$$

or,

$$(\alpha - \beta^2)\sigma^2 = e^2 \tag{11.32}$$

Using the definitions of α, β and σ^2, (11.32) can be rewritten as

$$E\delta^2 - \frac{[Ev_i(0)\delta]^2}{Ev_i(0)^2} = [E\delta]^2 \tag{11.33}$$

or

$$\frac{[\operatorname{cov} v_i(0)\delta]^2}{\operatorname{var} v_i(0) \cdot \operatorname{var} \delta} = 1 \tag{11.34}$$

which implies the colinearity of δ and $v_i(0)$.

By virtue of lemma 1, above, we can see that as long as $v_i(0)$ is not surely signed, δ will be zero with positive probability (when $v_i(0) < 0$) and δ will satisfy (11.16) with positive probability (when $v_i(0) \geqq 0$). As long as $v_i(\cdot)$ is not a constant function in the latter case, $\delta = 0$ will not satisfy (11.16). Therefore δ and $v_i(0)$ cannot be linearly related. Thus (11.34) is false and the matrix in (11.27) fails to have the property which would imply

$$\Pr(WV < 0) = 0.$$

Q.E.D.

The limiting probability that the mechanism will take a decision contrary to the sign of $\sum_i v_i(0)$ is:

$$\frac{|\Delta|^{-\frac{1}{2}}}{\pi} \int_{-\infty}^{0} \int_{0}^{\infty} e^{-\frac{1}{2}(V, W)\Delta^{-1}\binom{V}{W}} \mathrm{d}V \, \mathrm{d}W \tag{11.35}$$

which is a positive number, but of course less than $\frac{1}{2}$. The mechanism (11.12, 11.14) therefore makes some progress towards solving the incentive problem, as a naive procedure would make a mistake half the time, but it does not entirely overcome it.

Chapter 12

ECONOMIES IN ACQUIRING INFORMATION THROUGH SAMPLING

12.1. Costs of mechanisms

There are several types of direct costs associated with the use of the dominant-strategy mechanisms. First, the process of elicitation of responses and their compilation might involve the expenditure of real resources, which as a first approximation can be thought to be proportional to the population concerned, that is

$$C = cN \tag{12.1}$$

where C is the direct cost of the mechanism when the population is of size N.

A second type of cost is that which is due to the income transfers induced by the mechanism. If we assume that the procedure is used on a population for which the existing income distribution is optimal, then a bound on the expected costs of the income transfers is

$$C = \sum_i |t_i(w)| \tag{12.2}$$

An obvious way of limiting these costs is to ask only a sample of the population instead of the whole population; but this procedure risks making a wrong decision.

In this chapter we consider the potential value of taking a random sample from the population, eliciting the tastes of its members via a dominant strategy mechanism and making the decision for the entire group on the assumption that the sample's preferences are representative. We are concerned with two questions in this regard: First, we investigate properties of the optimal sample size. Second, we study whether and to what extent the resulting mechanism, whose outcomes and costs may be regarded as random functions of the true preferences, can be considered as a solution to the free-rider problem. More specifically, we compare the expected value of using this procedure, with the optimal sample size, to having perfect information or to being constrained to make the decision on the basis of the a priori beliefs of the decision maker. To the extent the imperfectness of information concerning tastes can be overcome by these methods, we can say that progress towards the solution of the free-rider problem has been made.

12.2. Imperfect information of the planner

The benefit of a large sample is to reduce the risk of accepting or rejecting the project when the opposite action would have been superior. To quantify this value, it is necessary to suppose the decision maker has prior beliefs concerning the preferences of the population which will be modified in light of the sample. For mathematical simplicity we assume that he believes that the N individuals' willingnesses-to-pay, v_i, are independently and identically normally distributed

$$v_i \sim N(m, s^2). \tag{12.3}$$

Moreover, he is uncertain about the mean, m, but (for simplicity) knows the variance. His prior belief on the parameter m is described by

$$m \sim N(\mu, \sigma^2). \tag{12.4}$$

When a sample of size n is taken, the planner revises his beliefs about m according to Bayes' theorem.

Letting $\overline{w} = \sum_i (w_i/n)$ be the mean of the sample's responses, we have that the posterior distribution is described by

$$m \sim N\left(\frac{s^2\mu + n\sigma^2\overline{w}}{s^2 + n\sigma^2}, \frac{\sigma^2 s^2}{s^2 + n\sigma^2}\right) \tag{12.5}$$

We will assume throughout that the planner is risk neutral. Thus the per capita expected value of constructing the project for the unsampled agents is $(s^2\mu + n\sigma^2\overline{w})/(s^2 + n\sigma^2)$.

The project is constructed only if the preferences stated responses indicate that it is valuable. That is, in the process of eliciting the true preferences, the planner announces the mechanism to be used, including the rule for project acceptance, $\sum_i w_i \geq 0$. The spirit of this analysis is that the mechanism is then followed, even though this might not be optimal

ex post, as would be the case if μ and w were of opposite signs but $|\mu| > (n\sigma^2/s^2)|\overline{w}|$, that is, an ex post mean of a different sign than $\overline{w}$.

An alternative procedure is to introduce the government as an artificial player whose expressed preferences for the project are minus its cost. If the decision is then made on the basis of the sign of the sum of willingnesses-to-pay, including the government's, the project will be accepted only if the value exceeds the cost. Transfer payments must also be calculated including the government's statement as well, but, if the costs are assumed to be proportional to the number of individuals served by the project, all of the results of this chapter are preserved. Subtracting the per capita cost from individuals' statements, as suggested in the text, leads to a system that is exactly analogous to this one as well.

Thus, the expected value of adopting the project for each of the $(N - n)$ individuals who were not sampled is given by:

$$\int_0^\infty \left(\frac{s^2\mu + n\sigma^2\overline{w}}{s^2 + n\sigma^2}\right) f(\overline{w})\,\mathrm{d}\overline{w} \tag{12.6}$$

where $f(\cdot)$ is the ex ante probability distribution of $\overline{w}$:

$$\overline{w} \sim N\left(\mu, \frac{s^2}{n} + \sigma^2\right) \tag{12.7}$$

For the sampled group, however, there is no residual uncertainty about their evaluations and, per capita, their expected willingness-to-pay for the decision that their responses induce is:

$$\textstyle\int_0^\infty \overline{w} f(\overline{w})\,\mathrm{d}\overline{w} \tag{12.8}$$

Thus the ex ante expected evaluation of the decision, when a sample of size n is drawn from a population of size N is

$$\int_0^\infty \left[(N - n)\frac{s^2\mu + n\sigma^2\overline{w}}{s^2 + n\sigma^2} + n\overline{w}\right] f(\overline{w})\,\mathrm{d}\overline{w} \tag{12.9}$$

or,

$$\int_0^\infty \left[\frac{(N - n)s^2\mu}{s^2 + n\sigma^2} + \frac{(N - n)n\sigma^2 + ns^2 + n^2\sigma^2}{s^2 + n\sigma^2}\overline{w}\right] f(\overline{w})\,\mathrm{d}\overline{w} \tag{12.10}$$

since the project is undertaken only if $\overline{w} \geq 0$.

One can show that this criterion is not strictly optimal in the sampling context, although it is the only way to produce a Pareto optimum relative to the preferences of the sample. Superior results could be attained by using an acceptance criterion of

$$\bar{w} \geqslant -\frac{(N-n)\mu}{Nnl+n} \qquad \text{with } l = \frac{\sigma^2}{s^2} \tag{12.11}$$

which will result in acceptance if and only if the *posterior evaluation of the government* is non-negative. (The cutoff point is then $-(N-n)\mu/(Nnl+n)$ instead of zero.)

In order to maintain the incentive compatibility of the mechanism, an additional subsidy of $(N-n)\mu/(Nnl+n)$ would have to be given to each member of the sampled group in the event that the project is accepted.

This modification would, therefore, affect the choice of the sample size. In general, the sample size and cutoff point would have to be jointly determined at an optimum. One can write down the expected value of the mechanism in this case, paralleling the method of this section, as a function of these two variables. It can be shown that the cutoff point converges to zero when the population grows to infinity and that the asymptotic rate of growth of the sample size is unchanged. Variations in the small sample results depend on the parameters used, but the magnitude of the modification is very minor in all cases.

Let us come back to the computation of the expected value of adopting the project using the intermediary result:

$$\int_0^\infty \bar{w} f(\bar{w})\,\mathrm{d}\bar{w} = \frac{1}{\sqrt{(2\pi)}\,s\left(\dfrac{1}{n}+l\right)^{\frac{1}{2}}} \int_0^\infty \bar{w}\,\mathrm{e}^{-\frac{1}{2}\frac{(\bar{w}-\mu)^2}{s^2\left(\frac{1}{n}+l\right)}}\,\mathrm{d}\bar{w}$$

$$= \frac{s\left(\dfrac{1}{n}+l\right)^{\frac{1}{2}}}{\sqrt{(2\pi)}}\mathrm{e}^{-\frac{\mu^2}{2s^2\left(\frac{1}{n}+l\right)}} + \mu\rho\left(\frac{\dfrac{\mu}{s}}{\left(\dfrac{1}{n}+l\right)^{\frac{1}{2}}}\right) \tag{12.12}$$

We finally obtain:

$$\frac{(N\sigma^2+s^2)}{\sqrt{(2\pi)}\left(\dfrac{s^2}{n}+\sigma^2\right)^{\frac{1}{2}}}\mathrm{e}^{-\frac{\mu^2}{2\left(\frac{s^2}{n}+\sigma^2\right)}} + \mu N\rho\left(\frac{\mu}{\left(\dfrac{s^2}{n}+\sigma^2\right)^{\frac{1}{2}}}\right) \tag{12.13}$$

for the value of (12.10),

where $\rho(\xi) = \dfrac{1}{\sqrt{(2\pi)}}\displaystyle\int_{-\infty}^{\xi} \mathrm{e}^{-\frac{u^2}{2}}\,\mathrm{d}u.$

Let $C(n)$ be the cost associated to a sample of size n. The $V(N, n)$ value of the experiment is given by the difference between (12.13) and $C(n)$.

A naive procedure would be to use only the prior information in deciding whether to do the project. This leads to an ex ante expected value of:

$$N \max[0, \mu]. \tag{12.14}$$

The expected per capita gain of the experiment is then:

$$G(N, n) = \frac{1}{N}V(N, n) - \max[0, \mu] \tag{12.15}$$

which represents the value of the information acquired per capita net of the costs of acquisition.

12.3. Sampling costs constant per capita

We consider first the case of constant per capita sampling cost, that is,

$$C(n) = cn \tag{12.16}$$

for a sample of size n. For simplicity, we treat only in this section, the case of a zero mean μ. Then, the expected value of the decision becomes:

$$\frac{(N\sigma^2+s^2)}{\sqrt{(2\pi)}\left(\dfrac{s^2}{n}+\sigma^2\right)^{\frac{1}{2}}} \tag{12.17}$$

Let $V^1(N, n)$ be the ex ante expected evaluation for this problem,

$$V^1(N, n) = \frac{(N\sigma^2+s^2)}{\sqrt{(2\pi)}\left(\dfrac{s^2}{n}+\sigma^2\right)^{\frac{1}{2}}} - cn \tag{12.18}$$

This problem is concave in n; taking the derivative of $V^1(N, n)$ with respect to n, we obtain:

$$\frac{(N\sigma^2+s^2)s^2}{2\sqrt{(2\pi)}\,n^2\left(\dfrac{s^2}{n}+\sigma^2\right)^{\frac{3}{2}}} - c \tag{12.19}$$

We are interested in computing the rate of growth of the optimal sample size. Therefore we set $n = N^\delta$ and look for the value of δ such that (12.19) is zero for large N. It is easy to show that the positive terms in (12.19) are increasing at the rate of $N^{1-2\delta}$, whereas all of the negative terms are non-increasing as functions of N. Therefore, $\delta = \frac{1}{2}$ is required for the first-order condition. The sample size at the optimum will be proportional to the square root of the population size for large economies.

At this point we should pause to consider this somewhat puzzling result. Sampling theory teaches us that the accuracy of a sample is independent of the population size. One might therefore expect a bounded sample size to be optimal, rather than one growing like the square root of the total population. However, as the population size increases, a given degree of precision is no longer optimal. Any error made now affects a larger group of unsampled agents. Therefore, a higher degree of precision becomes necessary.

The asymptotic per capita value (and here the asymptotic gain) of the experiment is then:

$$\lim_{N\to\infty} \frac{1}{N} V^{\mathrm{I}}(N, n) = \frac{\sigma}{\sqrt{(2\pi)}} \tag{12.20}$$

Clearly, the more uncertain the decision maker is, the more he gains by using the mechanism.

12.4. Sampling costs proportional to total transfer payments

We now neglect the cost of resources used in the sampling process and introduce the cost of distortions induced through the transfers mandated by the preference revelation mechanism. These transfers will, of course, depend upon the particular member of the class of preference revelation mechanisms being employed. In this section we consider two specific mechanisms; our choice is motivated by the results concerning the inducement of participation in the process presented in chapter 6. These mechanisms are defined by

$$h_i(w_{-i}) \equiv 0 \qquad \text{for all } i \tag{12.21}$$

and

$$h_i(w_{-i}) = \min\Big(-\sum_{j\neq i} w_j, 0\Big) \qquad \text{for all } i. \tag{12.22}$$

For the mechanism $h \equiv 0$, the transfer payments are $C(\overline{w}, n) = n(n-1)\overline{w}$ if $\overline{w} \geqslant 0$ and zero otherwise.

As a limiting hypothesis we regard the entire volume of these payments as a dead-weight loss. This is a very pessimistic viewpoint since we will be

neglecting their beneficial effects on the welfare of the sample while fully counting the cost that their collection will impose on the unsampled group.

Actually, using $C(\overline{w}, n)$ as an upper bound on the costs of the mechanism involves a slight underestimate in some cases. The sum of transfers may be written

$$\sum_i \sum_{j\neq i} w_j \tag{12.23}$$

whenever $\sum_i w_i \geqslant 0$. In some situations, $\sum_{j\neq i} w_i$ may be negative for some i even though $\sum_i w_i \geqslant 0$. These individuals are taxed instead of being subsidized by the mechanism. In computing an upper bound on the cost these transfers impose on the system, these taxes should not be counted as a net benefit to the economy. To obtain a bound, we should treat taxes as a dead-weight loss and not count subsidies at all. Therefore, the sum of the taxes imposed on the sampled group is

$$-\sum_i \min\Big(0, \sum_{j\neq i} w_j\Big) \tag{12.24}$$

while the unsampled group must pay a tax of

$$\sum_i \max\Big(0, \sum_{j\neq i} w_j\Big) \tag{12.25}$$

Therefore, total taxes are

$$\sum_i \Big|\sum_{j\neq i} w_j\Big| \tag{12.26}$$

which differs from $C(\overline{w}, n)$ by

$$2\sum_i \min\Big(0, \sum_{j\neq i} w_j\Big) \tag{12.27}$$

Incorporating this modification in our analysis would greatly complicate matters. However, it would not change the asymptotic results in any way.

Since $C(\overline{w}, n)$ grows like n^2, this correction term does not have a significant influence on the asymptotic value of the procedure and does not affect the rate of growth of the optimal sample size. Small sample results would display slightly smaller sample sizes, but the effect is extremely small for all parameter values we have explored.

12.5. Optimal sample sizes

Because of the non-concavity of the objective function for this problem, $V^{\mathrm{II}}(N, n)$, it is not possible to obtain analytically the optimal sample size

$n^*(N)$. However, we are able to bound the rate of increase of $n^*(N)$ from above and to find a lower bound for the value of the experiment which is shown to be a good one in the simulation of section 12.7. This is done by finding the change in the objective function with respect to n, and showing that it is negative in the limit if $n(N) = N^\delta$ with $\delta > \frac{1}{3}$.

Consider first the case where $\mu = 0$. The objective function is then:

$$V^{11}(N, n) = \frac{s}{\sqrt{(2\pi)}} \frac{(Nl + 2) + n(l-1) - n^2 l}{\left(\frac{1}{n} + l\right)^{\frac{1}{2}}} \tag{12.28}$$

The derivative of this expression with respect to n is:

$$\frac{\partial V^{11}}{\partial n}(N, n) = \left[\frac{s}{\sqrt{(2\pi)}}\left(\frac{1}{n} + l\right)^{-\frac{1}{2}}\right] \frac{1}{2n(1 + nl)}[Nl - 4l^2 n^3 + (2l^2 - 7l)n^2 + 3(l-1)n + 2] \tag{12.29}$$

Let

$$L = \frac{s/l}{2\sqrt{(2\pi)}}\left(\frac{1}{n} + l\right)^{-\frac{1}{2}} \tag{12.30}$$

Then

$$\overline{L} = \lim_{n\to\infty} L = \frac{\sigma}{2l^2\sqrt{(2\pi)}}.$$

If $n = N^\delta$, with $\delta > 0$ and N goes to infinity, the dominant terms are $LlN^{1-2\delta}$ among the positive ones and $-4Ll^2 N^\delta$ among the negative ones. Since the positive term is of the order of $N^{1-2\delta}$ and the negative one of the order of N^δ, we have that the derivative is negative whenever $\delta > 1 - 2\delta$ or $\delta > \frac{1}{3}$ as $N \to \infty$.

It is easy to verify that for $\mu = 0$, the objective function is quasi-concave. For N large but finite, an approximation for the zero of the derivative of the objective function is then obtained with:

$$\delta = \frac{1}{3} - \frac{\log 4l}{3 \log N} \tag{12.31}$$

When $\mu \neq 0$, the objective function is more complex. The derivative of $V^{11}(N, n)$ is then:

$$[Le^{-\frac{1}{2}\mu^2/[(s^2/n)+\sigma^2]}] \cdot \frac{1}{2n(1 + nl)} \cdot \left[Nl - 4l^2 n^3 + (2l^2 - 7l)n^2 + 3(l-1)n + 2 - \left\{(Nl - ln^2 + [l-1]n + 2)\frac{\mu^2/s^2}{l + (1/n)}\right\} + \frac{\mu^2}{s^2}(N - n^2 + n)\right] + \mu[1 - 2n]\rho\left\{\frac{\mu}{([s^2/n] + \sigma^2)^{\frac{1}{2}}}\right\} \tag{12.32}$$

We must distinguish two cases according to the sign of μ.

Case I $\mu > 0$

As N goes to infinity, the dominant terms are

$$l\overline{L}^{-\frac{1}{2}(\mu/\sigma)^2} N^{1-2\delta}$$

among the positive ones, and

$$\left[4k^2\overline{L}e^{-\frac{1}{2}(\mu/\sigma)^2} + 2\mu\rho\left(\frac{\mu}{\sigma}\right)\right] \tag{12.33}$$

among the negative ones.

Therefore, just as in the case of $\mu = 0$, we have that the derivative is negative whenever $\delta > \frac{1}{3}$ as $N \to \infty$; and an approximation of the *last zero* of the derivative is obtained with:

$$\delta = \frac{1}{3} - \frac{1}{3 \log N}\log\left[4l + 4l\sqrt{(2\pi)} \cdot \frac{\mu}{\sigma} \cdot \rho\left(\frac{\mu}{\sigma}\right)e^{\frac{1}{2}(\mu/\sigma)^2}\right] \tag{12.34}$$

Case II $\mu < 0$

As N goes to infinity, the dominant terms are

$$l\overline{L}e^{-\frac{1}{2}(\mu/\sigma)^2} N^{1-2\delta} \text{ and } 2|\mu|\rho\left(\frac{\mu}{\sigma}\right)N^\delta$$

among the positive ones, and

$$4l^2\overline{L}e^{-\frac{1}{2}(\mu/\sigma)^2} N^\delta \tag{12.35}$$

among the negative ones. Therefore, if

$$4l^2\overline{L}e^{-\frac{1}{2}(\mu/\sigma)^2} > 2|\mu|\rho\left(\frac{\mu}{\sigma}\right), \tag{12.36}$$

the result is the same as for $\mu > 0$.

The relation (12.36) can be rewritten

$$e^{-\frac{1}{2}(\mu/\sigma)^2} - \frac{|\mu|}{\sigma}\int_{-\infty}^{\mu/\sigma} e^{-\frac{1}{2}\alpha^2} d\alpha > 0 \tag{12.37}$$

Let $y = |\mu|/\sigma$. Condition (12.37) becomes

$$\phi(y) = e^{-\frac{1}{2}y^2} - y\int_{-\infty}^{-y} e^{-\frac{1}{2}\alpha^2}\,d\alpha > 0 \tag{12.38}$$

We have

$$\phi'(y) = -\int_{-\infty}^{-y} e^{-\frac{1}{2}\alpha^2}\,d\alpha < 0 \tag{12.39}$$

and

$$\phi(0) = 1.$$

Moreover, $\lim_{y\to\infty}\phi(y) = 0$.

Therefore, it is clear that ϕ can have no zeros, since it is a decreasing function with a zero asymptote. Hence, (12.36) is always satisfied.

To sum up, for $\mu \neq 0$, the derivative is negative whenever $\delta > \frac{1}{3}$ when $N \to \infty$. However, we do not know in this case if $V(N, n)$ is quasi-concave; we know only that an upper bound of the maximum sample size is of the order of $N^{1/3}$ as $N \to \infty$. The simulation shows that the objective function is often quasi-concave and that (12.34) is a good approximation of the optimal sample size.

12.6. A lower bound for the per capita gain of the experiment

According to section 12.5, an obvious lower bound for the per capita value of the experiment when $N \to \infty$ is obtained by replacing n by $N^{1/3}$ in (12.28):

$$V(N, N^{1/3}) = \frac{(N + N^{1/3} - N^{2/3})l + (2 - N^{1/3})}{\frac{1}{s}(N^{-\frac{1}{3}} + l)^{1/2}\sqrt{(2\pi)}}\, e^{-\frac{1}{2}\frac{(\mu/s)^2}{(N^{-1/3}+l)}} + \mu(N + N^{1/3} - N^{2/3})\rho\left(\frac{\mu}{s(N^{-\frac{1}{3}} + l)^{1/2}}\right) \tag{12.40}$$

or taking the limit when $N \to \infty$

$$\lim_{N\to\infty}\frac{1}{N}V(N, N^{1/3}) = e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2}\frac{\sigma}{\sqrt{(2\pi)}} + \mu\rho\left(\frac{\mu}{\sigma}\right) \tag{12.41}$$

and

$$\lim_{N\to\infty} G(N, N^{1/3}) = e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2}\frac{\sigma}{\sqrt{(2\pi)}} - \mu\left(1 - \rho\left(\frac{\mu}{\sigma}\right)\right) \quad \text{if } \mu > 0$$
$$= e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2}\frac{\sigma}{\sqrt{(2\pi)}} + \mu\rho\left(\frac{\mu}{\sigma}\right) \quad \text{if } \mu < 0. \tag{12.42}$$

Let $B(\mu, \sigma) = \lim_{N\to\infty} G(N, N^{1/3})$ be this bound on the asymptotic per capita gain. For $\mu > 0$, $B(\mu, \sigma)$ is

$$\frac{1}{\sqrt{(2\pi)}}\left[\sigma e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} - \mu\int_{\mu/\sigma}^{\infty} e^{-\frac{\alpha^2}{2}}\,d\alpha\right] \tag{12.43}$$

Writing $y = \mu/\sigma$, this is proportional to

$$e^{-\frac{y^2}{2}} - y\int_{y}^{\infty} e^{-\frac{\alpha^2}{2}}\,d\alpha \tag{12.44}$$

which is greater than or equal to

$$e^{-\frac{y^2}{2}} - \int_{y}^{\infty}\alpha e^{-\frac{\alpha^2}{2}}\,d\alpha = 0 \tag{12.45}$$

since $\alpha > y$ throughout the range of integration. The case of $\mu < 0$ can be treated in a parallel manner, obtaining the same result.

The loci of $B(\mu, \sigma) =$ constant can be analyzed as follows

$$\frac{\partial B(\mu, \sigma)}{\partial\sigma} = B_\sigma = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2} \geqslant 0$$
$$\frac{\partial B(\mu, \sigma)}{\partial\mu} = B_\mu = \frac{-1}{\sqrt{(2\pi)}}\int_{\mu/\sigma}^{\infty} e^{-\frac{\alpha^2}{2}}\,d\alpha \leqslant 0 \tag{12.46}$$

Since $B(\mu, \sigma)$ is homogeneous in μ/σ, the locus of $B = 0$ in the μ, σ-plane is linear. Therefore, because B_μ and B_σ are one-signed, the locus of $B = 0$ must be the horizontal axis. Furthermore,

$$\frac{\partial^2 B(\mu, \sigma)}{\partial\mu^2} = B_{\mu\mu} = \frac{1}{\sigma\sqrt{(2\pi)}}e^{-\frac{1}{2}(\mu/\sigma)^2}$$
$$\frac{\partial^2 B(\mu, \sigma)}{\partial\sigma^2} = B_{\sigma\sigma} = \frac{\mu^2}{\sigma^3}\cdot\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}(\mu/\sigma)^2} \tag{12.47}$$
$$\frac{\partial^2 B(\mu, \sigma)}{\partial\mu\partial\sigma} = B_{\mu\sigma} = -\frac{\mu}{\sigma^2}\cdot\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}(\mu/\sigma)^2}$$

Along $B(\mu, \sigma) =$ constant

$$\frac{d\sigma}{d\mu} = -\frac{B_\mu}{B_\sigma} > 0$$

$$\frac{d^2\sigma}{d\mu^2} = -\frac{B_\sigma^2 B_{\mu\mu} - 2B_\mu B_\sigma B_{\mu\sigma} + B_\mu^2 B_{\sigma\sigma}}{B_\sigma^3}$$

$$= -\frac{\frac{1}{\sigma}\left\{e^{-\frac{1}{2}(\mu/\sigma)^2} - \frac{\mu}{\sigma}\int_{\mu/\sigma}^{\infty} e^{-\frac{\alpha^2}{2}}\,d\alpha\right\}^2}{e^{-(\mu/\sigma)^2}} < 0 \qquad (12.48)$$

This gives rise to the following diagram, figure 12.1.

Thus, for every $\sigma > 0$, we have a positive lower bound for the *limiting* per capita value of the optimal elicitation procedure. However, as shown by the simulation, for large (absolute) values of μ, the mechanism may require a very large population to confirm this asymptotic result. When $\sigma = 0$, since the prior has no uncertainty, the optimal sample size is clearly zero, and any utilization of this mechanism can only waste resources.

12.7. Simulations

The main purpose of the simulation is to check if the upper bound obtained for the optimal sample size and the lower bound obtained for the per capita gain of the experiment are tight bounds which can be used as approximations. This is necessary because of the inherent non-convexities in the objective function $V^{11}(N, n)$.

We compute the value of $V^{11}(N, n)$ for $s = 1$, $\sigma = .1$, $l = .01$ and different

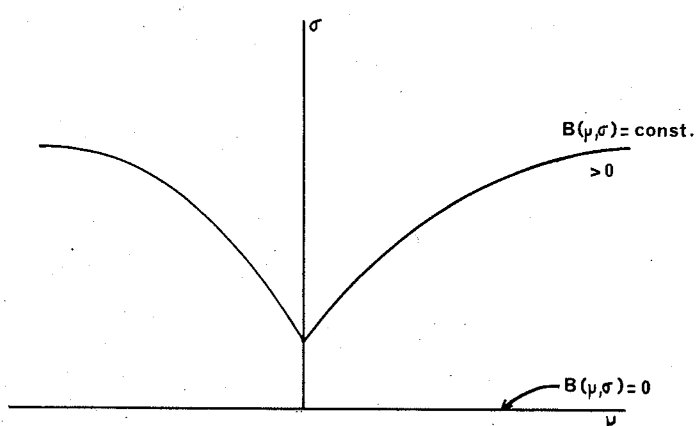

Figure 12.1.

values of μ and N. Then, we identify the optimal sample size and compute the per capita gain in information.

Figure 12.2 pictures the typical evolution of the optimal sample size. Figure 12.3 represents the per capita informational gain of the procedure.

We observe that the optimal sample size converges to $\alpha N^{1/3}$. The last zero which was approximated in the analytical derivations corresponds to the optimum (figure 12.2).

As expected, figure 12.3 shows that the experiment is the most valuable when μ is small in absolute value with respect to σ; this corresponds to very uncertain prior beliefs. For large values of μ, the experiment becomes valuable only for large populations. Note also that the per capita informational gain is almost symmetrical around the value of $\mu = 0$. However, to achieve the same informational gain, the sample size is larger for negative values of μ than for positive values. Therefore, if there were a per unit sample cost, the experiment would be less valuable for negative than positive values of μ. Sufficiently large values of N are checked to verify the finding of section 6, that the asymptotic per capita gain is always positive when $\sigma > 0$.

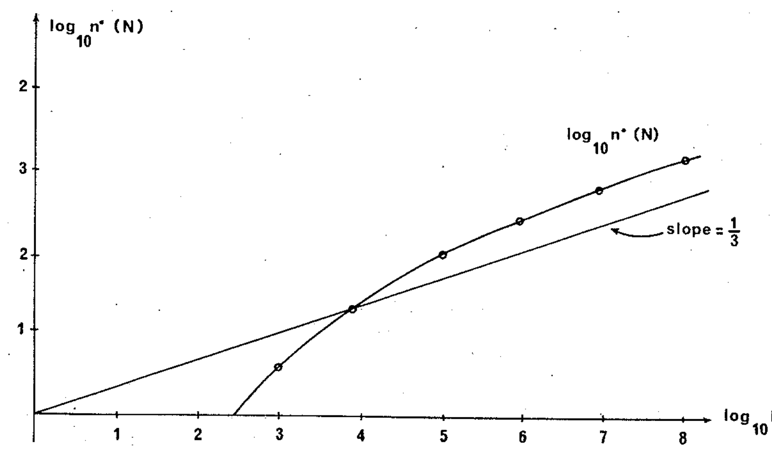

Figure 12.2.

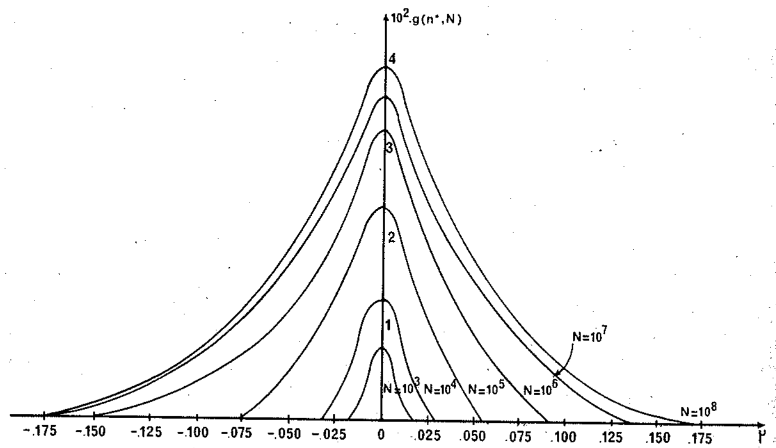

Figure 12.3.

12.8. Sample sizes and asymptotic gains for the Clarke mechanism

In order to improve upon the results obtained above, we study the Clarke mechanism defined by $h = \min(-\sum_{j \neq i} w_j, 0)$. This mechanism has the property that an individual is taxed whenever the decision is altered on account of his statement. In this case he pays $|\sum_{j \neq i} w_j|$, which is the cost that his participation and revealed preferences have imposed on the other members of the group. Note, however, that in those cases in which the project is overwhelmingly carried, that is, when the sum $\sum_i w_i$ is larger than any single evaluation, w_i, no taxes will be paid by anyone. This is to be contrasted with the situation for the $h \equiv 0$ mechanism which would produce a total transfer of $(n-1)\sum_i w_i$.

It is not true that the Clarke mechanism involves less transfer payments in every case than $h \equiv 0$ would. Consider the following three-person situation:

$$v_1 = +9; \quad v_2 = -5; \quad v_3 = -5 \tag{12.49}$$

A straightforward calculation reveals that under $h = \min(-\sum_{j \neq i} w_j, 0)$, the transfers received are:

$$t_1 = 0; \quad t_2 = -4; \quad t_3 = -4; \quad \text{so that } \sum_i |t_i| = 8 \tag{12.50}$$

while under $h = 0$, they are,

$$t_1 = 0; \quad t_2 = 0; \quad t_3 = 0, \quad \text{so that } \sum_i |t_i| = 0. \tag{12.51}$$

However, in an expected sense, the payments are much lower with the Clarke mechanism. We know from theorem 9.1, that if the v_i are independently identically distributed according to a distribution with finite variance, then the mechanism produces transfers t_i always non-positive such that $(1/\sqrt{(N)})E\sum_i t_i$ approaches a finite limit.

Following our earlier procedure, we can define the ex ante expected value of using this mechanism on a sample of size n. In the normal case, one can calculate for $\mu = 0$ that the expected transfers $C(n)$ are approximated by (see chapter 9)

$$\frac{s}{2\sqrt{(2\pi)}}\left[\sqrt{(n)} - \frac{1}{4\sqrt{(n)}} \cdots\right] \tag{12.52}$$

The ex ante value of the mechanism is therefore

$$V^{\mathrm{III}}(N, n) = \frac{N\sigma^2 + s^2}{\sqrt{(2\pi)}\left(\dfrac{s^2}{n} + \sigma^2\right)^{1/2}} - C(n) \tag{12.53}$$

Calculations analogous to the ones described in section 12.3 give the order of the optimal sample size, namely $N^{2/3}$. The expected transfer costs are much smaller with the pivotal mechanism than with the mechanism associated to $h \equiv 0$. Consequently, the optimal sample size is much higher. Indeed, with this mechanism, if we introduced simultaneously per capita sample costs and transfers costs, the per capita sample costs would dominate in the determination of the optimal sample size.

12.9. Conclusion

In concluding this chapter, we can safely say that as long as individuals have perfect information regarding their own tastes, the different costs of acquiring information through sampling with Groves mechanisms are not a significant obstacle to the use of these mechanisms, as a democratic basis for public decision making. Nevertheless, the lack of public perfect information reduces per capita welfare below what could be attained if it were available. The free rider problem has not disappeared, but its effects can be somewhat mitigated.

Chapter 13

IMPERFECT PERSONAL INFORMATION

13.1. Introduction

It is an important factual observation that, for most public projects, agents do not know their own willingness-to-pay precisely. This situation is certainly due in part to the fact that people's direct participation in public decision making is rarely solicited; however, it is also inherent to the problem for reasons hinted at in previous chapters. Public projects often provide new and complex consumption goods that it is difficult to appreciate without experiencing them. Moreover, a precise description of the public project is often not enough to have an idea of its ultimate consequences which might have to be worked out through complex general equilibrium channels.

We are then quickly faced with the following dilemma. On one hand, if the strength of incentives is too weak, agents will not invest sufficient real resources to ascertain their personal valuations accurately. In such a case, even the best revelation mechanisms would be pointless, since the agents' answers are essentially of no value. On the other hand, if the strength of incentives is made too great, excessive resources may be devoted to private information acquisition. Is it possible to design a mechanism with a strength of incentive strong enough so that there exists a valuable decentralized public decision making procedure between these two extremes?

Since the family of Groves mechanisms characterizes truthful mechanisms, a negative answer given to the above question, using Groves mechanisms, would be bad news for democratic public decision making. This is the reason why we study this question in great detail, using alternative formalizations in an attempt to cover all its aspects.

13.2. The strength of incentives

We consider an agent who is in a sample of size n and who is asked, through a Clarke mechanism, to reveal his evaluation of a given public project, the cost of which is assumed to be zero for simplicity of notation. Let x denote the sum of the answers of the $(n-1)$ other agents. From the definition of the Clarke mechanism we see that agent i's gain, when saying the truth, v_i, instead of w_i is:

$$0 \qquad \begin{matrix} \text{if } v_i + x \geqslant 0 \text{ and } w_i + x \geqslant 0 \quad \text{or} \\ \text{if } v_i + x < 0 \text{ and } w_i + x < 0 \end{matrix}$$

since v_i and w_i lead to the same social decision and private goods allocation;

$$v_i + x \qquad \text{if } v_i + x \geqslant 0 \text{ and } w_i + x < 0,$$

that is, if his statement defeats the project when he should not have done so; and

$$-v_i - x \qquad \text{if } v_i + x < 0 \text{ and } w_i + x \geqslant 0,$$

that is, when his answer leads to acceptance of the project when he should have killed it.

Let $P^n(x)$ be agent i's subjective distribution function over the sum of the answers of the $(n-1)$ other agents. If agent i is risk neutral, his expected gain of saying v_i instead of w_i is, for $v_i \geqslant 0$, given by,

$$\int_{\substack{-v_i \leqslant x < -w_i \\ -w_i \leqslant x < -v_i}} |v_i + x| \mathrm{d}P^n(x) \tag{13.1}$$

that is,

$$\int_{-v_i \leqslant x < -w_i} (v_i + x)\mathrm{d}P^n(x) \qquad \text{if } v_i > w_i \tag{13.2}$$

and

$$-\int_{-w_i \leqslant x < -v_i} (v_i + x)\mathrm{d}P^n(x) \qquad \text{if } v_i < w_i$$

Observe that (13.1) is bounded above by

$$\begin{matrix} [v_i - w_i] \cdot \Pr[-v_i \leqslant x < -w_i] & \text{if } v_i > w_i \\ [w_i - v_i] \cdot \Pr[-w_i \leqslant x < -v_i] & \text{if } v_i < w_i \end{matrix} \tag{13.3}$$

An analogous expression can be derived for $v_i < 0$. Clearly, as n grows, the probability that x belongs to a fixed interval $[-w_i, -v_i]$ (or $[-v_i, -w_i]$) decreases for n large enough. For example, if the subjective probability distribution of x is normal with mean zero and variance $n-1$, we have:

$$\Pr[-w_i \leqslant x < -v_i] = \frac{1}{\sqrt{(2\pi)}} \int_{-w_i}^{-v_i} \frac{\mathrm{e}^{-\frac{x^2}{2(n-1)}}}{\sqrt{(n-1)}} \mathrm{d}x \tag{13.4}$$

The density $\mathrm{e}^{-x^2/2(n-1)}/\sqrt{(n-1)}$ converges uniformly to zero as n goes to infinity. Consequently, $\Pr[-w_i \leqslant x < -v_i]$ converges to zero as n tends to infinity.

More generally, if $y = x/\sqrt{(n)}$, let $\psi(y)$ be the continuous density function of y

$$\Pr[-w_i \leqslant x < -v_i] = \Pr\left[-\frac{w_i}{\sqrt{(n)}} \leqslant y < -\frac{v_i}{\sqrt{(n)}}\right]$$
$$\leqslant \frac{(v_i - w_i)}{\sqrt{(n)}} \sup_{y \in \left[-\frac{w_i}{\sqrt{(n)}}, -\frac{v_i}{\sqrt{(n)}}\right]} \psi(y) = \frac{(v_i - w_i)k(n)}{\sqrt{(n)}} \tag{13.5}$$

where $k(n)$ goes to $\psi(0)$ as n goes to infinity.

Therefore, the strength of incentive, measured as the expected gain of saying the truth instead of any other fixed answer w_i, goes to zero as n goes to infinity, at the speed of $1/\sqrt{(n)}$.

13.3. Individual behavior

We suppose now that the agent does not know his own evaluation exactly, but that he can acquire costly information about the project itself and about his own preferences. This uncertainty is here formalized in the following way: each agent has a prior belief concerning his own evaluation given by a normal distribution $N(0, r^2)$. His information gathering process is simplified to the extreme; if he spends an amount c, he discovers his true evaluation v. Otherwise, he gives an answer which minimizes his expected regret, that is, the expected short-fall in utility attained versus perfect information. We assume that the ability to process information differs among the agents and that there exists in the sample a distribution function $J(c)$ which describes the distribution of costs necessary to discover the truth.

Let v_i be the true evaluation of agent i; if he buys information at the cost c_i, he answers v_i, since we know that the Clarke mechanism induces the response v_i as a dominant strategy.

The regret of saying w_i when the truth is v_i equals the expected gain of saying v_i instead of w_i when the truth is v_i. In view of section 13.2, this is,

$$R(v_i, w_i) = \int_{\substack{w_i < -x \leq v_i \\ -w_i \leq x < -v_i}} |v_i + x| p^n(x)\,dx \tag{13.6}$$

where $p^n(x)$ is agent i's subjective probability density over the sum of the others' answers. We assume that the agent's beliefs are restricted to the family of normal distributions. The mean and the variance depend on his expectations about who buys information. Suppose that he believes that an agent who does not buy information gives a non-random answer s and that there are n_1 such individuals. Let $S = n_1 s$. Then $x = S + \sum_{i=1}^{n_2} v_i$, where n_2 is the number of other agents in the sample who do buy information. Agent i believes that the others' true evaluations are independent and identical to his expectations concerning his own evaluation. Consequently,

$$\begin{aligned} Ex &= S \\ \operatorname{var} x &= n_2 r^2 \end{aligned} \tag{13.7}$$

Therefore,

$$p^n(x) = \frac{1}{r\sqrt{(2\pi n_2)}} e^{-\frac{1}{2}\frac{(x-S)^2}{n_2 r^2}} \tag{13.8}$$

Theorem 13.1. Whatever his expectations about the answers of the agents who do not buy information, if an agent i does not buy information, he answers the mean of his prior, 0. Consequently, he is led to believe that the other agents who do not buy information also answer 0; that is, $S = 0$.

Proof. Consider agent i who does not buy information. His optimal response is determined by the minimization of expected regret:

$$\min_{w_i} \int_{-\infty}^{+\infty} R(v_i, w_i) g(v_i)\,dv_i \tag{13.9}$$

where $g(v_i)$ is his prior density function, $(1/r\sqrt{(2\pi)})e^{-v_i^2/2r^2}$.

Let us first compute $R(v_i, w_i)$.

If $v_i > w_i$, then:

$$\begin{aligned} R(v_i, w_i) &= \int_{-v_i}^{-w_i} (v_i + x) \frac{1}{r\sqrt{(2\pi n_2)}} e^{-\frac{1}{2}\frac{(x-S)^2}{n_2 r^2}} dx \\ &= (v_i + S)\left[\rho\left(\frac{-w_i - S}{r\sqrt{(n_2)}}\right) - \rho\left(\frac{-v_i - S}{r\sqrt{(n_2)}}\right)\right] \\ &\quad + \frac{r\sqrt{(n_2)}}{\sqrt{(2\pi)}}\left[e^{-\frac{(v_i+S)^2}{2n_2r^2}} - e^{-\frac{(w_i+S)^2}{2n_2r^2}}\right] \end{aligned} \tag{13.10}$$

where

$$\rho(y) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{y} e^{-\frac{u^2}{2}} du$$

The same result obtains for $v_i < w_i$. Therefore, the expected regret is:

$$\int_{-\infty}^{+\infty} \left\{ (v_i + S)\left[\rho\left(\frac{-w_i - S}{r\sqrt{(n_2)}}\right) - \rho\left(\frac{-v_i - S}{r\sqrt{(n_2)}}\right)\right] + \frac{r\sqrt{(n_2)}}{\sqrt{(2\pi)}}\left[e^{-\frac{(v_i+S)^2}{2n_2r^2}} - e^{-\frac{(w_i+S)^2}{2n_2r^2}}\right]\right\} \frac{e^{-\frac{v_i^2}{2r^2}}}{r\sqrt{(2\pi)}} dv_i \tag{13.11}$$

We differentiate with respect to w_i and obtain the first-order condition,

$$\frac{e^{-\frac{(w_i+S)^2}{2n_2r^2}}}{\sqrt{(2\pi n_2)}} \int_{-\infty}^{+\infty} (w_i + S - v_i - S) \frac{e^{-\frac{v_i^2}{2r^2}}}{r\sqrt{(2\pi)}} dv_i = 0 \tag{13.12}$$

Hence, $w_i = 0$. Q.E.D.

Therefore, we know that if he buys information, the agent i answers the truth v_i, and if he does not buy information, he answers 0. The remaining problem is: how does he decide whether or not to buy information?

He will buy information if the expected regret when answering 0 is larger than his cost c_i. We take an equilibrium theoretic point of view in this matter, and assume that each individual makes a decision, based on his own cost, in the belief that all other individuals are performing the same calculation relative to their own cost. An equilibrium situation will be one in which these beliefs are verified by the actual decisions made. Recall that $J(\cdot)$ denotes the distribution of c, and let $j(\cdot)$ be its density.

If we let $S = 0$ in (13.11), we obtain:

$$\begin{aligned} E_{v_i} R(v_i, 0) &= 2\int_0^{\infty} \left[v_i\left[\frac{1}{2} - \rho\left(-\frac{v_i}{r\sqrt{(n_2)}}\right)\right] + \frac{r\sqrt{(n_2)}}{\sqrt{(2\pi)}}\left(e^{-\frac{v_i^2}{2n_2r^2}} - 1\right)\right] \frac{e^{-\frac{v_i^2}{2r^2}}}{r\sqrt{(2\pi)}} dv_i \\ &= \frac{r}{\sqrt{(2\pi)}}[\sqrt{(n_2 + 1)} - \sqrt{(n_2)}] \end{aligned} \tag{13.13}$$

as may be seen by complex but straightforward manipulations.

Since all agents face the same problem, there must be a cutoff point c^* beyond which it is not worth it to buy information. Then,

$$n_2 = n\int_0^{c^*} \mathrm{d}J(c) = nJ(c^*) \tag{13.14}$$

Therefore, c^* is a solution of the equation

$$\frac{r}{\sqrt{(2\pi)}}[\sqrt{(nJ(c^*)+1)} - \sqrt{(nJ(c^*))}] = c^* \tag{13.15}$$

The agent must obtain the cutoff point c^* and then compare his own c_i to c^*. If $c_i > c^*$, he does not buy information and answers 0. If $c_i < c^*$, he buys information, he obtains the truth v_i and he answers by revealing it.

Theorem 13.3. There is a unique equilibrium cutoff point.

Proof. Let

$$\phi(c) = \frac{r}{\sqrt{(2\pi)}}[\sqrt{(nJ(c)+1)} - \sqrt{(nJ(c))}]$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}c}(c) = \frac{r}{2\sqrt{(2\pi)}}\left[\frac{nj(c)}{\sqrt{(nJ(c)+1)}} - \frac{nj(c)}{\sqrt{(nJ(c))}}\right] < 0 \tag{13.16}$$

Therefore, there is a unique cutoff point (see fig. 13.1). Q.E.D.

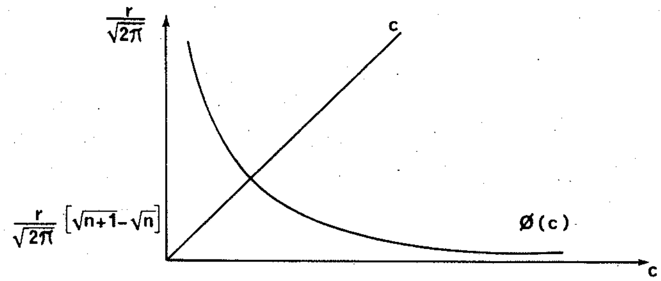

Figure 13.1.

Moreover, the cutoff point is stable in the following sense. $ER(v_i, 0)$ is a decreasing function of c. To obtain the equilibrium c^*, the agent can start from c, compute a number, n_2, of agents buying information by $n_2 = n\int_0^c \mathrm{d}J(c)$. Then, if $ER(v_i, 0) > c$, the expected cost of not buying information is larger than c, suggesting that the cutoff point is larger, etc.

Theorem 13.4. The cutoff point, c^*, goes to zero as the sample size goes to infinity; when the uncertainty of the agents increases (r^2 increases) the cutoff point increases.

Proof. It is easy to check by differentiation of (13.15) that $\mathrm{d}c^*/\mathrm{d}r > 0$. Suppose that c^* converges to $\bar{c} \neq 0$ when n goes to infinity; then the left-hand side of (13.15) goes to zero, a contradiction since the right-hand side goes to $\bar{c}$. Q.E.D.

In order to obtain a closed form relation between the equilibrium cutoff point and the sample size, we first use the approximation,

$$\sqrt{(n_2+1)} - \sqrt{(n_2)} \approx \frac{1}{2\sqrt{n_2}} \quad \text{for } n_2 \text{ large.} \tag{13.17}$$

Equation (13.15) then becomes:

$$\frac{r}{2\sqrt{(2\pi)}}\frac{1}{\sqrt{(nJ(c^*))}} = c^* \tag{13.18}$$

In the next section we consider the special cases of distributions of ability to process information such that:

$$J(c) = \frac{c^\alpha}{A_\alpha}, \ \alpha > 0, \ A_\alpha > 0 \tag{13.19}$$

For example, if $\alpha = 1$, this is the uniform distribution on $[0, A_1]$. More generally, the αth distribution is concentrated on $[0, a_\alpha]$ where $a_\alpha = A_\alpha^{1/\alpha}$ and c^* is given by

$$c^* = \left[\frac{r}{2\sqrt{(2\pi)}}\right]^{\frac{2}{2+\alpha}} A_\alpha^{\frac{1}{2+\alpha}} n^{-\frac{1}{2+\alpha}} \tag{13.20}$$

13.4. Optimal sample size

The role of the decision maker is to determine the sample size, on the basis of his prior information, to maximize the pre-sample expected valuation of the project minus the induced private informational costs.

We assume that the decision maker approximates the distribution of evaluations in the population of size N by a normal distribution $N(m, s^2)$ with an unknown mean m on which he has a prior $N(0, \sigma^2)$ such that $s^2 + \sigma^2 = r^2$, to ensure that the decision maker has the same information as any other agent.

The decision maker knows that the agents who do not buy information answer zero.

Let $\xi = (\sum_{i=1}^{n_2} v_i)/n_2$ the mean of the answers of the agents who buy information. The posterior mean in the subsample of size n_2 is ξ. The posterior mean for the entire population is

$$\frac{n_2\sigma^2\xi}{s^2 + n_2\sigma^2} \tag{13.21}$$

Therefore, for a decision maker who is assumed to be risk neutral, the *post sample* evaluation of the project is:

$$(N - n_2)\left(\frac{n_2\sigma^2\xi}{s^2 + n_2\sigma^2}\right) + n_2\xi = \frac{(N\sigma^2 + s^2)\xi}{\frac{s^2}{n_2} + \sigma^2} \tag{13.22}$$

The ex ante distribution of ξ is clearly a normal distribution of mean zero and variance $s^2/n_2 + \sigma^2$.

The decision maker accepts the project if the sum of the answers is positive, $\xi \geqslant 0$. Therefore, the presample evaluation of the project is:

$$\frac{1}{\sqrt{(2\pi)}} \int_{\xi \geqslant 0} \frac{(N\sigma^2 + s^2)}{\left(\frac{s^2}{n_2} + \sigma^2\right)^{\frac{3}{2}}} \xi \, e^{-\frac{1}{2}\frac{\xi^2}{\left(\frac{s^2}{n_2}+\sigma^2\right)}} \, d\xi \tag{13.23}$$

$$= \frac{N\sigma^2 + s^2}{\sqrt{(2\pi)}\sqrt{\left(\frac{s^2}{n_2} + \sigma^2\right)}}$$

with $n_2 = nJ(c^*(n))$.

The expected private informational costs are:

$$n \int_0^{c^*(n)} c\, dJ(c) \tag{13.24}$$

The net value of the experiment with a sample size n is:

$$V(N, n) = \frac{N\sigma^2 + s^2}{\sqrt{(2\pi)}\sqrt{\left(\frac{s^2}{nJ(c^*(n))} + \sigma^2\right)}} - n \int_0^{c^*(n)} c\, dJ(c) \tag{13.25}$$

The optimal n is a solution of

$$\frac{\partial V}{\partial n}(N, n) = 0 \qquad \text{if } N > n > 0 \tag{13.26}$$

To gain some intuition regarding this solution, we use the approximation developed at the end of section 2.

Since $J(c) = c^\alpha/A_\alpha$ for $0 \leqslant c \leqslant A_\alpha^{1/\alpha}$,

$$n \int_0^{c^*(n)} c\, dJ(c) = \frac{n\alpha}{\alpha + 1} \cdot \frac{c^*(n)^{\alpha+1}}{A_\alpha} \tag{13.27}$$

The presample evaluation of the project is then:

$$V(N, n) = \frac{N\sigma^2 + s^2}{\sqrt{(2\pi)}\sqrt{\left(\sigma^2 + \frac{8\pi}{r^2}s^2c^*(n)^2\right)}} - \frac{n\alpha}{\alpha + 1} \cdot \frac{c^*(n)^{\alpha+1}}{A_\alpha} \tag{13.28}$$

with

$$c^*(n) = \left[\frac{r}{2\sqrt{(2\pi)}}\right]^{\frac{2}{2+\alpha}} A_\alpha^{\frac{1}{2+\alpha}} n^{\frac{-1}{2+\alpha}}$$

Differentiating with respect to n and equating to zero, we obtain the optimal sample size. For $\alpha = 1$ (uniform distribution) this is:

$$n \approx \frac{8\sqrt{(2\pi)}A_1 s^2}{\sigma r^2} \cdot N \tag{13.29}$$

More generally, we can ascertain the rate of growth of the optimal sample size for $\alpha < 1$ by differentiating $V(N, n)$ with respect to n and setting $n = N^\delta$. The value of δ for which this derivative is zero asymptotically is $\delta = (2 + \alpha)/3$.

13.5. The informational value of processes with imperfect personal information

In the previous sections we have deduced that the optimal sample size should grow at the rate of $N^{(2+\alpha)/3}$, where N is the population size and $0 \leqslant \alpha \leqslant 1$ is the parameter of the structure of private information costs as we have specified it. In this way the three conflicting problems of inaccurate answers, sampling variance and private information gathering costs are traded off to best advantage.

Because it involves real resource costs, however, it is necessary to demonstrate that the procedure is of positive value to society. More specifically, as N is large, we want to ascertain the asymptotic rate of growth of the net value of the procedure. A genuine informational advantage of this method can be said to follow if this value is positive and if its limit is still strictly positive on a per capita basis. As we increase the population size in our conceptual experiment, we are implicitly increasing the scale of the public project under consideration – therefore the *per capita* informational value is a sign of the actual informational value of the mechanism in the presence of inaccurate, costly, personal information regarding preferences.

Substituting into V, $n = N^{(2+\alpha)/3}$, we obtain only a lower bound on the net gain, since the true optimal sample size is of this order but may differ from it by lower order terms. This gives:

$$V(N, n^*) \geqslant \frac{N\sigma^2 + s^2}{\sqrt{(2\pi)}\sqrt{\left(\sigma^2 + \frac{8\pi A^2 s^2}{r^2} N^{-2/3}\right)}} - \frac{\alpha}{\alpha+1}\frac{A^{\alpha+1}}{A_\alpha} N^{1/3} \tag{13.30}$$

where

$$A = \left(\frac{r}{2\sqrt{(2\pi)}}\right)^{\frac{2}{2+\alpha}} A_\alpha^{-\frac{1}{2+\alpha}}$$

The first term is positive and its denominator is decreasing in N, approaching $\sigma\sqrt{(2\pi)}$. The second, a negative term, is growing at the rate $N^{1/3}$.

Therefore, $V(N, n^*)$ is asymptotically positive and, moreover,

$$\lim_{N\to\infty} \frac{V(N, n^*)}{N} = \frac{\sigma}{\sqrt{(2\pi)}} > 0. \tag{13.31}$$

13.6. Summary of this model

We can now summarize the working of this simple model as follows. When the sample size n increases, the proportion of agents buying information decreases since $c^*(n)$ decreases. However, the absolute number of agents buying information increases since $nJ(c^*(n)) \approx n^{2/(2+\alpha)}$. Because in this model the agents who do not buy information do not introduce noise in the answers (they are easily recognized since they answer 0 and hence are effectively deleted from the sample), it is clear that without informational costs the optimal sample size would be N. However, the total private informational costs increase as $n^{1/(\alpha+2)}$, so that there is an optimal size of the order of $N^{(\alpha+2)/3}$.

Since we use only answers given by agents with low c_i, there may still be a difficulty if there is a correlation between the v_i and the c_i. The conditional expectation of v, given that c is larger than a cutoff point c^*, may be different from the unconditional expectation in which we are interested. It is not easy to correct the induced bias. We might draw a subsample to estimate the correlation, by compensating everybody for his information costs, and use this estimation to suppress the bias in the large sample. But then we would face the difficulty that in the subsample, truthful revelation would no longer be a dominant strategy. The costly solution of paying everybody in the random sample faces another type of difficulty. Information costs can be divided into information gathering costs and information processing costs. It is clear that the decision maker can pay the information gathering costs, or by disseminating information take advantage of some informational returns to scale. But no increasing returns seem to exist in the activity of information processing, and more importantly there is no way to be sure that, even if they were compensated for these, the agents would undertake them.

13.7. A more general model

The particular formalization of the problem that we have used so far in this chapter simplified the matter greatly, because the agents who do not buy information were easily identified and did not introduce biases in the answers. In the remainder of this chapter, we consider a more complex formulation of the information acquisition process, where these two properties do not hold.

As before, the agent, i, is represented as a Bayesian statistician who has prior beliefs concerning his own evaluation, summarized by a normal distribution $N(v, r^2)$. This information gathering process is formalized as the opportunity to draw an observation $\tilde{w}_i$ from a normal probability distribution $N(v_i, t^2(c_i))$, whose mean is the true value and whose variance t^2 decreases with the expense c_i incurred. The desirability of information is easily derived from Blackwell's theorem.

The behavior of the agent follows a two-step procedure: first, he defines his purchase of information; with this acquired knowledge, he then gives his optimal answer concerning his own evaluation. Using the method of dynamic programming, the solution is obtained by solving the problem in reverse. For each outcome of the first step, that is, for each value of $\tilde{w}_i$, the agent determines what his conditional optimal answer concerning his own evaluation will be. This first step defines, therefore, an indirect utility function in $\tilde{w}_i$. In the second step, the maximization of the expected value of this indirect utility function defines the purchase of information.

Let us now specify the procedure in greater detail. For a given observation $\tilde{w}_i$, the posterior distribution over his own evaluation v of the Bayesian agent is therefore:

$$N\left[\frac{t^2(c_i)v + r^2\tilde{w}_i}{t^2(c_i) + r^2}, \frac{t^2(c_i)r^2}{t^2(c_i) + r^2}\right] \tag{13.32}$$

As in section 3, the expected regret of saying w_i instead of the truth v_i is

$$R(v_i, w_i) = \int_{\substack{w_i < -x \leqslant v_i \\ -w_i \leqslant x < -v_i}} |v_i + x|\, p''(x)\,\mathrm{d}x \tag{13.33}$$

where $p''(\cdot)$ is the agent's prior probability density over the sum of the others' answers.

For each value of $\tilde{w}_i$, his optimal response is then determined by the minimization of expected regret

$$\min_{w_i} \int_{-\infty}^{+\infty} R(v_i, w_i) g_{\tilde{w}_i}(v_i)\,\mathrm{d}v_i \tag{13.34}$$

where $g_{\tilde{w}_i}(\cdot)$ is the density of (13.32).

Let $\psi(\tilde{w}_i)$ be the optimal response; for each value of $\tilde{w}_i$ and v_i the expected regret is then $R(v_i, \psi(\tilde{w}_i))$.

In the second step of the procedure the agent determines the optimal purchase of information. The joint distribution of v_i and $\tilde{w}_i$ for a given level of information expenditure c_i is:

$$f_{c_i}(v_i, \tilde{w}_i) = \frac{1}{2\pi r t(c_i)} \mathrm{e}^{-\frac{1}{2}\left[\frac{(v_i - v)^2}{r^2} + \frac{(\tilde{w}_i - v_i)^2}{t^2(c_i)}\right]} \tag{13.35}$$

because v_i is distributed $N(v, r^2)$ and $\tilde{w}_i$ is distributed $N(v_i, t^2(c_i))$, given v_i.

The optimization program of the agent is then:

$$\min_{c_i \geqslant 0} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} R(v_i, \psi(\tilde{w}_i)) f_{c_i}(v_i, \tilde{w}_i)\,\mathrm{d}v_i\,\mathrm{d}\tilde{w}_i + c_i \tag{13.36}$$

The solution to program (13.36) determines the optimal purchase of information.

13.8. Optimal purchase of information

In order to study the comparative static properties of the optimal response $\psi(\tilde{w}_i)$ and of the optimal purchase of information, we choose the following specification[1] of the agents' perceptions $p''(\cdot)$.

$$\begin{aligned} p''(x) &= \frac{1}{2(n-1)A} \qquad -(n-1)A \leqslant x \leqslant (n-1)A \\ &= 0 \qquad \text{otherwise} \end{aligned} \tag{13.37}$$

Theorem 13.4. Under (13.37), the optimal response is the mean of the posterior distribution, $g_{\tilde{w}_i}(v_i)$; that is,

$$\frac{t^2(c_i)v + r^2\tilde{w}_i}{t^2(c_i) + r^2}$$

Proof. If $w_i \leqslant v_i$

$$\begin{aligned} R(v_i, w_i) &= \frac{1}{2(n-1)A} \int_{-v_i}^{-w_i} (v_i + x)\,\mathrm{d}x \\ &= \frac{1}{4(n-1)A}(v_i - w_i)^2 \end{aligned} \tag{13.38}$$

Equation (13.38) can also be seen to hold when $v_i < w_i$.

Therefore, the optimal response is determined by

$$\min_{w_i} \int_{-\infty}^{\infty} \frac{1}{4(n-1)A}(v_i - w_i)^2 \frac{1}{\sqrt{(2\pi)}\sigma^*} \mathrm{e}^{-\frac{1}{2\sigma^{*2}}\left[v_i - \frac{t^2(c_i)v + r^2\tilde{w}_i}{t^2(c_i) + r^2}\right]}\,\mathrm{d}v_i \tag{13.39}$$

where

$$\sigma^{*2} = \frac{t^2(c_i)r^2}{t^2(c_i) + r^2}$$

Since the loss function is quadratic, the optimal response can easily be shown to be as in the statement of the theorem. Q.E.D.

[1] This is the only case in which we have been able to obtain analytic results in this more general system where the non-purchasers of information cannot be immediately detected. It should be observed that this p'' cannot be obtained as the sum of $n - 1$ independently identically distributed random variables for all n. Some dependence is, therefore, being implicitly assumed here.

The optimal response takes this form for some other distributions, but not generally. Notably, in the case in which each agent's beliefs about the others is normally distributed around his own true evaluation, the optimal answer is between zero and the posterior mean. In this case, it can in fact be shown that the optimal answer approaches zero for all agents as the number of others tends toward infinity.

For the system with expectations given by (13.37), however, the following attractive property can be proven.

Theorem 13.5. Under (13.37), the optimal purchase of information is a decreasing function of the sample size.

Proof. The optimal expenditure on information purchased is determined by the minimization problem:

$$\min_{c_i \geq 0} c_i + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{4(n-1)A}\left(v_i - \frac{t^2(c_i)v + r^2\tilde{w}_i}{t^2(c_i)+r^2}\right)^2 \frac{1}{2\pi r t(c_i)} e^{-\frac{1}{2}\left[\left(\frac{v_i - v}{r^2}\right)^2 + \frac{(\tilde{w}_i - v_i)^2}{t^2(c_i)}\right]} dv_i d\tilde{w}_i \tag{13.40}$$

Integrating, this can be shown to be equal to

$$c_i + \frac{1}{4(n-1)A} \cdot \frac{r^2 t^2(c_i)}{t^2(c_i)+r^2} \tag{13.41}$$

Therefore, the minimum is defined by

$$1 + \frac{r^2 t(c_i)}{(n-1)A}\frac{dt}{dc_i} = 0 \tag{13.42}$$

Differentiating totally with respect to c_i and n, we have

$$0 = \left[-\frac{-t(c_i)}{(n-1)^2}\frac{dt}{dc_i}\right]dn + \frac{1}{n-1}\Delta dc_i \tag{13.43}$$

where Δ is proportional to the second derivative of (13.41) with respect to c_i and is therefore positive by the second-order condition for an optimal c_i.

Therefore,

$$\frac{dc_i^*}{dn} = \frac{t(c_i)\frac{dt}{dc_i}}{(n-1)\Delta} < 0 \tag{13.44}$$

since $dt/dc_i < 0$, by assumption. Q.E.D.

It is also true that $c_i^*(n)$ approaches zero for large n, which can be seen from the monotonicity of t in c_i and (13.44), (13.42).

Appropriate differentiations give also

$$\frac{dc_i^*}{dA} = \frac{t\frac{dt}{dc_i}}{2A\Delta} < 0 \tag{13.45}$$

$$\frac{dc_i^*}{dr^2} > 0 \tag{13.46}$$

The more uncertain the agent is about his own evaluation (i.e., the larger is r^2), the more information he buys. Also, the more spread out his beliefs about the other people's evaluations are (the larger is A), the less information he buys (since the larger A is the less likely he is to be pivotal).

Consider in (13.42), n as a function of c_i. Integrating, we obtain

$$n - 1 = \frac{n^4}{4Ac_i}\log\left[\frac{t^2(0)+r^2}{t^2(c_i)+r^2}\right]. \tag{13.47}$$

What we would like is the inverse function, that is, c_i as a function of n. Let us take an example in which $c_i^*(n)$ can be computed exactly.

Suppose that $t(c_i)$ is given by

$$t^2(c_i) = Ke^{-\sqrt{c_i}} - r^2. \tag{13.48}$$

We then can obtain

$$c_i^*(n) = K^1(n-1)^{-1/2}, \tag{13.49}$$

where K^1 is a positive constant independent of n.

13.9. Optimal sample sizes

The decision maker determines the sample size which maximizes the pre-sample expected evaluation of the project based on his prior information. He is assumed to be a Bayesian statistician who approximates the distribution of evaluations in the population of size N with a normal distribution of unknown mean m and known variance s^2. He has a prior $N(\mu, \sigma^2)$ on the unknown parameter m. In addition, he believes that the means v_i of the agents' prior beliefs are randomly distributed around the true evaluations with variance p^2, when in fact they have a systematic bias $b > 0$. Consequently, he believes that the answers of the sampled agents are distributed around the true evaluations with a normal distribution of variance $(t^4(c_i)p^2 + r^4t^2(c_i))/(t^2(c_i) + r^2)^2$, which is considered as a constant t^2, because all agents are alike in this model and differ only through their information acquisition experience.[2]

A random sample of size n is drawn from the population; let $\overline{w}$ be the mean of the answers given by the agents in this sample.

The posterior mean of the *sample* is then:

$$\frac{t^2\mu + n\sigma^2\overline{w}}{t^2 + n\sigma^2} \tag{13.50}$$

The posterior mean of the *entire population* is:

$$\frac{(t^2 + s^2)\mu + n\sigma^2\overline{w}}{(t^2 + s^2) + n\sigma^2} \tag{13.51}$$

Therefore, for the decision maker who is assumed to be risk neutral, the *post sample* evaluation of the project is:

$$(N - n)\frac{(t^2 + s^2)\mu + n\sigma^2\overline{w}}{t^2 + s^2 + n\sigma^2} + n\frac{t^2\mu + n\sigma^2\overline{w}}{t^2 + n\sigma^2} \tag{13.52}$$

Let $f(\overline{w})$ be the ex ante distribution attributed by the center to $\overline{w}$; it is a normal distribution with mean μ and variance $[(s^2 + t^2)/n] + \sigma^2$.

Therefore, the presample expected evaluation of the project is:

$$V(N, n) = \int_{\overline{w} \gtrless 0} \left[(N - n)\frac{(t^2 + s^2)\mu + n\sigma^2\overline{w}}{t^2 + s^2 + n\sigma^2} + n\frac{t^2\mu + n\sigma^2\overline{w}}{t^2 + n\sigma^2}\right] f(\overline{w})\,\mathrm{d}\overline{w} - c(n)n$$

or

$$\frac{V(N, n)}{N} = \frac{\sigma^2 + k(n)/N}{\sqrt{(2\pi)}\cdot\left(\dfrac{s^2 + t^2}{n} + \sigma^2\right)^{1/2}} \mathrm{e}^{-\frac{1}{2}\frac{\mu^2}{\left(\frac{s^2+t^2}{n}+\sigma^2\right)}} + \mu\rho\left(\frac{\mu}{\left(\dfrac{s^2 + t^2}{n} + \sigma^2\right)^{1/2}}\right) - \frac{nc(n)}{N} \tag{13.53}$$

where $\lim k(n) = \overline{k}$ as n goes to infinity.

[2] The decision maker assumes that the agent answers with the posterior mean.

The optimal sample size $n^*(N)$ is obtained from the maximization of the presample expected evaluation of the project $V(N, n)$ computed on the basis of expectations distorted by a bias.

Let $\tilde{V}(N, n)$ denote the expected value of the project computed as if the decision maker knew the right distribution of $\overline{w}$, i.e., a normal distribution with mean $(\mu + b)$ and variance $[(s^2 + t^2)/n] + \sigma^2$. Then,

$$\frac{\tilde{V}(N, n)}{N} = \frac{\sigma^2 + k(n)/N}{\sqrt{(2\pi)}\cdot\left(\dfrac{s^2 + t^2}{n} + \sigma^2\right)^{1/2}} \mathrm{e}^{-\frac{1}{2}\frac{(\mu+b)^2}{\left(\frac{s^2+t^2}{n}+\sigma^2\right)}} + \mu\rho\left(\frac{\mu + b}{\left(\dfrac{s^2 + t^2}{n} + \sigma^2\right)^{1/2}}\right) - \frac{nc(n)}{N} \tag{13.54}$$

To evaluate the expected gain of the procedure, we use the expected value of the project computed with correct expectations, i.e.,

$$\tilde{V}(N, n^*(N)) - \max[\mu,0] \tag{13.55}$$

From section 8, we know that $c(n)$ goes to zero as n goes to infinity; let us say that $c(n)$ is of the order of $1/n^\lambda$. Calculations analogous to those of chapter 11 yield an "optimal" sample size $n^*(N)$ of the order of $N^{1/(2-\lambda)}$ if λ is less than one or of the order of N otherwise. On this basis, we can therefore evaluate the asymptotic gain of the procedure. We obtain:

$$\begin{aligned} &\mu\left[\rho\left(\frac{\mu + b}{\sigma}\right) - 1\right] + \frac{\sigma}{\sqrt{(2\pi)}}\mathrm{e}^{-\frac{1}{2}\left(\frac{\mu+b}{\sigma}\right)^2} && \text{if } \mu > 0 \\ &\mu\rho\left(\frac{\mu + b}{\sigma}\right) + \frac{\sigma}{\sqrt{(2\pi)}}\mathrm{e}^{-\frac{1}{2}\left(\frac{\mu+b}{\sigma}\right)^2} && \text{if } \mu < 0 \end{aligned} \tag{13.56}$$

Now, from figure 13.2 we see that for negative values of the mean μ, there is a range of values of σ, corresponding to a not too large uncertainty, for which the "true" expected gain of the procedure is negative, because the sample size was computed with wrong expectations.

In other words, when the prior information of agents on their evaluation of public goods is biased, and when the uncertainty of the decision maker is not too large with respect to this bias, it may be better for the society if the decision maker goes ahead with his prior beliefs than to try to extract information from agents with a costly mechanism which has such weak incentives that the answers will be of little value.

Of course, in any given problem, it is difficult to know in advance in which region of figure 13.2 we are.

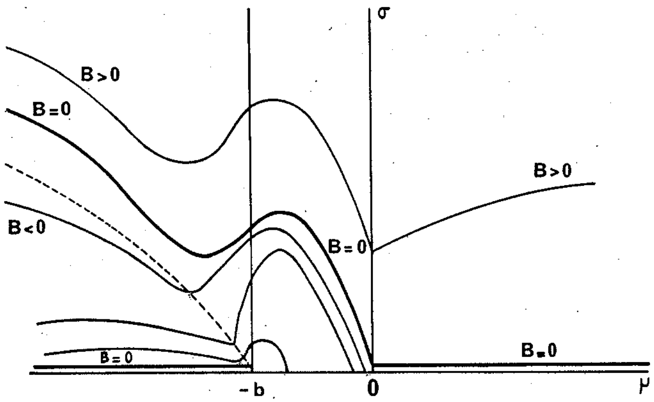

Figure 13.2.

This result shows, however, that it is important to question the value of decentralized information before using democratic decision processes.

One might be tempted to conclude with Hume [1888] who suspected the failure of decentralized decision mechanisms:

> Political society easily remedies both these inconveniences.[3] Magistrates find an immediate interest in the interest of any considerable part of their subjects. They need consult nobody but themselves to form any scheme for the promoting of that interest.

But, how do magistrates get an unbiased idea of the interest of their subjects? For clearly valuable or not valuable public projects (σ small with respect to μ), this task appears simple enough, insofar as the magistrate can through his relationships with the population extract this information. For more delicate decisions, he himself may well face a free-rider problem and a decentralized truthful mechanism would appear necessary.

[3] That is free-rider behavior and organizational problems linked to large number societies.

PART IV

Chapter 14

DYNAMIC APPROACHES

14.1. Procedures

Parts II and III of the book have concentrated on "global" mechanisms, that is, mechanisms which select the social state in a single play, in which the strategy spaces of the agents were the spaces of allowable preference relations or even larger spaces for non-revelation mechanisms. Under the assumption of separability, we found classes of mechanisms which were quite satisfactory on incentives grounds; the truth was a dominant strategy for Groves mechanisms, a maximin strategy for the Dubins mechanism, or an expected utility maximizing strategy for the D'Aspremont–Gérard–Varet mechanism. In each case an optimal outcome would result.

For the dominant strategy mechanisms, it was shown (see chapters 5 and 13) that the separability assumption was necessary if Pareto optimality and strategic non-manipulability were to be maintained. We showed in chapter 5 that, by weakening the requirement of Pareto optimality slightly, we could relax the separability assumption to some extent. In this part we will show that it is possible to replace separability by a limitation on the type of strategic manipulation considered.

We consider mechanisms for environments with non separable utility functions which are not revelation mechanisms in the ordinary sense. We will first restrict attention to models with one public good and one private good, and to utility functions that are differentiable. Agents are assumed to play the game defined by the mechanism by revealing their marginal rate of substitution at each point in time, as the process evolves through alternative patterns of resource allocation. Formally, a strategy for each agent is a function which gives, for each date t, the marginal rate of substitution he would announce conditional on the evolution of the process up to that date. Generally speaking, this is a highly complex entity. It can be greatly simplified by the assumption of myopic behavior. The strategy played at t is independent of the history of the process.

Myopic behavior implies that each agent states a marginal rate of substitution at each instant that is optimal under the assumption that the process is just about to terminate. More formally, since the state of the system is known to him at that date, it is assumed that he optimizes the rate of change of his utility at the current data.

One could use this type of formalization either in discrete or continuous time models. We chose the later because of its greater technical simplicity and elegance.

A dynamic mechanism as described above will be called a *procedure*.

In the remainder of this chapter, we will describe more precisely the two commodity model to be used. Then we discuss the Bowen procedure and the naïve procedure to point out some difficulties. Most of what will be said in this part can be generalized to several public goods. We mention the cases where such extensions are not possible, which arise mainly in studying the dynamic stability of the procedures.

14.2. The model

We consider an economy with N agents (sometimes a continuum $[0, 1]$ of agents will be used), one private good and one public good, both being infinitely divisible.

Let $u_i(K, x_i)$ be the utility function of the ith consumer as a function of his consumption of public good and private good respectively. $u_i(\cdot,\cdot)$ is assumed to be strictly quasi-concave and twice continuously differentiable with:

$$u_{ix_i}(K, x_i) = \frac{\partial u_i}{\partial x_i}(K, x_i) > 0$$

$$u_{iK}(K, x_i) = \frac{\partial u_i}{\partial K}(K, x_i) \geqslant 0, \quad \frac{\partial u_i}{\partial K}(0, x_i) > 0 \tag{14.1}$$

We assume that the utility function $u_i(\cdot,\cdot)$ is defined for any value x_i in $\mathbf{R}$, and we ignore, for the moment, possibilities of bankruptcy. (Below we will discuss several alternative ways of treating bankruptcy). The marginal rate of substitution between public good and private good is denoted:

$$q_i(K, x_i) = \frac{u_{iK}(K, x_i)}{u_{ix_i}(K, x_i)}. \tag{14.2}$$

Let $\overline{x}_i > 0$ be the initial endowment of agent i in private good.

The production possibilities are described by a twice continuously differentiable increasing function $\Gamma(K)$ which specifies the amount of private good necessary to produce any given amount K of public good. The marginal cost of producing an additional unit of public good is denoted $\gamma(K)$ with

$$\gamma(K) > 0 \text{ and } \gamma'(K) > 0 \quad \text{if } K > 0;$$

$$\gamma(0) = 0, \qquad \gamma'(0) = 0.$$

Theorem 14.1. An allocation $(K, x_1, \ldots, x_N)$ is a Pareto optimum iff

$$\sum_i q_i(K, x_i) = \gamma(K) \tag{14.3}$$

$$\Gamma(K) + \sum_i x_i = \sum_i \overline{x}_i \tag{14.4}$$

The proof is immediate in view of our assumptions which avoid a corner solution.

Condition (14.3) is the famous Bowen–Lindahl–Samuelson condition and (14.4) is the feasibility constraint.

14.3. Bowen's procedure

We first describe the dynamic version of Bowen's voting mechanism, referred to as the Bowen procedure (see Bowen [1943]).

Starting at $t = 0$ with a zero level of public good production, there is, at each instant t, a vote on the opportunity to increase or decrease the production of the public good. A uniform imputation of the cost is announced in advance, and the decision to increase (or decrease) the quantity of public good is taken if a majority of agents vote for an increase (or decrease).

For any agent i and any instant $t \geqslant 0$, let $s(i, t)$ be the vote of agent i at instant t.

$$\begin{aligned} s(i,t) &= +1 \text{ if he votes for an increase} \\ &= 0 \text{ if he votes for a status quo} \\ &= -1 \text{ if he votes for a decrease.} \end{aligned} \tag{14.5}$$

The Bowen procedure can then be formalized as follows:

$$\begin{aligned} \dot{K}(t) &= +1 \quad \text{if } \textstyle\sum_i s(i,t) > 0 \\ &= 0 \qquad\qquad\quad = 0 \\ &= -1 \qquad\qquad\quad < 0 \end{aligned} \tag{14.6}$$

$$\dot{x}_i(t) = -\frac{\gamma(t)}{N}\cdot \dot{K}(t) \qquad i = 1, \ldots, N. \tag{14.7}$$

The relations (14.6) describe the evolution of the produced quantity of public good according to the majority rule; (14.7) describes the imputation of the cost of the public good to the agents, chosen to be uniform by assumption.

Assumption 1. The distribution of marginal rates of substitution net of imputed costs, $q_i(K, x_i) - (1/N)\gamma(K)$, is such that at each instant the median equals the mean.

A vote by agent i at instant t is said to be *truthful* iff

$$\begin{aligned} s(i,t) &= 1 \quad \text{if } q_i(K(t), x_i(t)) > \frac{\gamma(t)}{N} \\ &= 0 \qquad\quad q_i(K(t), x_i(t)) = \frac{\gamma(t)}{N} \\ &= -1 \qquad\quad q_i(K(t), x_i(t)) < \frac{\gamma(t)}{N} \end{aligned} \tag{14.8}$$

An agent is said to have a myopic behavior when he defines his vote at date t by considering only the increase of utility he can obtain at date t, namely $\dot{u}_i = (\partial u_i/\partial K)\dot{K} + (\partial u_i/\partial x_i)\dot{x}_i$.

Theorem 14.2. Truthful voting is a dominant strategy for each agent at each instant (even if he is not myopic).

The proof is obvious since he cannot influence his share of the cost, and any untruthful vote can only hurt him by stopping the procedure too soon.

Theorem 14.3. Under assumption 1, for every t

$$\begin{aligned} \textstyle\sum_i s(i,t) &> 0 \quad \text{iff } \textstyle\sum_i q_i(K(t), x_i(t)) > \gamma(t) \\ &= \qquad\qquad\qquad\qquad\qquad = \\ &< \qquad\qquad\qquad\qquad\qquad < \end{aligned} \tag{14.9}$$

The result follows directly from theorem 14.2 and assumption 1.

Assumption 2. There exists a level of public good production $\overline{\overline{K}}$ such that for any $(x_1, \ldots, x_N)$, $\sum_i q_i(\overline{\overline{K}}, x_i) - \gamma(\overline{\overline{K}}) < 0$.

Theorem 14.4. Under assumptions 1 and 2, the Bowen procedure converges to a Pareto optimum.

Proof. At a stationary point t^* of the procedure

$$\sum_i q_i(K(t^*), x_i(t^*)) = \gamma(t^*) \qquad \text{since } \dot{K}(t^*) = 0 \tag{14.10}$$

Moreover, $\sum_i \dot{x}_i(t) + \gamma(t) = 0$ at each instant t implies $\Gamma(K(t^*)) + \sum_i x_i(t^*) = \sum_i \overline{x}_i$. Therefore, from section 2, a stationary point of the procedure is a Pareto optimum.

At date 0, by the assumptions of section 2, we have $\dot{K}(0) > 0$. Thus, $\dot{K}(t)$ cannot become negative without having the procedure stop beforehand, since, from the continuity of $\sum_i q_i(K(t), x_i(t)) - \gamma(t)$, $\dot{K}(t)$ should become zero by the intermediate value theorem.

Therefore $K(t)$ is increasing and bounded by $\overline{\overline{K}}$ from assumption 2. Consequently, $K(t)$ converges to a stationary value K^*. Let $(x_1^*, \ldots, x_N^*)$ be the associated distribution of private good, then $(K^*, x_1^*, \ldots, x_N^*)$ is a Pareto optimal allocation. Q.E.D.

Assumption 2 is used to bound the optimal level of public good, while allowing unbounded negative levels of private good consumption. Of course, this boundedness of public good production could, in a more classic way, be obtained from the feasibility constraint (14.4) by imposing positive private good consumption.

In that case, however, the procedure may not be well defined if the trajectory hits the constraint $x_i(t) \geqslant 0$, for an agent i, as the imputation of the cost which is fixed ex ante may cause him to become bankrupt. Either one assumes that the procedure stops at such points at the risk of losing Pareto optimality, or one redefines the shares of the cost so that it is borne by solvent agents only. The latter possibility will destroy truthful voting unless myopic behavior is assumed.

Another point of view is that the center can observe initial endowments, which are more likely to be observable than preferences, and, knowing an upper limit to the production of public good, defines a cost sharing scheme which guarantees that no agent can become bankrupt. This will be possible in general if the upper limit $\overline{\overline{K}}$ is not too large. Assumption 1 must then be interpreted to apply to the modified net willingnesses to pay. This is clearly the more realistic point of view even though it is theoretically clumsy.

Were it not for the necessity of assumption 1, the results of the Bowen procedure would be remarkably good. What is the best interpretation to be given to assumption 1? Bowen clearly had in mind a large population, considered as a sample from a distribution (chosen to be normal in his paper) for which the median of the net willingnesses to pay was equal to the mean. Of course, this property would not be strictly true in the sample obtained, except by coincidence.

However, we can easily prove a continuity theorem of the following kind. If the distribution of net willingnesses to pay converges to a distribution for which the median and mean are the same, then the outcome of the Bowen procedure converges to a Pareto optimum. The Bowen procedure is then quite satisfactory for approximately "symmetric" distributions of willingnesses to pay[1].

The procedure stops for a median of net willingnesses to pay equal to zero. Then, $\sum_i q_i(K, x_i) - \gamma(K)$ is close to ε; for example let us take $\phi(K, x) = \sum_i q_i(K, x_i) - \gamma(K) > 0$. We want to find conditions such that $\phi(K, x)$ can be decreased to restore Pareto efficiency through a feasible reallocation

$$\mathrm{d}\phi(K, x) = \left(\sum_i \frac{\partial q_i}{\partial K}\right)\mathrm{d}K + \sum_i \frac{\partial q_i}{\partial x_i}\mathrm{d}x_i - \gamma'(K)\mathrm{d}K \qquad (14.11)$$

From feasibility $\gamma(K)\mathrm{d}K + \sum_i \mathrm{d}x_i = 0$.

Hence,

$$\mathrm{d}\phi(K, x) = \sum_i \left[\frac{\partial q_i}{\partial x_i} + \frac{\gamma'(K)}{\gamma(K)} - \frac{1}{\gamma(K)}\sum_j \frac{\partial q_j}{\partial K}\right]\mathrm{d}x_i$$

A sufficient condition to reach our result is therefore that there exists an i such that

$$\frac{\partial q_i}{\partial x_i} + \frac{\gamma'(K)}{\gamma(K)} - \frac{1}{\gamma(K)}\sum_j \frac{\partial q_j}{\partial K} \qquad (14.12)$$

be uniformly bounded below by a positive number δ.

[1] Note that this assumption of approximately symmetric distributions must hold along the path so that it appears as a reasonable assumption only in the separable case, for which the net willingnesses to pay are independent of the transfers.

For example, if the utility function of agent i is separable and concave, we have:

$$\frac{\partial q_i}{\partial x_i} = 0, \quad \frac{\partial q_i}{\partial K} \leqslant 0 \qquad (14.13)$$

the condition (14.12) is fulfilled if either $\gamma'(K)$ is bounded below by a positive number (a positive slope of the marginal cost function), or there is an agent j such that $\partial q_j/\partial K$ is bounded above by a negative number (strictly positive curvature of the valuation function).

With several public goods, the same results hold except that stability may fail to obtain if the utility functions are not separable (see the discussion in chapter 16, section 3).

14.4. The naïve procedure

The large number assumption is often formalized in economic theory by a continuum of agents, for example [0, 1]; each agent is really negligible; competitive behavior is then studied in such models in a rather elegant way. However, the validity of approximating a large number economy by a continuum needs proof, and indeed, such proofs for the study of the core (Hildenbrand [1974]) or for the study of incentives in private good economies (Roberts and Postlewaite [1973]) have been provided through appropriate continuity results. In this section, we would like to draw attention to a lack of continuity of the incentives properties with respect to the size of the population.

Consider the following naïve procedure for the continuum [0, 1] with the Lebesgue measure. Let $\psi_i(t)$ be the stated marginal rate of substitution of agent i, $i = 1, \ldots, N$.

$$\dot{K}(t) = \int_0^1 [\psi_i(t) - \delta(i)\gamma(t)]\,\mathrm{d}i \qquad (14.14)$$

$$\dot{x}_i(t) = -\delta(i)\gamma(t) \text{ for all } i, \text{ with } \int_0^1 \delta(i)\mathrm{d}i = 1. \qquad (14.15)$$

Equation (14.15) describes an ex-ante imputation of the cost and (14.14) describes the evolution of the planned public good level proportional to the excess of the aggregate willingness to pay over the marginal cost.

Clearly, in this procedure, there is no incentive to misrepresent preferences since there is no opportunity to influence the cost share defined ex ante and no possibility of changing the decision because each agent is of measure zero. Consequently, there is, strictly speaking, no incentive to lie, even though there is no incentive to tell the truth either. Such a situation may appear good enough if we endow the agents of the model with a sense of solidarity and group awareness. On the other hand it is clearly not enough if preference information is personally uncertain as in chapter 13.

Is this asymptotic model a valid representation of a large number economy? The large number analog of the naïve procedure is:

$$\dot{K}(t) = \sum_i [\psi_i(t) - \delta(i)\gamma(t)] \qquad (14.16)$$

$$\dot{x}_i(t) = -\delta(i)\gamma(t) \text{ for all } i, \text{ with } \sum_i \delta(i) = 1. \qquad (14.17)$$

It is still true that the agent has no incentive to lie to change his cost share since it is defined ex ante. However if his net willingness to pay is positive, $q_i - \delta(i)\gamma > 0$, his best interest is to overstate his willingness to pay *as much as possible* and similarly if $q_i - \delta(i)\gamma < 0$, he should understate his willingness to pay as much as possible. Indeed, the gain in utility of agent i at instant t is:

$$\begin{aligned} \dot{u}_i &= \frac{\partial u_i}{\partial x_i}(q_i\dot{K} + \dot{x}_i) \\ &= \frac{\partial u_i}{\partial x_i}[q_i - \delta(i)\gamma]\,\dot{K} \qquad (14.18) \end{aligned}$$

By overstating his marginal rate of substitution at instant t when $q_i - \delta(i)\gamma$, he increases $\dot{K}(t)$ and therefore $\dot{u}_i$.

Thus, for any N, as large as desired, there is an incentive to lie enormously even though for the continuum model there is no incentive to lie.

14.5. Conclusions

A number of other procedures have been suggested (see Tulkens [1976] for a survey of the literature) which use personalized prices for public goods revised during the course of the procedure on the basis of the demands expressed at these prices. These procedures require very strong assumptions to insure stability. An exception to this is a procedure given by Malinvaud [1971] that formalizes Lindahl mechanisms in economies where lump sum transfers are possible. None of these procedures has been studied from the point of view of incentives, and even though one may be

justifiably skeptical about the performance of these procedures on this ground, further work is needed in this direction.

A second class of procedures which directly specifies a dynamics in quantities has been the topic of a considerable amount of work in the last few years. We turn to it in the next chapter.

Chapter 15

THE MDP PROCEDURE

15.1. Definition of the MDP procedure

Drèze and de la Vallée Poussin [1971] and Malinvaud [1972a] have constructed a planning procedure for economies with public goods which has come to be called the MDP process. The essential idea of the MDP procedure can also be found in Tideman [1972] who suggested to follow a two stage process. In the first the Bowen procedure as described in chapter 14 is used to approximate the optimum. Then a single iteration of a discrete version of the MDP procedure is employed to sharpen the approximation.

The continuous version of the MDP procedure is described by the following system of differential equations:

$$\dot{K} = \sum_i q_i(K, x_i) - \gamma(t) \tag{15.1}$$

$$\dot{x}_i = -q_i(K, x_i)\dot{K} + \lambda_i\Big(\sum_i q_i(K, x_i) - \gamma(t)\Big)\dot{K}$$

$$\sum_i \lambda_i = 1, \ \lambda_i \geqslant 0 \qquad i = 1, \ldots, N \tag{15.2}$$

The speed of adjustment of the quantity of public good is defined by the difference between the sum of the willingnesses to pay of the agents and the marginal cost of public good production.

Agent i is taxed (if $\dot{K} > 0$) or subsidized (if $\dot{K} < 0$) by a first amount $-q_i \cdot \dot{K}$ aimed at keeping him on the same indifference curve. This quantity would exactly compensate him for the change in the level of public good given his announced preferences. If $\dot{K} > 0$, the amount of taxes levied in this operation $\sum_i q_i \dot{K}$ is larger than the cost of producing the marginal unit by virtue of the definition of $\dot{K}$, so that a surplus in private good is made available at each instant t namely $[\sum_i q_i(K, x_i) - \gamma(t)]\dot{K}$. The MDP procedure divides this surplus according to a sharing rule defined by the coefficients λ_i. Similarly when $\dot{K} < 0$, the amount of compensating payments $-\sum_i q_i \dot{K}$ is less than the saving arising through decreasing the production of public good so that a surplus has also to be distributed in this case as well.

Consequently, it is obvious that the MDP procedure is *individually rational* at each instant, and therefore globally. The fact that the utility of each agent is monotone non-decreasing throughout the process, whenever he reveals his correct preferences is the essential feature of this procedure that is responsible for all of its desirable properties. We will return to it repeatedly in this chapter.

15.2. Stability

The starting point of the procedure is the allocation defined by the endowments in private good and a zero level of public good; any other feasible allocation could also be used.

Under correct revelation of preferences, the first byproduct of individual rationality is stability, a simple proof of which is given below under the assumptions of section 14.2.

Theorem 15.1. Under assumption 2, the MDP procedure converges to a Pareto optimum.

Proof. Consider the sum, Σ, of the utility functions

$$\Sigma = \sum_i u_i(K, x_i) \tag{15.3}$$

$$\dot{\Sigma} = \sum_i \dot{u}_i(K, x_i) = \sum_i \frac{\partial u_i}{\partial x_i}[q_i \dot{K} + \dot{x}_i]$$

$$= \sum_i \frac{\partial u_i}{\partial x_i} \lambda_i (\sum_i q_i(K, x_i) - \gamma(t))^2 \tag{15.4}$$

Therefore, from (15.4) we have

$\dot{\Sigma}(t) \geqslant 0$ and from (15.1, 15.2) we have:

$\dot{\Sigma}(t) = 0$ implies $\dot{K} = 0$, and also $\dot{x}_i = 0$, for all i.

From assumption 2, and the feasibility constraint, K and x_i, $i = 1, \ldots, N$ are bounded above so that $\Sigma(t)$ is also bounded above. $\Sigma(t)$ is thus an appropriate choice for a Lyapunov function.

The MDP procedure is therefore quasi-stable in Uzawa's sense [1961] that is, any limit point of the trajectory is a stationary point. It is then immediate to obtain stability by using the strict quasi-concavity of the utility functions and the fact that any stationary point is a Pareto optimum, and hence the process converges to such a point. Q.E.D.

In the case of a single public good, a proof using the quantity of public good as a Lyapunov function as in theorem 14.1, could have been given; however, the above proof generalizes immediately to several public goods.

For simplicity of exposition we have neglected the possibility of boundary problems, at $K = 0$. In the case of a single public good, it is not a problem since K increases monotonically throughout the convergence process but for several public goods it is a serious difficulty which has been solved by Henry [1972].

This is done by reformulating (15.1) as:

$$\dot{K}_j = \sum_i q_{ij}(K, x_i) - \gamma_j(t) \qquad \text{if } K_j > 0$$

$$= \max(0, \sum_i q_{ij}(K, x_i) - \gamma_j(t)) \qquad \text{if } K_j = 0 \tag{15.5}$$

where the index j denotes the jth public good $j = 1, \ldots, J$.

In the system of equations (15.5) there is a discontinuity in the right hand side. Henry has shown how the existence of solutions to this non-classical differential system could be obtained and Champsaur [1976] has provided a general proof of stability in that case.

Without assumption 2, boundary problems due to $x_i \geqslant 0$, are easily avoided with:

Assumption 3. $u_i(K, 0) < u_i(0, x_i)$, for any $x_i \in \mathbf{R}_+$, for any $K \in \mathbf{R}_+$.

Then, when the private good allocation of agent i decreases, his evaluation of the public good also decreases so that the compensating payments he might have to make can never bankrupt him.

15.3. Maximin property

Drèze and de la Vallée Poussin [1972] have studied the MDP procedure from the point of view of incentives. They observed that the truthful strategy is the maximin strategy. Any misrepresentation may result in a decrease of utility if the other agents' strategies are in a particular configuration.

Consider an agent at instant t, and let us assume that he announces a marginal rate of substitution $\psi_i(t)$ different from the true one $q_i(t)$. His instantaneous change in utility is:

$$\dot{u}_i = \frac{\partial u_i}{\partial x_i}[q_i(t) - \psi_i(t) + \lambda_i[\sum_j \psi_j(t) - \gamma(t)]]\dot{K} \tag{15.6}$$

Suppose that $q_i(t) = \psi_i(t) + \varepsilon, \varepsilon > 0$. Then, if $\sum_{j \neq i}\psi_j(t) = \gamma(t) - \psi_i(t) - \varepsilon/2$, we will have $\dot{K} < 0$. Hence $\dot{u}_i = (\partial u_i/\partial x_i)(1 - \lambda_i/2)\,\varepsilon\,\dot{K} < 0$. A similar result follows for $\varepsilon < 0$.

Therefore, for any misrepresentation of his marginal rate of substitution, the answers of the other agents may be such that this agent's utility function decreases.

Since in a continuous time procedure, misrepresentation at an instant of time can only lead to an infinitesimal decrease in utility, one might not think that the incentives to follow such a maximin behavior are compelling, even for extremely risk averse agents. The real risk that an agent faces, however, is that by distorting his preferences the process will stop prematurely. He thus is risking the loss of all the gain in utility that would have occurred to him during the remainder of the process that has been precluded from taking place. A counter argument to this risk is that it is extremely unlikely, indeed an event of measure zero given diffuse beliefs, that a distortion at a single point in time will cause termination. Even an extremely risk averse agent will not therefore act to avoid such a negligible risk of a bounded loss. However, if each agent's responses are restricted to be continuous functions of time, then a distortion at one instant necessitates distortions throughout a non-degenerate interval of time. In this way the risk that the process can stop during such an interval is assured to be a significant one that will be avoided by sufficiently risk averse agents. These considerations justify the assumption of the maximin behavior of revealing true marginal rates of substitution throughout the process.

Recall that in chapter 7 we studied a global mechanism, the Dubins mechanism, for which, when utility functions are separable, revelation of the true willingnesses to pay is a maximin strategy and a Pareto optimal outcome is achieved. The MDP process can be seen to be a dynamic version of a Dubins mechanism as follows:

Consider an instant t. There are three possible projects $\dot{K} = +1$, $\dot{K} = 0$, $\dot{K} = -1$, that is the space of projects is $\mathscr{K} = \{-1, 0, +1\}$.

In a Dubins mechanism, the agent is assigned an ex ante defined cost share $\delta_i\gamma(t)\dot{K}$ and receives a transfer:

$$t_i(w) = -w_i(\bar{K}^*) + \int_{\mathscr{K}} w_i(K)\mathrm{d}\pi(K) + \lambda_i\sum_j w_j(\bar{K}^*)$$
$$-\lambda_i \int_{\mathscr{K}} \sum_j w_j(K)\mathrm{d}\pi(K) \qquad \text{with } \sum_i \lambda_i = 1 \tag{15.7}$$

where $w_i(K)$ is the net willingness to pay for project K, i.e. here at instant t, $(q_i - \delta_i\gamma)\dot{K}$.

$\bar{K}^*$ is the maximizing project which is here $+1, 0, -1$ according to the sign of $\sum_i q_i - \gamma$ and $\pi(K)$ is an arbitrary measure on the space $\mathscr{K}$ of projects that we choose here to be uniform.

Then:

$$\int_{\mathscr{K}} w_i(K)\mathrm{d}\pi(K) = \tfrac{1}{3}[w_i(+1) + w_i(0) + w_i(-1)]$$
$$= \tfrac{1}{3}[(q_i - \delta_i\gamma) - (q_i - \delta_i\gamma)] = 0. \tag{15.8}$$

The transfer reduces to:

$$t_i = -(q_i - \delta_i\gamma)\dot{K} + \lambda_i(\sum_j q_j - \gamma)\dot{K} \tag{15.9}$$

The net change in the private good endowment of agent i is therefore:

$$\dot{x}_i = -q_i\dot{K} + \lambda_i(\sum_j q_j - \gamma)\dot{K}$$

with

$$\begin{aligned} \dot{K} &= +1 && \text{if } \sum_i q_i - \gamma > 0 \\ &= 0 && \phantom{\text{if } \sum_i q_i - \gamma} = 0 \\ &= -1 && \phantom{\text{if } \sum_i q_i - \gamma} < 0 \end{aligned} \tag{15.10}$$

The procedure so obtained is almost the MDP procedure. In the latter the speed of adjustment of the public good level is determined by $\dot{K} = \sum_i q_i - \gamma$ instead of (15.10). However, it is clear that any sign preserving transformation of the speed of adjustment (15.10) has the same properties in terms of maximin behavior. Indeed such a transformation amounts to multiplying the whole utility function at instant t, $\dot{u}_i$, by a positive number and since truthful revelation is the only strategy securing a minimum value of zero for $\dot{u}_i$, this remains true for any speed of adjustment chosen. The MDP procedure is therefore a speed transform of the dynamic version of the Dubins mechanism.

Note that truthful revelation is also a maximin property of the MDP procedure, considered as a revelation mechanism. Each agent announces his entire utility function $u_i(K, x_i)$ and the outcome is given by the asymptotic solution of the MDP system of differential equations.[1]

The mechanism of Malinvaud and Drèze–de la Vallée Poussin appears then as an alternative to Dubins. Dubins obtains a successful mechanism under maximin behavior when the space of preferences on arbitrary project spaces is restricted to separable utility functions, but no further assumptions are made. They obtain a successful mechanism under maximin behavior when the utility functions on continuous project spaces are assumed to be twice continuously differentiable and strictly quasi-concave, under the additional assumptions made in section 14.2.

15.4. Neutrality

For any choice of $\lambda = (\lambda_1, \ldots, \lambda_N)$ such that $\sum_i \lambda_i = 1$ and $\lambda_i \geqslant 0$, $i = 1, \ldots, N$, the MDP procedure converges to an individually rational Pareto optimum. Let IRP denote the set of individually rational Pareto optima. The inverse question is of interest. Is it possible to reach any point in IRP by an appropriate choice of λ, that is, by an appropriate method for sharing of the surplus? A positive answer to this question has been given by Champsaur [1976], ensuring the "distributive neutrality" of the procedure.

In the case of the model used in this part with one public good, the only additional assumption needed is the strict concavity of the utility functions[2]. The interpretation of this result is somewhat delicate. It is not true that it is possible to choose ex ante λ in order to achieve a given Pareto optimum, since at the beginning of the procedure the center has no global information on utility functions. If on the other hand, λ is changed during the procedure on the basis of available information up to that point of time, the incentives for truthful answers at each instant under maximin behavior might be disturbed[3].

If the center knows the initial endowments, λ can be based on them; but then there will be however an incentive for misrepresentation of these data as well.

[1] However, the truth is not the only maximin property since all utility functions which have the same indifference curve (as the true utility function) going through the initial endowment are also maximin strategies.

[2] In the case of several public goods, Champsaur [1976] used an additional assumption of uniqueness of the trajectory (solution of the system of differential equations). This assumption has been later relaxed by Cornet [1976].

[3] Tulkens and Zamir [1976] have suggested to consider λ as determined as the outcome of a local bargaining game at each instant.

The redistributive capability of the MDP procedure, being limited to individually rational allocations, can provide an outcome satisfactory from the distributional point of view only in so far as initial endowments are equitably distributed. If λ could be chosen at the beginning of the procedure to reach any point in IRP the distributional and allocational roles of the center would be clearly separated. Because this property is not really valid further work on this question is needed.

15.5. Nash equilibria

Revelation of the truth under maximin behavior is of course a very weak incentive property. For this reason, further work has been done on the robustness of the MDP procedure with respect to alternative behavioral assumptions.

Roberts [1976a], following a suggestion of Malinvaud [1971] assumes myopic behavior and considers the Nash equilibrium obtained at each instant in the revelation game. The trajectory so obtained still converges to a Pareto optimum even though misrepresentation of marginal rates of substitution occurs at the Nash equilibrium.

At each instant, agent i maximizes

$$\dot{u}_i = \frac{\partial u_i}{\partial x_i}[q_i(t) - \psi_i(t) + \lambda_i(\sum_{j \neq i} \psi_j(t) + \psi_i(t) - \gamma(t))]\dot{K}(t)$$

$$\text{with } \dot{K}(t) = \sum_{j \neq i} \psi_j(t) + \psi_i(t) - \gamma(t) \qquad (15.11)$$

The optimal announcement $\phi_i(t)$ as a function of the answers of the other agents is (if $\lambda_i < 1$):

$$\phi_i(t) = \frac{1}{2(1 - \lambda_i)}[q_i - (1 - 2\lambda_i)(\sum_{j \neq i} \psi_j - \gamma)]. \qquad (15.12)$$

If we neglect the fact that $\phi_i(t)$ may become negative, a statement which would be rejected by a planner who knew that preferences were monotonic, there is a unique Nash equilibrium obtained from equations (15.12)[4]:

$$\phi_i = q_i - \left(\frac{1 - 2\lambda_i}{N - 1}\right)(\sum_i q_i - \gamma) \qquad (15.13)$$

when $\lambda_i < \frac{1}{2}$, which is surely the interesting case when there are many

[4] The procedure must start at $K > 0$ to avoid multiple Nash equilibria at $K = 0$.

agents, this equilibrium results in an underreporting of willingness to pay and a tax proportional to the stated willingness to pay, $-\psi_i \dot{K}$, when the level of the public good is increasing, and an overreporting of willingness to pay and a subsidy proportional to the stated willingness to pay, $-\psi_i \dot{K}$, when the public good is decreasing.

The differential equations which describe now the evolution of the system are obtained by replacing q_i by ϕ_i in (15.1) and (15.2).

$$\dot{K} = \frac{1}{N-1}\left[\sum_i q_i - \gamma\right] \quad (15.14)$$

$$\dot{x}_i = \left[-q_i + \tilde{\lambda}_i\left(\sum_j q_j - \gamma\right)\right]\dot{K} \quad (15.15)$$

with $\tilde{\lambda}_i = (1 - \lambda_i)/(N - 1)$, $i = 1, \ldots, N$.

With a slight change in the definition of the sharing rule and a decrease of the adjustment speed of the public good level this system is analogous to (15.1) (15.2) as far as stability and convergence to a Pareto optimum are concerned. Note, however, that now the speed of convergence decreases with the number of agents.[5]

The last question concerns the stability of the Nash equilibrium at each instant t, since the meaningfulness of this solution concept depends to a large extent on the economy's ability to reach it at each instant. It is easy to see that a continuous formalization of this local game is stable while a discrete formalization leads to an unstable game, leaving the question somewhat unsettled.

These mixed results have led Henry [1977][6] to investigate the stability properties of a discrete version of the MDP procedure as it would evolve if agents adopted the following behavior. At date t_0, agent i behaves as if the other agents were going to play the same strategies as at date $t_0 - 1$. This approach disentangles the double infinity implicit in the Roberts formulation (since at each instant a Nash game had to be played). Henry then shows that the procedure is stable provided that, when agents are indifferent between the truth and another strategy, they tell the truth. Paradoxically, the same condition needed to insure Pareto optimality in the infinite player case of the naive procedure reappears as a prerequisite for stability in the MDP process with strategic interactions taken into account.

[5] Henry [1977] has shown that the qualitative results described above subsist even if the positivity constraint $\psi_i(t) \geqq 0$ is imposed.
[6] See also F. Schoumaker [1977].

If one is not willing to make such an assumption (as for example in chapter 12), the descriptive relevance of this Nash approach is weakened, leaving the MDP procedure with relatively poor incentive properties. In the next chapter, we attempt to use the results of parts II and III to construct a class of procedures with very strong incentive properties.

Chapter 16

A NEW PROCEDURE

16.1. The procedure

In this chapter, we construct procedures for which truth telling is a dominant strategy at each instant t. We allow for non-separable utility functions, but we limit manipulability by assuming myopic behavior, as in ch. 15.

There is no a priori restriction on the stated marginal rate of substitution ψ_i for any i and any t. As in the Bowen procedure, the shares of the cost of the project are defined ex-ante; let δ_i be the share imputed to agent i[1]

$$\sum_i \delta_i = 1$$

At each instant t, we define the set of *pivotal agents*, $P(t)$, as the union of two sets: those whose statement changes the sign of the aggregate net willingness to pay ($P^2(t)$) and those whose statements cause the aggregate net willingness to pay to be non-zero when it would otherwise be zero or vice versa ($P^1(t)$). More formally:

$$\begin{aligned} i \in P^1(t) &\leftrightarrow \Big[\sum_j \psi_j(t) - \gamma(t)\Big]\Big[\sum_{j \neq i} \psi_j(t) - \delta_j \gamma(t)\Big] = 0 \\ i \in P^2(t) &\leftrightarrow \Big[\sum_j \psi_j(t) - \gamma(t)\Big]\Big[\sum_{j \neq i} \psi_j(t) - \delta_j \gamma(t)\Big] < 0 \end{aligned} \qquad (16.1)$$

The procedure is then defined by the following system of differential equations:

$$\begin{aligned} \dot{K}(t) &= +1 \quad \text{if } \sum_i \psi_i(t) - \gamma(t) > 0 \\ &= \ \ 0 \quad \qquad\qquad\qquad\quad\ \ = 0 \\ &= -1 \quad \qquad\qquad\qquad\quad\ \ < 0 \end{aligned} \qquad (16.2)$$

[1] The shares can be time dependent if the agents still behave myopically. However, myopia is then even more difficult to accept.

If $\sum_i \psi_i(t) - \gamma(t) \neq 0$,

$$\dot{S}(t) = -2 \sum_{i \in P(t)} \sum_{j \neq i} [\psi_j(t) - \delta_j \gamma(t)] \dot{K}(t) \tag{16.3}$$

$$\dot{x}_i(t) = -\delta_i \gamma(t) \dot{K}(t) + 2 \sum_{j \neq i} [\psi_j(t) - \delta_j \gamma(t)] \dot{K}(t) + \frac{1}{N} \dot{S}(t) \text{ if } i \in P(t)$$

$$= -\delta_i \gamma(t) \dot{K}(t) + \frac{1}{N} \dot{S}(t) \text{ if } i \notin P(t) \tag{16.4}$$

If $\sum_i \psi_i(t) - \gamma(t) = 0$, that is if every agent is in $P(t)$, then

$$\dot{x}_i(t) = -\left| \sum_{j \neq i} \psi_j(t) - \delta_j \gamma(t) \right| + \frac{1}{N} \left| \sum_l \sum_{j \neq l} \psi_j(t) - \delta_j \gamma(t) \right| \tag{16.5}$$

Equation (16.2) specifies the adjustment speed of the quantity of public good, according to the sign of the aggregate net willingness to pay. Equations (16.4) and (16.5) describe the change in the endowment of private good for pivotal and non-pivotal agents. $\dot{S}(t)$ defined in equation (16.3) is referred to as the surplus at instant t. This procedure is referred to as *procedure A*.

Note that equation (16.5) above defines the payments to be made when the stated willingness to pay would indicate that a Pareto optimum had been reached. This is necessary in order to preserve the correct incentive properties at this last instant. However we introduce a rule for stopping the process which mandates that $\dot{K}$ and $\dot{x}_i$ be identically zero for all t after a time such that $\dot{K} = 0$ has been reached. Thus, although individuals take (16.5) into account, it never really influences the path followed since it terminates at the same instant that (16.5) is applied.

Without the surplus $\dot{S}(t)$ being rebated to the agents, (16.3), (16.4) and (16.5) define precisely the transfers associated with the pivotal mechanism[2] when applied to the three projects $\dot{K} = +1, 0, -1$.

[2] From the definition of the pivotal mechanisms we can write the following system:

$$\dot{K}(t) = \begin{cases} +1 & \text{if } \sum_i \psi_i(t) - \gamma(t) > 0 \\ 0 & \phantom{\text{if } \sum_i \psi_i(t) - \gamma(t)} = 0 \\ -1 & \phantom{\text{if } \sum_i \psi_i(t) - \gamma(t)} < 0 \end{cases}$$

$$\dot{K}_i(t) = \begin{cases} +1 & \text{if } \sum_{j \neq i} (\psi_j(t) - \delta_j \gamma(t)) > 0 \\ 0 & \phantom{\text{if } \sum_{j \neq i} (\psi_j(t) - \delta_j \gamma(t))} = 0 \\ -1 & \phantom{\text{if } \sum_{j \neq i} (\psi_j(t) - \delta_j \gamma(t))} < 0 \end{cases}$$

$$\dot{S}(t) = -\sum_i \sum_{j \neq i} (\psi_j(t) - \delta_j \gamma(t)) (\dot{K}(t) - \dot{K}_i(t))$$

$$\dot{x}_i(t) = -\delta_i \gamma(t) \dot{K}(t) + \sum_{j \neq i} [\psi_j(t) - \delta_j \gamma(t)] [\dot{K}(t) - \dot{K}_i(t)] + \frac{1}{N} \dot{S}(t)$$

Rewriting this system somewhat we obtain precisely procedure A.

In order to avoid having the procedure continually generate a surplus, the payments made $(-\sum_i \dot{x}_i(t))$ must be rebated, and this rebate must be perceived by each agent to be independent of his own statement. Indeed, we know that

$$\sum \dot{x}_i(t) + \gamma(t) \dot{K}(t) \leqq 2 \sum_{i \in P(t)} \sum_{j \neq i} [\psi_j(t) - \delta_j \gamma(t)] \dot{K}(t) < 0 \tag{16.6}$$

by definition of pivotal agents, instead of

$$\sum_i \dot{x}_i(t) + \gamma(t) \dot{K}(t) = 0 \text{ from the feasibility constraint.} \tag{16.7}$$

For simplicity we suppose that the surplus is shared equally among all agents. It could be done more arbitrarily except that we want it to be distributed among a fairly large number of agents, as will be seen below.

In order to determine the agents' answers, $\psi_i(t)$, we have to make some behaviorial assumptions.

Assumption 4. Each agent behaves myopically.

Assumption 5. For large N, each agent neglects the impact of his answer on $(1/N)\dot{S}(t)$ in determining his optimal answer.

For agent i, let us denote $\dot{u}_i(\psi_i(t), \psi_{-i}(t))$ his *anticipated* change in utility, if he says $\psi_i(t)$ when the other agents answer $\psi_{-i}(t) = [\psi_1(t), \ldots, \psi_{i-1}(t), \psi_{i+1}(t), \ldots, \psi_N(t)]$. It is only the anticipated change, because the agent is assumed to neglect the impact of his answer on $(1/N)\ \dot{S}(t)$.

Theorem 16.1. Under assumptions 4 and 5, revelation of the truth at each instant t, is the only dominant strategy.

Proof. We want to show that:

$$\Delta_i = \dot{u}_i(q_i(t), \psi_{-i}(t)) - \dot{u}_i(\psi_i(t), \psi_{-i}(t)) \geqslant 0 \tag{16.8}$$

for any $\psi_i(t)$, $\psi_{-i}(t)$, with a strict inequality for some $\psi_i(t)$.

Below, we check that it is true for several cases leaving symmetric situations to the reader.

If:

$$\sum_{j \neq i} \psi_j(t) + q_i(t) > \gamma(t) \tag{16.9}$$

$$\sum_{j \neq i} \psi_j(t) + \psi_i(t) > \gamma(t)$$

and,

$$\sum_{j\neq i}(\psi_j(t) - \delta_j\gamma(t)) < 0$$

then,

$$\Delta_i = u_{ix}[K(t), x_i(t)]\Big[q_i(t) - \delta_i\gamma(t) + 2\sum_{j\neq i}(\psi_j(t) - \delta_j\gamma(t)) + \frac{1}{N}\dot{S}(t) - q_i(t) + \delta_i\gamma(t) - 2\sum_{j\neq i}(\psi_j(t) - \delta_j\gamma(t)) - \frac{1}{N}\dot{S}(t)\Big] = 0$$

if

$$\sum_{j\neq i}\psi_j(t) + q_i(t) > \gamma(t)$$
$$\sum_{j\neq i}\psi_j(t) + \psi_i(t) = \gamma(t) \tag{16.10}$$

and,

$$\sum_{j\neq i}[\psi_j(t) - \delta_j\gamma(t)] < 0$$

then,

$$\Delta_i = u_{ix}[K(t), x_i(t)]\Big[\sum_{j\neq i}\psi_j(t) + q_i(t) - \gamma(t)\Big] \geqslant 0$$

if

$$\sum_{j\neq i}\psi_j(t) + q_i(t) = \gamma(t)$$
$$\sum_{j\neq i}\psi_j(t) + q_i(t) = \gamma(t) \tag{16.11}$$

and,

$$\sum_{j\neq i}[\psi_j(t) - \delta_j\gamma(t)] > 0 \text{ (and therefore } q_i(t) < \delta_i\gamma(t))$$

then,

$$\Delta_i = u_{ix}[K(t), x_i(t)][-q_i(t) + \delta_i\gamma(t)] > 0$$

if

$$\sum_{j\neq i}\psi_j(t) + q_i(t) = \gamma(t)$$
$$\sum_{j\neq i}\psi_j(t) + \psi_i(t) < \gamma(t) \tag{16.12}$$

and,

$$\sum_{j\neq i}[\psi_j(t) - \delta_j\gamma(t)] > 0$$

then,

$$\Delta_i = u_{ix}[K(t), x_i(t)]\Big[q_i(t) - \delta_i\gamma(t) + \sum_{j\neq i}(\psi_j(t) - \delta_j\gamma(t))\Big] = 0$$

Q.E.D.

Thus, if the agents neglect the impact of their answers on their share of the surplus generated they will respond truthfully. A justification of this behavior is that their share in the surplus is very small. In chapter 9, we have shown that, if the tastes of the population are smoothly distributed, the per capita surplus goes to zero as the square root of the size of the population N when the mean of the net willingnesses to pay in the population is zero, and converges to zero faster than any power of N when the mean is not zero. Using these results we assume that for large N the contribution of the per capita surplus is so minimal that it justifies the bounded rationality behavior we have just specified.

Theorem 16.2. Under assumptions 2, 4 and 5, procedure A converges to a Pareto optimum.

Proof. See proof of theorem 14.4.

In this procedure it is necessary to have a constant speed of adjustment. At each instant, we could elicit the linear utility function $q_i\dot{K} + \dot{x}_i$, but the maximization of welfare by the center would give infinite adjustment speeds whenever $\sum_i q_i \neq \gamma$.

16.2. On individual rationality

It is important to observe that this procedure is not individually rational, in the sense of improving the welfare of everybody with respect to the initial situation, unless very strong assumptions are made (such as "the public good is necessary") which de facto make the initial position the worst for every agent. When the public good is desired by everybody it is possible to insure that some agents will see their final utility at least as large as the initial utility by equating their imputed share of the project to zero. This allows us to have a procedure which can favor a subset of agents. However, for a given distribution of agents the procedure is certainly not neutral, in the sense of the last chapter.

However, one can conjecture a kind of approximate neutrality in large economies, in so far as it is shown that the transfers in private good necessary to run this process are of small order with respect to the increase of welfare brought about by the production of public good.

In an economy where the initial situation corresponding to a distribution of the private good has no reason to be fair, the requirement of individual rationality looses much of its appeal. It is clear that the spirit of the procedures proposed here is to go beyond the Pareto optimal criterion by using a cardinal representation of the utility functions and by maximizing the social welfare represented by the sum of utilities. By choosing equal shares of the cost, the procedure can be made equitable in the following sense: if the agents consider the procedure before knowing their own preferences, in the spirit of the Rawlsian approach, no particular agent is favored. However, this procedure is certainly not optimal for extremely risk averse agents, as required by Rawls [1971]. This is in the same spirit as our results of chapter 6 on the concept of individual rationality on average.

The lack of individual rationality also creates the possibility of bankruptcy. Two points of view can be taken on this issue. First one may consider that we are dealing with small projects relatively to the wealth of the agents and, then, there is no real problem of bankruptcy. If one is willing to handle big public projects with such a method, bankruptcy must be faced. An agent may be bankrupted for two reasons. On the one hand he may be unable to pay his ex ante imputed share of the project; in that case one may legitimately assume that the planner has sufficient information on the endowments so that he does not impute shares which could bankrupt some agents. On the other hand, he may have to make pivotal payments beyond his wealth. Note however, that the mechanism is such that a pivotal payment is always less than twice his marginal willingness to pay. Therefore, if one bounds the marginal rate of substitution, it is possible to limit the eventual pivotal payments to an amount below the endowment, net of the cost share.

16.3. Several public goods

The generalization of the above results to the case of several public goods faces the difficulty of proving stability. In subsection 16.3.1 we construct a procedure converging to a Pareto optimum when the utility functions are separable. In subsection 16.3.2, we discuss the difficulties of proving stability without such an assumption.

16.3.1. Separability

We assume here that the utility functions of the N agents can be written:

$$u_i(K, x_i) = x_i + v_i(K_1, \dots, K_L) \qquad (16.13)$$

where L is the number of public goods and $K = (K_1, \dots, K_L)$. The other assumptions and notations of section 2 are trivially generalized here.

A Pareto optimal allocation $(K_1, \dots, K_L, x_1, \dots, x_N)$ is now characterized by:

$$\sum_i q_{il}(K, x_i) = \gamma_l(K), \qquad l = 1, \dots, L \qquad (16.14)$$

$$\Gamma(K) + \sum_i x_i = \sum_i \bar{x}_i \qquad (16.15)$$

We define imputed cost shares for each public project δ_{il}, $i = 1, \dots, N$; $l = 1, \dots, L$, as well as a set of pivotal individuals, $P_l(t) = P_l^1(t) \cup P_l^2(t)$; for each public project.

$$\begin{aligned} i \in P^1(t) &\leftrightarrow [\sum_j \psi_{jl}(t) - \gamma_l(K(t))]\,[\sum_{j \neq i} [\psi_{jl}(t) - \delta_{jl}\gamma_l(K(t))]] < 0 \\ i \in P_l^2(t) &\leftrightarrow [\sum_j \psi_{jl}(t) - \gamma_l(K(t))]\,[\sum_{j \neq i} [\psi_{jl}(t) - \delta_{jl}\gamma_l(K(t))]] = 0 \end{aligned} \qquad (16.16)$$

The procedure is then defined by

$$\begin{aligned} \dot{K}_l(t) &= +1 && \text{if } \sum_i \psi_{il}(t) - \gamma_l(K(t)) > 0 \\ &= 0 && \text{if } \qquad = 0 \\ &= -1 && \text{if } \qquad < 0 \qquad l = 1, \dots, L \end{aligned} \qquad (16.17)$$

$$\begin{aligned} \dot{K}_{il}(t) &= +1 && \text{if } \sum_{j \neq i} [\psi_{jl}(t) - \delta_{jl}\gamma_l(K(t))] > 0 \\ &= 0 && \text{if } \qquad = 0 \\ &= -1 && \text{if } \qquad < 0 \qquad l = 1, \dots, L \end{aligned} \qquad (16.18)$$

$$\dot{S}_l(t) = -\sum_{i \in P_l(t)} d_{il}\,[\sum_{j \neq i} (\psi_{jl}(t) - \delta_{jl}\gamma_l(K(t)))]\,[K_l(t) - K_{il}(t)] \qquad l = 1, \dots, L \qquad (16.19)$$

where

$$\begin{aligned} d_{il} &= 2 && \text{if } i \in P_l^2(t) \\ &= 1 && \text{if } i \in P_l^1(t) \end{aligned}$$

$$\dot{x}_i(t) = -\sum_l \delta_{il}\gamma_l(K(t)) \cdot \dot{K}_l(t) + \sum_{\{l/i \in P_l(t)\}} d_{il} \sum_{j \neq i} [\psi_j(t) - \delta_{jl}\gamma_l(K(t))]\,[\dot{K}_l(t) - \dot{K}_{il}(t)] + \frac{1}{N}\sum_l \dot{S}_l(t) \qquad (16.20)$$

If $\dot{K}_l(t) = 0$ for all l, then the process stops; but if $\dot{K}_l(t) \neq 0$ for any l, then the process continues for all commodities, as stated.

It is immediate to show that under the analog of assumptions 4 and 5 (assumptions 4 and 6) truthful revelation of preferences is a dominant strategy for each agent.

Assumption 6. For any K, there exists a level of public good production $\overline{\overline{K}}_l$ such that for any $K_{-l} = (K_1, \ldots, K_{l-1}, K_{l+1}, \ldots, K_L)$ and any $x_1, \ldots, x_N$

$$\sum_i q_{il}(\overline{\overline{K}}_l, K_{-l}, x_i) - \gamma_l(\overline{\overline{K}}_l, K_{-l}) \text{ is negative.}$$

The system of differential equations so described is a complex one, in particular because of a number of discontinuities in the right-hand side.

First, the speed of adjustment for the different public goods is discontinuous. In addition, when the speed of adjustment for a given public good changes sign at instant t_0, the set of pivotal agents changes, and even though the sum of pivotal payments is the same when

$$\underset{t > t_0}{t \to t_0} \text{ or when } \underset{t < t_0}{t \to t_0}, \text{ at } t = t_0$$

this sum is zero because $\dot{K} = 0$. Also, the rate of monetary transfer to an agent who at $\dot{K} = 0$ ceases to be (or becomes) pivotal is discontinuous. Note however that when the speed of adjustment is unchanged, the transfer of an agent who ceases to be (or becomes) pivotal is unchanged. Therefore, the system of differential equations has discontinuities only at instants when the quantities of some (and not of all) public goods are stationary. It may be hoped that the methods of Henry [1972] can be used to establish the following result.

Under appropriate assumptions, there exists a solution to the system of differential equations (16.4), (16.5), (16.6) and the correspondence which associates to an allocation the set of finite time trajectories starting from this allocation is included in a compact convex set and is upper hemi continuous.

Assuming that these results are true, we can then prove:

Theorem 16.3. Under assumptions 2, 4, 6, the procedure A with many public goods converges to a Pareto optimum if preferences are "separable".

Proof. Clearly a stationary point of the procedure is a Pareto optimum. Consider the following Lyapunov function:

$$V(t) = \sum_i u_i(K, x_i) \tag{16.21}$$

$$\dot{V}(t) = \sum_i [\dot{x}_i + \sum_l q_{il}\dot{K}_l] \tag{16.22}$$

$$\dot{V}(t) = \sum_i \sum_l [q_{il} - \delta_{il}\gamma_l(K(t))]\dot{K}_l > 0 \tag{16.23}$$

$$\dot{V}(t) = 0 \quad \text{iff the allocation is Pareto optimum.} \tag{16.24}$$

From assumption 3, with (16.23) K_l must be bounded above for every l. Therefore $V(t)$ must be bounded above.

Clearly, $V(t)$ is an increasing function of time which is bounded above. From theorem 15.2 and theorem 6.2 in Champsaur, Drèze, Henry [1977] we know that the procedure is quasi stable – that is, any limit point of any trajectory is a stationary point. It is immediate then to obtain stability by using the strict quasi-concavity of the utility functions and the fact that any stationary point is a Pareto optimum. Q.E.D.

16.3.2. Instability without separability

Without separability the transfers in private good may disturb the stability of the procedure. Suppose there are two public goods. With appropriate concavity assumptions, it is clear that the change in the production of public goods brings us closer to the Pareto surface as $|\sum_i q_{il} - \gamma_l|$ $l = 1, 2$ decrease. We approach a Pareto allocation with given levels of public good productions. But, in the procedure, we are obliged to make transfers in private good to an agent who will favor a different structure of public goods. Then, as we approach this different mixture, it could be that transfers are now made to another agent who reverses the trajectory towards the previous combination. During a cycle, the utility level of one agent increases and then decreases, while for another agent changes in utility have the opposite characteristics. This is made possible by the lack of individual rationality of the procedure. Such a phenomenon cannot occur with separability since the transfers in private good do not affect the decisions about public goods.

An example having the qualitative features just described utilizing piece-wise linear utility functions is now presented:

The example involves two public goods with constant marginal costs equal to 2 and two agents who share equally the costs of the public goods. Let $\bar{x}_i$ be the initial amount of private good held by each agent, $i = 1, 2$. $\bar{x}_i$ is large enough so that no bankruptcy occurs along the trajectories.

The preferences are locally defined along the trajectory as follows; appropriate increasing transformations can be used to make these preferences continuous.

Agent 1

For $\bar{x}_1 - 2 \leqslant x_1 \leqslant \bar{x}_1$
$0 \leqslant K_1 = K_2 < 1$ then $U_1 = 2K_1 + 2K_2 + x_1$

For $\overline{x}_1 - 2.75 \leqslant x_1 \leqslant \overline{x}_1 - 2$
$K_1 = 1$
$1 \leqslant K_2 < 2$ then $U_1 = K_1 + \frac{3}{4}K_2 + x_1$

For $\overline{x}_1 - 3.25 \leqslant x_1 \leqslant \overline{x}_1 - 2.75$
$K_1 + K_2 = 3$
$1 \leqslant K_1 < 2, 1 < K_2 \leqslant 2$ then $U_1 = 3K_1 + x_1$

For $\overline{x}_1 - 3.25 \leqslant x_1 \leqslant \overline{x}_1 - 2$
$K_2 = 1$
$1 \leqslant K_1 \leqslant 2$ then $U_1 = 1.25K_1 + K_2 + x_1$

Agent 2

For $\overline{x}_2 - 2 \leqslant x_2 \leqslant \overline{x}_2$
$0 \leqslant K_1 = K_2 < 1$ then $U_2 = 2K_1 + 2K_2 + x_2$

For $\overline{x}_2 - 3.25 \leqslant x_2 \leqslant \overline{x}_2 - 2$
$K_1 = 1$
$1 \leqslant K_2 < 2$ then $U_2 = K_1 + 4K_2 + x_2$

For $\overline{x}_2 - 3.25 \leqslant \leqslant \overline{x}_2 - 2.75$
$K_1 + K_2 = 3$
$1 \leqslant K_1 < 2, 1 < K_2 \leqslant 2$ then $U_2 = 0.75K_1 + 1.25K_2 + x_2$

For $\overline{x} - 3.25 \leqslant x_2 \leqslant \overline{x}_2 - 2$
$K_2 = 1$
$1 \leqslant K_1 \leqslant 2$ then $U_2 = 0.25K_1 + K_2 + x_2$.

In the space of public goods the trajectory is as shown figure 16.1.

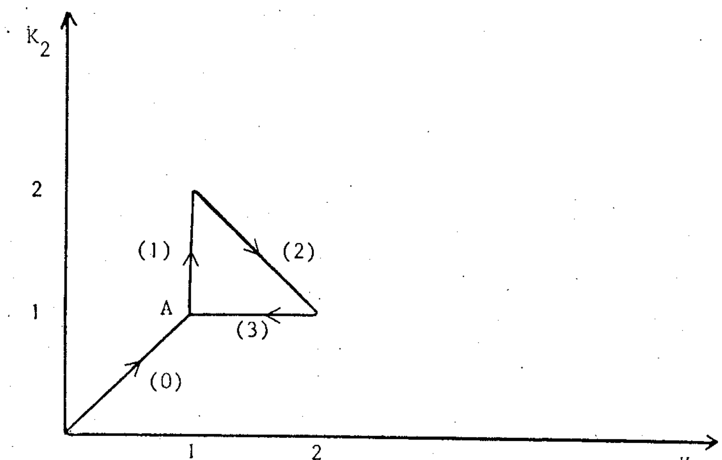

Figure 16.1.

In the region (0) (see figure 16.1), there is no pivot, and both projects are constructed in increasing quantity. Then we enter region (1) with incomes $\overline{x}_1 - 2$ and $\overline{x}_2 - 2$ for the two agents. Then agent 2 is a pivot for the second public good and he has to pay $\frac{1}{2}$ with a rebate of $\frac{1}{4}$ i.e. $\frac{1}{4}$. In region (1) his income decreases from $\overline{x}_2 - 2$ to $\overline{x}_2 - 3.25$ since he has also to pay his share of the cost of the second public good i.e. 1. Agent 1 on the other hand receives $\frac{1}{4}$ and has to pay his share of the cost. In region (2), agent 1 is a pivot for public good 1 because he causes it to be increased, and a pivot for public good 2 but in the opposite sense. Finally in region 3, agent 2 is a pivot in the downward direction. Then we are back at A and cycle indefinitely in (1), (2), (3).

16.4. Other interpretations of procedure A

Instead of redistributing the surplus one may choose to destroy it in order to be sure of not distorting incentives.

The loss in efficiency can then be measured in private good as:

$$L = \int_0^{T^*} \dot{S}(t)\,\mathrm{d}t \leqq \int_0^{T^*} 2 \sum_{i \in P(t)} \Big[\sum_{j \neq i} (\psi_j(t) - \delta_j \gamma(t))\Big]\,\mathrm{d}t \tag{16.25}$$

where T^* is the convergence time of the procedure with

$$T^* = \int_0^{T^*} \dot{x}(t)\,\mathrm{d}t = K^*, \text{ optimal quantity of public good.} \tag{16.26}$$

L expresses the loss to incur in order to obtain perfect incentive compatibility. It is intuitively clear from the results of chapter 4 that the per capita loss can be made as small as desired for N large enough. We show it below in a special case with a simple argument.

Consider a distribution of net willingnesses to pay whose support is $[-A, +A]$. The population is considered as a rth replicate of a sample of size n drawn from this distribution. Let $v_i(t) = q_i(t) - \delta_i \gamma(t)$ be the net willingness to pay at instant t in the separable case. Let $z(t) = \sum_{i=1}^{rn} v_i(t)$.

We start from a situation where $\sum_{i=1}^{n} v_i(0) > 0$; for r large enough $\sum_{i=1}^{rn} v_i(0) > A$. The convergence time which is equal to the optimal size of the public good is clearly independent of r.

Let $T_1(r)$ be the first instant at which $z(t) = A$. By a Taylor expansion, we have:

$$z(t) = (t - T^*)\dot{z}(T^*) + \frac{(t - T^*)^2}{2}\ddot{z}(\tilde{t}) \text{ with } \tilde{t} \in [t, T^*] \qquad (16.27)$$

Since $\dot{z}(T^*) = 0$ and $\ddot{z}(\tilde{t}) = r\sum_{i=1}^{n} \ddot{v}_i(\tilde{t})$

$$T^* - T_1(r) = \frac{\sqrt{(2A)}}{\sqrt{\left(r \sum_{i=1}^{n} \ddot{v}_i(\tilde{t})\right)}} \qquad (16.28)$$

If $\sum_{i=1}^{n} \ddot{v}_i(\tilde{t})$ is bounded below by a positive number, that is, if there is a minimum degree of decrease in the marginal rate of substitution, $T^* - T_1(r)$ is of the order of $1/\sqrt{(r)}$.[3]

When $z(t)$ becomes smaller than A, it is possible that some agent becomes pivot; however after $T_1(r)$, the transfer paid by an agent is always less than $2A$, since his own net willingness to pay is less than A. A weak bound on L is therefore:

$$L \leqslant (T^* - T_1(r))rnA \propto \sqrt{(r)} \qquad (16.29)$$

Hence, the per capita loss in efficiency converges to zero as r goes to infinity.

Another point of view is to assume that agents attempt to take into account how their answers influence their rebates, $(1/N)\dot{S}(t)$, even though it is very small. Suppose that the mean of the decision maker's expectations is the same as the mean of the agent's expectations. Then, the government should go ahead as long as his mean stays positive, reaching a level of public good production $\bar{K}$, without asking the population. Then, he should start the procedure (procedure B); a complete formalization of the problem should specify the way expectations evolve over time; we will make the simplifying:

Assumption 7. In procedure B, expectations stay fixed with a zero mean.

This is reasonable only because we start from a good first guess. Another solution might be to use in a first phase the Bowen procedure until convergence and then start procedure B. However, agents may want to manipulate the outcome of Bowen procedure, in order to minimize their

[3] For separable utility functions, this amounts to put a lower bound on the curvature on the evaluation function; for non-separable utility functions the matter is more delicate and assumptions on the curvature of the cost function as well as on the elasticity of substitution of the utility function are needed.

expected pivotal payments in the second phase. For example, if the Bowen procedure with truthful answers yields a level of public good production larger than the optimal level (well defined in the separable case), they will try to stop Bowen procedure at the optimum to avoid making pivotal payments when in the second phase procedure B will drive back the level of public good production to the optimal level. However if we maintain the myopic assumption in the first phase too, this type of manipulation is avoided.

Assume for example as in section 4.7, that at each instant the willingnesses to pay belong to the same compact set. From theorem 9.2, we know that for large N, the maximization problem of any agent is concave, and, therefore, we can restrict the allowable stated marginal rates of substitution to be continuous functions of time. From theorem 9.3, we know that for any t, $\psi_i(t)$ is as closed as wished to $q_i(t)$; for any ε, there exists N^* such that $N > N^*$ implies

$$|\psi_i(t) - q_i(t)| < \varepsilon, \quad \text{for all } i. \qquad (16.30)$$

With assumption 6, we can conclude that the quantity of public good which may be produced is bounded (at least for large N). Therefore we can immediately state.

Theorem 16.4. Under assumptions 4 and 7, procedure B with a single public good converges.

It is clear however that because of the slight distortions made by the agents in reporting their preferences, a stationary point of procedure B is not a Pareto optimum.

Let z^* be the allocation at which procedure B stops. From chapter 9, we know that for all $\varepsilon > 0$, and $\eta > 0$, there exists N_0 such that $N > N_0$ implies

$$\Pr\left[\left|\sum_i q_i(z^*) - \sum_i \psi_i(z^*)\right| > \varepsilon\right] < \eta. \qquad (16.31)$$

Since at z^*, $\sum_i \psi_i(z^*) = \gamma(z^*)$, we can say that for large populations with probability close to one we stop at an allocation such that it is not possible to increase the sum of willingnesses to pay at a rate faster than ε. This is only a local result. We have seen however in chapter 14 how lower boundedness assumptions on the derivative of marginal productivity or on the absolute value of second derivatives of utility functions can transform this local result into approximate global Pareto optimality.

REFERENCES

Arrow, K.J. [1951] *Social Choice and Individual Values*, J. Wiley and Sons, New York.

Arrow, K.J. [1969] "The organization of economic activity: Issues pertinent to the choice of market versus nonmarket allocation", in *U.S. Congress Joint Economic Committee, 91st Congress, 1st. Session.*

Arrow, K.J. and F.H. Hahn [1971] *General Competitive Analysis*, Holden-Day, San Francisco.

Arrow, K.J. and L. Hurwicz [1960] "Decentralization and computation in resource-allocation", in R. Pfouts (ed.). *Essays in Economics and Econometrics*, in honor of Harold Hotelling, University of North Carolina Press, Chapel Hill, N.C.

d'Aspremont, C. and L. Gérard-Varet [1977], "Incentives and incomplete information", CORE, D.P. No. 7705.

Balasko, Y. [1975] "Some results on uniqueness and on stability of equilibrium in general equilibrium theory", *Journal of Mathematical Economics*, 2, 93–118.

Bergson, A. [1967] "Market socialism revisited", *Journal of Political Economy*, 75, 655–673.

Berliner, J. [1975] *Innovation in Soviet Industry*, MIT Press, Cambridge, MA.

Black, D. [1948] "On the rationale of group decision-making", *Journal of Political Economy*, 56, 23–24.

Bohm, P. [1972] "Estimating demand for public goods: an experiment", *European Economic Review*, 3, 111–130.

Bowen, H. [1943] "The interpretation of voting in the allocation of economic resources", *Quarterly Journal of Economics*, 58, 27–48.

Brown, C., G. Fane and J. Medoff [1973] "The income distribution as a pure public good: comment", *Quarterly Journal of Economics*, 87, 296–303.

Brown, D.J. [1973] "Acyclic choice", Cowles Foundation Discussion Paper, Yale University.

Champsaur, P. [1976] "Neutrality of planning procedures in an economy with public goods", *Review of Economic Studies*, 43, 293–300.

Champsaur, P. [1977] "Lindahl equilibria and rationing", INSEE. DP. Paris.

Champsaur, P., J. Drèze, and C. Henry [1977] "Stability theorems with economic applications", *Econometrica*, 45, 273–294.

Clarke, E.H. [1971] "Multipart pricing of public goods", *Public Choice*, 19–33.

Coase, R. [1960] "The problem of social cost", *Journal of Law and Economics*, 3, 1–44.

Cornet, B. [1977] "On planning procedures defined by multivalued differential equations" in *Systèmes Dynamiques et Modèles Economiques*, CNRS.

David, H.A. [1970] *Order Statistics*, Wiley, New York.

Debreu, G. [1970] "Economies with a finite set of equilibria", *Econometrica*, 38, 387–392.

Drèze, J. and D. de la Vallée Poussin [1971] "A tâtonnement process for public goods", *Review of Economic Studies*, 38, 133–150.

Dubins, L. [1974] "Group decision devices", mimeo, Berkeley.

Dutta, B. and P.K. Pattanaik [1975] "On nicely consistent voting schemes", mimeo, Delhi School of Economics.

Fishburn, D. [1972] "Lotteries and social choice", *Journal of Economic Theory*, 5, 189–207.

Foley, D. [1967] "Resource allocation and the public sector", *Yale Economic Essays*, 7, 45–98.

Gibbard, A. [1973] "Manipulation of voting schemes. A general result", *Econometrica*, 41, 587–601.
Gibbard, A. [1977] "Manipulation of schemes that mix voting with chance", *Econometrica*, 45, 665–682.
Grandmont, J.-M. [1978] "Intermediate preferences and the majority rule", *Econometrica*, 46, 317–330.
Groves, T. [1973] "Incentives in teams", *Econometrica*, 41, 617–631.
Groves, T. [1974] "Information, incentives, and the internalization of production externalities", D.P. No. 87, Center for Mathematical Studies in Economics and Management Science, Northwestern University.
Groves, T. and J. Ledyard [1975] "An incentive mechanism for efficient resource allocation in general equilibrium with public goods", D.P. No. 119, Center for Mathematical Studies in Economics and Management Science, Northwestern University.
Groves, T. and M. Loeb [1975] "Incentives and public inputs", *Journal of Public Economics*, 4, 311–226.
Heal, G. [1973] *The Theory of Economic Planning*, North-Holland Press, Amsterdam-London.
Henry, C. [1972] "Differential equations with discontinuous right-hand side for planning procedures", *Journal of Economic Theory*, 4, 545–557.
Henry, C. [1977] "The free rider problem in the MDP procedure", Ecole Polytechnique D.P. No. A1640577, Paris.
Hildenbrand, W. [1974] *Core and equilibria of a large economy*. Princeton University Press, Princeton, N.J.
Hume, D.A. [1888] *Treatise of Human Nature*, Oxford University Press, Oxford.
Hurwicz, L. [1960] "Optimality and informational efficiency in resource allocation processes", in K. Arrow, S. Karlin and P. Suppes (eds.) *Mathematical Methods in the Social Sciences*, Stanford, 27–46.
Hurwicz, L. [1972] "On informationally decentralized systems", in R. Radner and C.B. McGuire (eds.), *Decision and Organization*, North-Holland, Amsterdam.
Hurwicz, L. [1976] "Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points", mimeo, Berkeley.
Hurwicz, L. [1977] "On allocations attainable through Nash equilibria", mimeo, Berkeley.
Hurwicz, L. and D. Schmeidler [1976] "Construction of outcome functions guaranteeing existence and Pareto optimality of Nash equilibria", mimeo, University of Minnesota.
Johansen, L. [1977] "The theory of public goods: misplaced emphasis", *Journal of Public Economics*, 7, 147–152.
Kalai, E. and E. Muller [1976] "Characterization of domains admitting non-dictatorial social welfare functions and non-manipulable voting procedures", The Center for Mathematical Studies in Economics and Management Science, Northwestern University.
Kelly, J. [1974] "Voting anomalies, the number of voters and the number of alternatives", *Econometrica*, 42, 239–257.
Kurz, M. [1974] "Experimental approach to the determination of the demand for public goods", *Journal of Public Economics*, 3, 329–348.
Lange, O. [1936–37] "On the economic theory of socialism", *Review of Economic Studies*, 4.
Lerner, A.P. [1966] *The Economics of Control*, Macmillan, New York.
Liberman, E. [1962] "Plan, Pribil, Premia" in *Pravda*.
Lindahl, E. [1919] *Die Gerechtigkeit der Besteuerung*, Lund. English translation in Musgrave and Peacock [1958].
Loève, M. [1963] *Probability Theory*, Van Nostrand, Princeton, N.J.
Malinvaud, E. [1971] "A planning approach to the public good production", *Swedish Journal of Economics*, 73, 96–112.
Malinvaud, E. [1972a] "Prices for individual consumption, quantity indicators for collective consumption", *Review of Economic Studies*, 39, 385–405.

Malinvaud, E. [1972b] "The allocation of individual risks in large markets", *Journal of Economic Theory*, 4, 312–328.
Marwell, G. [1976] "The illogic of collective action: extensions and tests of a theory", Department of Sociology, Wisconsin, mimeo.
Mas-Colell, A. and H. Sonnenschein [1972] "General possibility theorems for group decisions", *Review of Economic Studies*, 39, 185–192.
Maskin, E. [1976a] "Social choice on restricted domains", unpublished Harvard Ph.D. Thesis.
Maskin, E. [1976b] "Game forms and allocation mechanisms: incentive compatibility, manipulation, truthful revelation" mimeo, Jesus College, Cambridge.
Meade, J. [1952] "External economies and diseconomies in a competitive situation", *Economic Journal*, 62, 54–67.
Musgrave, R.A. [1939] "The voluntary exchange theory of public economy", *Quarterly Journal of Economics*, 53, 213–237.
Musgrave, R.A. [1959] *The Theory of Public Finance*, McGraw Hill, New York.
Musgrave, R.A. and A.T. Peacock, eds., [1958] *Classics in the Theory of Public Finance*, MacMillan, London.
Newbery, D. [1974] "Experimental approach to the determination of the demand for public goods; a comment", *Journal of Public Economics*, 3, 425–429.
Pattanaik, P.K. [1975] "Strategic voting without collusion under binary and democratic group decision rules", *Review of Economic Studies*, 42, 93–104.
Peleg, B. [1976a] "On consistent voting systems", Center for Research in Mathematical Economics and Game Theory, no. 21, The Hebrew University, Jerusalem.
Peleg, B. [1976b] "A note on manipulation in large voting schemes", Center for Research in Mathematical Economics and Game Theory, No. 22, The Hebrew University, Jerusalem.
Postlewaite, A. and D. Schmeidler [1978] "Approximate efficiency of non-Walrasian Nash equilibria", *Econometrica*, 46, 129–135.
Rao, C.R. [1973] *Linear Statistical Inference and its Applications*, John Wiley, London.
Rawls, J. [1971] *A Theory of Justice*, Harvard University Press, Cambridge, Mass.
Renyi, A. [1966] *Calcul des Probabilités*, Dunod, Paris.
Roberts, J. [1976a] "Incentives in planning procedures for the provision of public goods", CORE, D.P. No. 7611.
Roberts, J. [1976b] "The incentives for correct relation of preferences and the number of consumers", *Journal of Public Economics*, 6, 359–374.
Roberts, J. and A. Postlewaite [1976] "The incentives for price-taking behavior in large exchange economies", *Econometrica*, 44, 115–128.
Samuelson, P. [1954] "The pure theory of public expenditures", *Review of Economics and Statistics*, 36, 387–389.
Samuelson, P. [1955] "Diagrammatic exposition of a theory of public expenditures", *Review of Economics and Statistics*, 37, 350–356.
Samuelson, P. [1969] "Pure theory of public expenditure and taxation", in *Public Economics*, J. Margolis and H. Guitton (eds.), Macmillan-St. Martin Press, New York, 98–123.
Sandmo, A. [1976] "Optimal taxation. An introduction to the literature", *Journal of Public Economics*, 6, 37–54.
Satterthwaite, M. [1975] "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions", *Journal of Economic Theory*, 10, 187–217.
Savage, L.J. [1971] "Elicitation of personal probabilities and expectations", *Journal of the American Statistical Association*, 66, 783–801.
Schmeidler, D. [1976] "A remark on microeconomic models of an economy and on a game theoretic interpretation of Walras equilibria", mimeo.
Schmeidler, D. and H. Sonnenschein [1974] "The possibility of a cheat-proof social choice

function: A theorem of A. Gibbard and M. Satterthwaite", Northwestern Discussion Paper, No. 89.
Schoumaker, F. [1976] "Best reply strategies on the Champsaur, Drèze and Henry procedure", Center for Mathematical Studies in Economics and Management Science, D.P. No. 262, Northwestern University.
Sen, A. [1970] *Collective Choice and Social Welfare*, Holden Day, San Francisco.
Sen, A. [1977] "Social choice theory: a reexamination", *Econometrica*, 45, 53–90.
Shapley, L. [1976] "Non-cooperative general exchange", Rand Corporation, P–5286.
Sherr, B.A. and E.M. Babb [1975] "Pricing public goods: an experiment with two proposed pricing systems", *Public Choice*, 23, 35–48.
Smets, H. [1972] "Le principe de la compensation reciproque: un instrument économique pour la solution de certains problèmes de pollution transfrontière", O.C.D.E., Direction de l'environnement.
Smith, V. [1976] "Incentive-compatible experimental processes for the provision of public goods", NBER Conference on Decentralization.
Sonnenschein, H. [1971] "Demand theory without transitive preferences, with applications to the theory of competitive equilibrium", Chapter 10 in *Preferences, Utility and Demand: a Minnesota Symposium*, J.S. Chipman et al, eds., Harcourt, Brace, Jovanovich, New York.
Stiglitz, J.E. [1974] "The theory of local public goods", Stanford Tech. Report No. 144.
Sweeney, J.W. Jr. [1973] "An experimental investigation of the free rider problem", *Social Science Research*, 2, 277–92.
Thompson, E.A. [1967] "A Pareto optimal group decision process", *Papers on Non-Market Decision Making*, University of Virginia.
Thurow, L. [1971] "The income distribution as a pure public good", *Quarterly Journal of Economics*, 85, 327–336.
Tideman, N. [1972] "The efficient provision of public goods", in *Public Prices for Public Products*, S. Mushkin (ed.), The Urban Institute, Washington.
Tiebout, C.M. [1956] "A pure theory of local expenditures", *Journal of Political Economy*, 64, 416–424.
Tulkens, H. [1977] "Allocation and preference revelation mechanisms for public goods" mimeo.
Tulkens, H. and S. Zamir [1976] "Local games in dynamic exchange processes", CORE discussion paper No. 7606.
Tullock, G. [1959] "Problems of majority voting", *Journal of Political Economy*, 67, 571–579.
Uzawa, H. [1961] "The stability of dynamic processes", *Econometrica*, 29, 617–631.
Vickrey, W. [1961] "Counterspeculation, auctions and competitive sealed tenders", *Journal of Finance*, 16, 1–17.
Walker, M. [1978] "A note on the characterization of mechanisms for the revelation of preferences", *Econometrica*, 46, 147–152.
Weitzman, M.L. [1976] "The new Soviet incentive model", *The Bell Journal of Economics*, 7, 251–257.
Westhoff, F. [1977] "Existence of equilibria in economies with a local public good", *Journal of Economic Theory*, 1, 84–112.
Wicksell, K. [1896] Finanztheoretische Untersuchungen and das Steuerwesen Schweden's, Jena, Germany, English translation in Musgrave and Peacock [1958].
Wilson, R. [1978] "Competitive Exchange", *Econometrica*, 46, 577–585.
Zeckhauser, R. [1969] "Majority rule with lotteries on alternatives", *Review of Economic Studies*, 42, 696–703.
Zeckhauser, R. [1973] "Voting systems, honest preferences and Pareto optimality", *The American Political Science Review*, 67, 934–966.

AUTHOR INDEX

SUBJECT INDEX

Note: Bold face number shows the page where the definition of the concept is given.