North-Holland

Cost Allocation: Methods, Principles, Applications

edited by
H. Peyton YOUNG

COST ALLOCATION:
Methods, Principles, Applications

COST ALLOCATION: Methods, Principles, Applications

edited by

H. Peyton YOUNG

School of Public Affairs
University of Maryland
College Park, Maryland 20742
U.S.A.

formerly at the
International Institute for
Applied Systems Analysis

1985

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD

ISBN: 0 444 87863 7

Publishers

ELSEVIERS SCIENCE PUBLISHERS B.V.
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

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

ELSEVIERS SCIENCE PUBLISHING COMPANY, INC.
52 Vanderbilt Avenue
New York, N.Y. 10017
U.S.A.

Library of Congress Cataloging-in-Publication Data
Main entry under title:

Cost allocation.

Outgrowth of a meeting on cost allocation, held by the International Institute for Applied Systems Analysis at Laxenburg, Austria, in August 1983.
Bibliography: p.
1. Cost accounting--Congresses. I. Young, H. Peyton, 1945- . II. International Institute for Applied Systems Analysis.
HF5686.C8C698 1985 657'.42 85-20911
ISBN 0-444-87863-7 (U.S.)

PRINTED IN THE NETHERLANDS

THE INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS

is a nongovernmental research institution, bringing together scientists from around the world to work on problems of common concern. Situated in Laxenburg, Austria, IIASA was founded in October 1972 by the academies of science and equivalent organizations of twelve countries. Its founders gave IIASA a unique position outside national, disciplinary, and institutional boundaries so that it might take the broadest possible view in pursuing its objectives:

To promote international cooperation in solving problems arising from social, economic, technological, and environmental change

To create a network of institutions in the national member organization countries and elsewhere for joint scientific research

To develop and formalize systems analysis and the sciences contributing to it, and promote the use of analytical techniques needed to evaluate and address complex problems

To inform policy advisors and decision makers about the potential application of the Institute's work to such problems

The Institute now has national member organizations in the following countries:

Austria
The Austrian Academy of Sciences

Bulgaria
The National Committee for Applied Systems Analysis and Management

Canada
The Canadian Committee for IIASA

Czechoslovakia
The Committee for IIASA of the Czechoslovak Socialist Republic

Finland
The Finnish Committee for IIASA

France
The French Association for the Development of Systems Analysis

German Democratic Republic
The Academy of Sciences of the German Democratic Republic

Federal Republic of Germany
Association for the Advancement of IIASA

Hungary
The Hungarian Committee for Applied Systems Analysis

Italy
The National Research Council

Japan
The Japan Committee for IIASA

Netherlands
The Foundation IIASA–Netherlands

Poland
The Polish Academy of Sciences

Sweden
The Swedish Council for Planning and Coordination of Research

Union of Soviet Socialist Republics
The Academy of Sciences of the Union of Soviet Socialist Republics

United States of America
The American Academy of Arts and Sciences

PREFACE

How should the common costs of an enterprise be shared "fairly" among its beneficiaries? This problem is widespread, both in public enterprises and also within private firms. It arises in the pricing policies of public utilities providing telephone services, electricity, water, and transport. It occurs in the cost–benefit analyses of public works projects designed to serve different constituencies, such as a multipurpose reservoir. It is implicit in the determination of access fees or user charges for common facilities such as airports or waterways. In private corporations it occurs in the form of internal accounting schemes to allocate common and overhead costs among different divisions of the firm. It even crops up in such routine matters as how to allocate travel expenses among the sponsors of a business trip, or how to share the expenses of running a club among its members.

Such *cost allocation* problems typically exhibit two features:

(i) costs must be allocated exactly, with no profit or deficit;
(ii) there is no objective basis at hand for attributing costs directly to specific products or services.

The goal of analysis is to devise criteria and methods for solving these problems in a just, equitable, fair, and reasonable manner. Cost allocation is thus ultimately concerned with fairness. The methods and principles of cost allocation that are likely to find acceptance must somehow be grounded in primitive, common-sense ideas of fairness and equity.

But precisely what is meant by the word fair? According to Webster (1981), it stems from ***fagar***, an Old High German term meaning "beautiful". ***Fair*** means, firstly, "attractive in appearance: pleasant to view". Significantly, a secondary meaning is "pleasing to hear: inspiring hope or confidence often delusively ... specious". It is closely connected to such ideas as ***just, equitable, impartial, unbiased, objective***. "Fair ... implies a disposition in a person or group to achieve a fitting and right balance of claims or considerations that is free from undue favoritism even to oneself ... *Just* stresses, more than ***fair***, a disposition to conform with, or conformity with the standard of what is right, true, or lawful, despite strong, especially personal, influences tending to subvert that conformity" (Id).

Fairness involves a balancing of demands, equal treatment, a concern with legitimacy, lack of coercion. In effect, it provides a basis for voluntary consent. Indeed, one could propose the following empirical test of a "fairness principle": is it sufficiently compelling to cause parties with diverse interests to *voluntarily* agree to its application? Whether the reader finds the cost allocation principles proposed in subsequent chapters compelling will of course depend to some degree on taste, for fairness is partly in the eye of the beholder.

Two approaches to achieving fair allocations may be distinguished. One is strictly normative: all of the objective data are at hand and the problem is to devise an appropriate formula for making an allocation based on these data. Such techniques are typically encountered in cost–benefit analysis or internal accounting schemes in corporations. The second approach is to design a procedure – e.g., a court trial, an arbitration rule, an auction, or a competitive market – that seems fair and impartial *a priori*, and by its functioning produces an allocation (which might or might not seem fair *a posteriori*). Most of the chapters in this volume deal with normative criteria and solutions, but some are concerned with procedural allocation mechanisms. In both cases game theory plays an important role: cooperative game theory in the design of normative formulas, and noncooperative game theory in the design of procedures.

Despite its numerous practical applications, there has until recently been relatively little theory, and even less empirical observation, about how cost allocations are, or ought to be, accomplished. Traditionally the subject has been regarded by economists and accountants as at best a necessary evil. In the economics literature one solution is prominent – "Ramsey pricing" (Ramsey, 1927; Baumol and Bradford, 1970) – but, despite over 50 years of writing on the subject by economists, Ramsey pricing has not found general acceptance by rate-makers and cost–benefit analysts. Similarly, the accounting literature has, until recently, given the problem short shrift. It is even claimed by some that cost allocation is an irrational activity that should be indulged in rarely if at all (Thomas, 1974). (For a more balanced range of views within the accounting profession, see Moriarity, 1981, and also Chapter 2 below.)

The aim of this volume is to explore three issues from diverse points of view. First, what methods are in current use by firms, public utilities, government agencies, accountants, and managers of public facilities? Second, are there general principles of allocative fairness that are supported by observation, common sense, and logic? Third, can theory characterize different methods in terms of their fairness properties and identify new ones?

The volume is divided into two parts. Part I deals with the mathematical formulation of cost allocation problems, the definition of solution concepts, and their characterization axiomatically by general principles of equity. Part II deals with applications to airports, reservoirs, telephone networks, and public rate-making in general. The first part is fairly technical mathematically, while the second is more accessible to the general reader.

Chapter 1 focuses on game-theoretic methods of cost allocation. Four cases are used to illustrate how cost allocation problems can be modeled as cooperative games: sharing municipal water costs, imputing costs to purposes in multipurpose reservoirs, setting airport landing fees, and allocating business trip expenses among sponsors. The most important allocation methods are defined, including the Shapley value, the core, the nucleolus, the separable costs remaining benefits method, Aumann–Shapley pricing, and Ramsey pricing. The *pros* and *cons* of these methods are weighed from a normative standpoint, and then certain *principles* of allocative fairness distilled that axiomatically characterize particular methods. The argument suggests that the Shapley value (or Aumann–Shapley pricing) and the nucleolus are the most robust and satisfactory methods. Which is most appropriate depends on the structure of the problem at hand.

Chapter 2, by Gary Biddle and Richard Steinberg, gives a scholarly survey and critique of the literature on cost allocation in the firm. They observe that objectivity and verifiability of the data on which allocations are based is of paramount importance to accountants, who can be held legally liable for the accuracy of financial statements. However, firms also employ internal allocation schemes for purposes of accounting and control that are subject to less stringent accounting standards. Biddle and Steinberg cite a 1981 study which found that 80 percent of the firms surveyed allocated some portion of the costs associated with corporate management and service departments among the firm's constituent divisions. They describe several methods that figure prominently in the accounting literature, including those of Moriarity, Louderback, Balachandran–Ramakrishnan, and Gangolly. The core and the Shapley value are also analyzed from an accounting perspective. They conclude that, while cost allocation forms an integral part of managerial incentive structures and financial accounting systems in firms, the methods advanced in the theoretical literature have not had much practical impact as yet on accounting practices.

In Chapter 3, Mirman, Tauman, and Zang treat the problem of allocating the joint costs of production among the outputs of a firm. They demonstrate why the Aumann–Shapley pricing mechanism is particularly desirable from a normative standpoint. They explain the mechanics of Aumann–Shapley pricing using simple examples, and show the difficulties inherent in other approaches, such as modified forms of marginal cost pricing. They also point out that Aumann–Shapley prices can be computed relatively easily when the firm's cost function is estimated by linear programming techniques. Finally, they extend their axiomatic framework to handle the problem of allocating *fixed* costs of production among different product lines.

Martin Shubik, who in the early 1960s originated the idea of applying game-theoretic ideas to cost accounting, raises in Chapter 4 a number of important questions about the modeling of cost allocation problems. He points out both the advantages and the pitfalls of the characteristic function as a modeling device: while economical, it ignores crucial distinctions in the information available to the parties, the specifics of coalition formation, the sequencing of moves, and the possibility of threats. In some cases it may be more appropriate to model the problem by a strategic or extensive form game, although there are often difficulties in specifying such models. Shubik suggests a novel approach to bounding the set of reasonable allocations in a cooperative game by defining the notion of an upper and a lower characteristic function, based on the partition function form. The central point, however, is that cost allocation is an exercise in modeling; thus the full diversity of interested parties, their purposes, and the goals of allocation must be recognized before meaningful solutions can be obtained. He predicts that accounting for the combinatorics of joint costs and revenues may prove to be the next major breakthrough in accounting – analogous to the development of input–output and national income accounts – and, as in those cases, the potential usefulness of the theory will stimulate the collection of necessary data to carry out the calculations.

In Chapter 5 Ted Groves addresses the issue of how cost allocation in the firm may affect the reliability of information reported to central management by decentralized divisions. Groves treats the case of a firm providing some common good or service (e.g., computing facilities) for the nonexclusive use of its autonomous divisions, the full cost of which is allocated by a prescribed method amoung the various divisions. It is assumed management does not know the true demands of the divisions for the service. He shows the impossibility of concocting a method that allocates costs exactly *and* implements the optimal level of service by inducing the divisions to report their true demands. In other words, no full cost allocation scheme is both efficient and incentive-compatible.

This result stands in contrast to a somewhat different situation portrayed in Chapter 1 (Section 6). Suppose that management can designate specific amounts of the service for the exclusive use of each division. In this case a mechanism can be designed in which divisions bid the amounts they are willing to pay for the designated levels of service. It can be shown that a noncooperative equilibrium exists that results in an efficient level of service and covers all costs. Moreover, if the cost game has a nonempty core, then full costs can be allocated exactly.

The concluding chapter of Part I, by Terje Lensberg, views cost allocation within the broader framework of economic welfare. He poses the following question: under what circumstances is the decentralized allocation of costs and benefits within public enterprises and firms "consistent" with an allocation that is fair for society as a whole? In other words, what characterizes societal allocations that have the property that they seem fair when viewed by any *subgroup* of society (assuming the others' allocations are fixed)? This "consistency" or "stability" principle (together with several regularity properties) implies that the allocation must maximize some additively separable social welfare function on the space of feasible alternatives. In other words, in order for an allocation to be locally stable, it must meet some global optimization criterion. Commonly advocated social welfare functions of this type include classical utilitarianism, the Nash social welfare function, and a refinement (due to Sen) of Rawls's maximin criterion. Within this framework, Lensberg provides new axiomatizations of particular allocation rules, such as those of Nash and Rawls, and a normative framework for treating decentralized allocation problems in relation to global concepts of social welfare.

The first chapter of Part II represents a scholarly, fascinating exploration by Edward Zajac into the public's perception of what constitutes justice and fairness in economic allocation. Drawing on his experience with the outcomes of public utility rate cases in the United States, combined with examples culled from economic history, politics, and the law, Zajac formulates six Propositions of perceived economic *injustice* – on the theory that examples of injustice are easier to pinpoint than justice *per se*. He finds that these Propositions not only conflict in some cases with economists' notions of economic efficiency, but to some extent with each other. Zajac's point is that justice (or the lack of justice) means balancing conflicting principles, the quest for economic efficiency being only one ingredient in achieving that balance.

Chapters 8 and 9 focus on the specific issue of allocating costs and setting prices in the telecommunications industry. In Chapter 8, William Sharkey summarizes the economic, technological, and regulatory aspects of telecommunications in the United States. He then shows how concepts from cooperative game theory can be used to clarify several important issues, including cross-subsidization between markets, and inefficiencies that result from the fragmentation of markets by the competitive entry of other firms. He also discusses the special structure of cost allocation problems on fixed networks.

Chapter 9, by Nicolas Curien, summarizes recent French telecommunications pricing policy. Using estimates of marginal costs as a reference point, he describes a methodology for computing the amount of cross-subsidization between different services such as long-distance traffic, local traffic, and network access, and between different classes of users such as businesses, households, and coin telephones. He concludes, based on recent French data, that local traffic tends to subsidize long-distance traffic, that general traffic tends to subsidize network access, and that businesses subsidize household and coin telephone use. He also draws attention to cross-subsidizations between peak and off-peak users, between urban and rural customers, and between different uses of the same network such as voice and data transmission. He concludes by pointing out that the goals of economic efficiency and subsidy-free pricing may be outweighed by other political, economic, and equity considerations in setting French pricing policy.

In Chapter 10, Michel Balinski and Francis Sand describe a procedure for allocating landing rights at congested airports. The number of scheduled landings or take-offs in a particular hourly period (called "slots") are limited by the capacity of the airport. In the United States, the Federal Aviation Administration (FAA) establishes quotas on the number of these slots at certain airports, and parcels them out among the different airlines by "scheduling committees" composed of industry representatives. Balinski and Sand investigate an auctioning procedure for achieving a more economically efficient allocation of these slots. In this method, repeated bidding for slots occurs simultaneously across different markets, which allows airlines to take into account the complex interdependence among the slots that they require for scheduling flights. It thus represents an approach to allocating public rights of access by competitive bidding.

In the final chapter, Norio Okada describes in detail how the "separable costs remaining benefits method" is employed in Japan to attribute costs to the different purposes served by multipurpose reservoirs: flood control, irrigation, power, and industrial and municipal supplies. Using Japanese engineering data from the Sameura Dam project, which was built between 1963 and 1970, he gives a detailed account of the cost allocation process, including all of the codicils and *ad hoc* assumptions needed to make the analysis complete. It thus forms a sobering antidote to the earlier theory, which often takes the problem of estimating the relevant data for granted.

This diverse compendium suggests that theory has moved rapidly in recent years, but is still far in advance of actual practice. Indeed the challenge to theory is to become both simple and robust enough to be practical. For this effort to be successful more careful experimentation and observation in conjunction with theorizing seems to be called for. If this volume serves no other purpose, it is hoped that it may stimulate further research of this type.

H. Peyton Young
Washington, D.C.

References

Baumol, W.J. and Bradford, D. (1970) Optimal departures from marginal cost pricing. *American Economic Review* 60: 265–283.

Moriarity, S. (Ed.) (1981) *Joint Cost Allocations* (Tulsa, OK: University of Oklahoma Press).

Ramsey, F. (1927) A contribution to the theory of taxation. *Economic Journal* 37: 47–61.

Thomas, A.L. (1974) *The Allocation Problem: Part II*. Studies in Accounting Research No. 9, National Accounting Association.

Webster's Third New International Dictionary of the English Language Unabridged (1981) (Springfield, MA: Merriam).

ACKNOWLEDGMENTS

This volume grew out of a meeting on Cost Allocation that was held by the International Institute for Applied Systems Analysis at Laxenburg, Austria, in August 1983. The editor gratefully acknowledges the help of the IIASA staff in making the meeting and this book possible. Special thanks are due to Jan Moffat and Bonnie Riley for preparing the manuscript, to Robert Duis for supervising the editing, to Claudia Staindl for transcribing the discussion, and to Elisabeth Both for so graciously arranging the details of the meeting itself. The editor is also grateful to the contributing authors who assisted in reviewing and proofreading the articles.

Finally, special mention should be made of the important role that IIASA itself played over the past dozen years in promoting scientific collaboration across cultural and disciplinary divides. Hopefully some of that spirit is reflected in these pages.

H. Peyton Young
Washington, D.C.
May, 1985

CONTENTS

PART I THEORY

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 1

METHODS AND PRINCIPLES OF COST ALLOCATION

H. Peyton Young

1 Introduction

A central problem in planning the provision of goods or services by a public enterprise is determining a "fair" or "just" allocation of the common costs of production. Examples include setting fees for the use of a common facility such as an airport, a transit system, a communications network, a canal, or a reservoir. Such enterprises, whether strictly public (like a reservoir), or publicly regulated (like telephones) are generally required to "stand on their own bottoms": prices must be set to exactly cover costs, possibly including a mark-up to cover the costs of capital. The economic ideal – marginal cost pricing – does not have this property except in very special cases.

In practice, agencies and firms use simple costing formulas and criteria to solve such problems. A case in point is the "separable costs remaining benefits" method employed by water resource planning agencies to cost out multipurpose reservoir projects. This method is defined in Section 3. Other less sophisticated devices include allocating costs in proportion to some criterion such as number of customers, revenues, profits, or usage rates if the latter can be unambiguously defined. The defect with these methods is that they almost completely ignore the problem of motivation: why, for example, should agents accept an allocation that exceeds their opportunity costs or willingness to pay?

The aim of this chapter is to survey cost allocation methods from an axiomatic perspective. What are the essential principles and properties that characterize different methods? In answering this query we must bear in mind that principles which seem compelling in one context may not be so in another. There is no "method for all seasons". Nevertheless, the axiomatic method does narrow down the plethora of possibilities to a handful of reasonable choices: the Shapley value, the core and certain particular core solutions like the nucleolus, Ramsey prices when demands are known, and outcomes of demand revelation mechanisms when demands are not known. General context, the level of information, and – not least – precedent, all play a role in determining which principles and methods seem most apt for a particular problem.

2 Problem Formulation and Examples

Many of the salient features of cost allocation can be captured in the following simple format. Let $N = \{1, 2, \ldots, n\}$ represent a set of potential customers of a public service or public facility. Each customer will either be served at some targeted level or not served at all. In other words, a customer $i \in N$ will either get a telephone or not, take a train ride or not, hook up to the local water supply or not. The problem is to determine how much to charge for the service, based on the costs of providing it.

The cost data are summarized by a *joint cost function* $c(S)$, which is defined for all subsets $S \subset N$ of potential customers. $c(S)$ represents the least cost of serving the customers in S by the most efficient means. The cost of serving no one is assumed to be zero: $c(\phi) = 0$. c is called the *characteristic function* of a *cost game*.

A *cost allocation method* is a function φ defined for all N and all joint cost functions c on N such that

$$\varphi(c) = (x_1, \ldots, x_n) \in R^N \text{ and } \sum_N x_i = c(N) \quad , \tag{1}$$

where x_i is the charge assessed customer i.

Example 1: Multipurpose reservoirs. One of the most perplexing examples of joint cost allocation is the multipurpose reservoir. Suppose that a dam on a river is planned to serve several different regional interests, such as flood control, hydro-electric power, navigation, irrigation, and municipal supply. The dam can be built to different heights, depending on which purposes are to be included. The cost function associated with such a problem typically exhibits decreasing marginal costs per acre-foot of water impounded up to some critical height of the dam, after which increasing marginal costs set in due to technological limitations. The water resource planning problem is how to apportion the costs among the different purposes.

This problem has a rich history dating back to the creation of the Tennessee Valley Authority (TVA) in the 1930s (see the historical accounts by Ransmeier, 1942 and Parker, 1943). Certain cost allocation formulas suggested for the TVA system are still in use today (in modified form) by water resource agencies, including the Bureau of Reclamation in the United States Department of the Interior.

Table 1.1 shows a joint cost function, based on actual TVA data, for three "purposes": navigation (n), flood control (f), and power (p). (Ransmeier, 1942, p 329).

Table 1.1 TVA cost data for navigation (n), flood control (f), and power (p) (thousands of dollars).

Coalition	*Cost*
ϕ	0
$\{n\}$	163,520
$\{f\}$	140,826
$\{p\}$	250,096
$\{n,f\}$	301,607
$\{n,p\}$	378,821
$\{f,p\}$	367,370
$\{n,f,p\}$	412,584

Example 2: Municipal cost sharing. The second example is one of investment planning, and is based on actual cost data derived from an engineering study (Young *et al.*, 1982). The Skåne region of southern Sweden consists of 18 municipalities, including the city of Malmö (see Figure 1.1). Each municipality requires a certain minimum supply of water that it can either buy from outside sources, pump from its own wells, or obtain by cooperating with some or all of the other municipalities in a regional system. Each municipality i has a minimum ***alternative cost*** or ***opportunity cost*** of supply $c(i)$, assuming that no agreement with the others is reached. Similarly, each potential coalition S of municipalities has a minimum alternative cost $c(S)$ of supplying just the members of S by the most efficient means available, independently of the others.

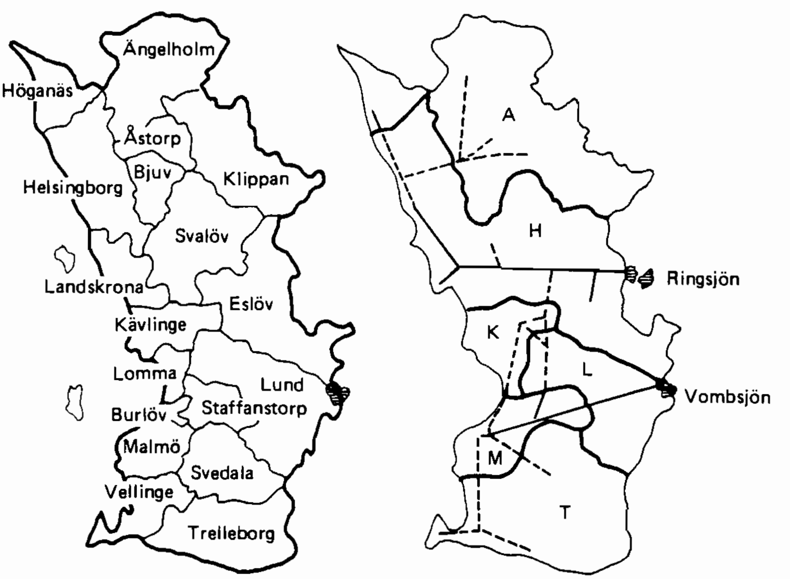

Figure 1.1 Skåne, Sweden, and its division into groups of municipalities.

The value $c(S)$ is determined by engineering considerations, i.e., by estimating the least-cost routing of pipes and pumping stations. $c(S)$ is defined so that it includes the possibility that some or all members of S develop independent on-site sources if this is the least-cost alternative of supplying S. Under these circumstances c will be ***subadditive***:

$$c(S)+c(T) \geqq c(S \cup T) \quad \text{for all disjoint } S, T \quad ,$$

since the cost of serving two disjoint groups includes the possibility of serving them separately. This amounts to a reasonable convention in defining the cost function, but we shall not assume it in the sequel without special notice.

In practice estimating the costs of $2^{18}-1 = 262{,}143$ coalitions is impossible. However, the municipalities fall into natural groupings based on geographical location, existing water transmission systems, and hydrological features that determine the best routes for transmission networks. These conditions can be used to aggregate the 18 municipalities into six units, denoted by A, H, K, L, M, T, as shown in Figure 1.1. The cost function is given in Table 1.2.

Table 1.2 Costs of alternative supply systems (millions of Swedish crowns). Coalitions are separated by commas if there are no economies of scale from their combination.

Group	*Total Cost*	*Group*	*Total Cost*	*Group*	*Total Cost*
A	21.95	AHK	40.74	AHKL	48.95
H	17.08	AHL	43.22	AHKM	60.25
K	10.91	AH,M	55.50	AHK,T	62.72
L	15.88	AH,T	56.67	AHL,M	64.03
M	20.81	A,K,L	48.74	AHL,T	65.20
T	21.98	A,KM	53.40	AH,MT	74.10
		A,K,T	54.84	A,K,LM	63.96
AH	34.69	A,LM	53.05	A,K,L,T	70.72
A,K	32.86	A,L,T	59.81	A,K,MT	72.27
A,L	37.83	A,MT	61.36	A,LMT	73.41
A,M	42.76	HKL	27.26	HKL,M	48.07
A,T	43.93	HKM	42.55	HKL,T	49.24
HK	22.96	HK,T	44.94	HKMT	59.35
HL	25.00	HL,M	45.81	HLMT	64.41
H,M	37.89	HL,T	46.98	KLMT	56.61
H,T	39.06	H,MT	56.49	AHKL,T	70.93
K,L	26.79	K,LM	42.01	AHKLM	69.76
KM	31.45	K,L,T	48.77	AKHMT	77.42
K,T	32.89	K,MT	50.32	AHLMT	83.00
LM	31.10	LMT	51.46	AKLMT	73.97
L,T	37.86			HKLMT	66.46
MT	39.41				
				AHKLMT	83.82

Example 3: Airport landing fees. Landing fee schedules at airports are often established to cover the costs of building and maintaining the runways. The capital cost of a runway is essentially determined by the size of the largest aircraft using it. Suppose that there are m different types of aircraft using an airport and that c_k $(1 \leq k \leq m)$ is the cost of building a runway to accommodate an aircraft of type k. Index the types so that $0 < c_1 < \cdots < c_m$, and let $c_0 = 0$. Let N_k be the set of all aircraft landings of type k in a given year (say n_k in number) and let

$$N = \bigcup_{k=1}^{m} N_k ,\ n = \sum_{k=1}^{m} n_k .$$

Each "player" i in N represents an aircraft using the airport exactly once. The cost game is defined as follows:

$$c(S) = \max \{ c_k : S \cap N_k \neq \phi \} \quad .$$

Table 1.3 gives cost and landing data for Birmingham airport in 1968–69, as reported by Littlechild and Thompson (1977).

Table 1.3 Aircraft landings, runway costs, and charges at Birmingham airport, 1968–69.

Aircraft Type	*Number of Aircraft Landings*	*Annual Capital Cost*	*Charges (Shapley Value)*
	n_i	c_i	x_i
Fokker Friendship 27	42	65,899	4.86
Viscount 800	9,555	76,725	5.66
Hawker Siddeley Trident	288	95,200	10.30
Britannia	303	97,200	10.85
Caravelle VI R	151	97,436	10.92
BAC 111 (500)	1,315	98,142	11.13
Vanguard 953	505	102,496	13.40
Comet 4B	1,128	104,849	15.07
Britannia 300	151	113,322	44.80
Corvair Corronado	112	115,440	60.61
Boeing 707	22	117,676	162.24

Example 4: Travel expenses. An itinerant mathematician is invited on a lecture tour originating in Washington, with stops in New York, Boston, and Chicago. The total fare is $540. The problem is how to allocate the fare "fairly" among the trip's three sponsors. The data of the problem shown below are the alternative costs that would be incurred if the traveler had to make either three separate trips, or one trip involving two of the sponsors and a separate trip for the third, etc.

{NY} + {BOS} + {CHI}	=	120 + 180 + 420	=	720
{NY,BOS} + CHI	=	180 + 420	=	600
{NY} + {BOS,CHI}	=	120 + 510	=	630
BOS + {NY,CHI}	=	180 + 480	=	660
{NY,BOS,CHI}	=	540	=	540

Suppose instead that the traveler gives three lectures in Boston, two in New York, and one in Chicago, the cost data remaining as before. A problem in modeling arises: Are the different *lectures* the primary entities to which costs are assigned? Or is it the sponsors? (Or is it the minutes of lecture time?) The answer depends on the context. If, for example, the lectures are deemed to be the primary objects of allocation, then the cost function would be defined on the set N of all six lectures as follows. $c(i)$ would be $180 if i is a lecture in Boston,

\$120 if i is a lecture in New York and \$420 if it is the lecture in Chicago. $c(i,j)$ would be \$480 if i is in New York and j is in Chicago, and \$120 if both i and j are in New York, and so forth. In short, $c(S)$ is the ***least cost*** of supplying each set S of lectures, and the allocation of costs is made (by whatever method) on a per-lecture basis.

3 Cost Allocation and Cooperative Game Theory

3.1 *The Core*

The foundations of cooperative game theory were laid down in the treatise of Von Neumann and Morgenstern, ***The Theory of Games and Economic Behavior*** (1944). The idea of the "core" of a game, which received only passing mention from these two founding fathers, was later developed by Gillies (1953) and Shapley. Interestingly, the core was foreshadowed in the early literature on cost–benefit analysis. A case in point is the Tennessee Valley Authority Act of 1933 as analyzed by Ransmeier (1942) (see also the excellent review by Straffin and Heaney, 1981). The Act stipulated that the costs of TVA projects be specifically allocated among the purposes involved, the principal ones being navigation, flood control, and power. Ransmeier suggested several criteria for judging cost allocation methods:

> The method should have a reasonable logical basis ... It should not result in charging any objective with a greater investment than would suffice for its development at an alternate single purpose site. Finally, it should not charge any two or more objectives with a greater investment then would suffice for alternate dual or multiple purpose development. (p 220)

In terms of the joint cost function $c(S)$ these requirements state that, if x_i is the charge to purpose i, then in addition to the break-even requirement $\sum_N x_i = c(N)$ the following inequality should hold for every subset S of purposes (including singletons),

$$\sum_S x_i \leqq c(S) \quad . \tag{2}$$

This condition is known as the ***stand-alone cost test***. Its rationale is evident: if cooperation among the parties is to be voluntary, then the calculus of self-interest dictates that no participant – or group of participants – be charged more than their "stand-alone" (opportunity) costs. Otherwise they would have no incentive to agree to the proposed allocation.

A related principle of cost allocation is known as the ***incremental cost test***. It states that no participant should be charged ***less*** than the marginal cost of including him. For example, in Table 1.1, the cost of including n at the margin is

$$c(n,f,p) - c(f,p) = 45{,}214 \quad .$$

In general, the ***incremental*** or ***marginal cost*** of any set S is defined to be $c(N) - c(N-S)$, and the incremental cost test requires that the allocation $\boldsymbol{x} \in R^N$ satisfy

$$\sum_S x_i \geq c(N) - c(N-S) \quad \text{for all } S \subseteq N \quad . \tag{3}$$

Whereas (2) provides incentives for voluntary cooperation, (3) arises from considerations of equity. For, if (3) were violated for some S, then it could be said that the coalition $N-S$ is ***subsidizing*** S. In fact, (2) and (3) are equivalent given the assumption of full cost allocation $\sum_N x_i = c(N)$.

The *core* of c, written Core (c), is the set of all allocations $\boldsymbol{x} \in R^N$ such that (2) [equivalently (3)] holds for all $S \subseteq N$. The core is a closed, compact, convex subset of R^N. Unfortunately, it may be empty, even if c is subadditive.

The core of the TVA cost game is illustrated in Figure 1.2. In this figure, the top vertex x_n represents the situation where all costs are allocated to n; the right-hand vertex the case where all costs are allocated to p, etc. The core is fairly large, reflecting the rapidly decreasing marginal costs of building higher dams. To illustrate other possibilities, let us modify the TVA cost data in one respect: imagine that total costs $c(n,f,p)$ increase to 515,000 due to a cost overrun (the other costs remaining as before). The core of this modified TVA cost game is shown in enlarged form in Figure 1.3.

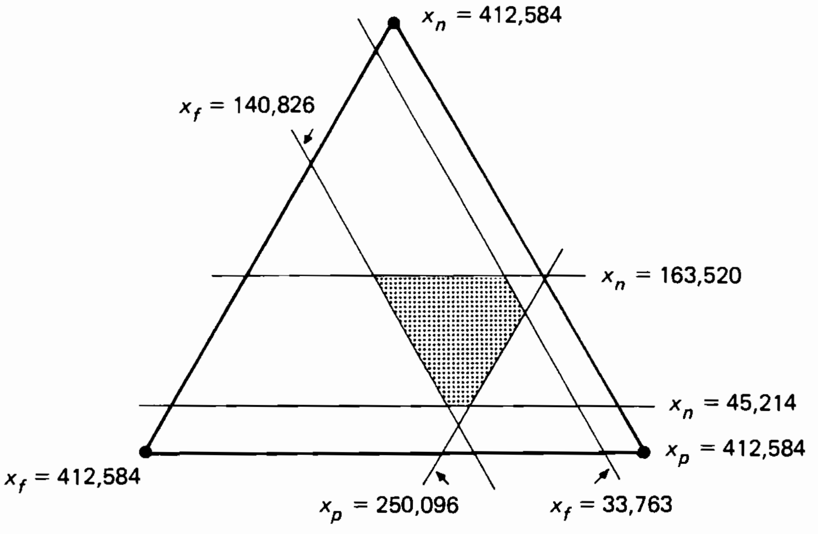

Figure 1.2 The core of the TVA cost game.

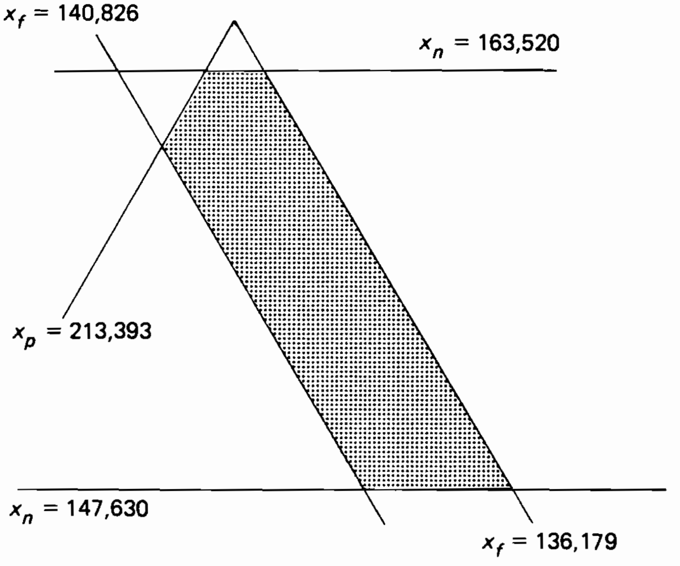

Figure 1.3 The core of the modified TVA cost game (enlarged).

3.2 *Nonemptiness of the core*

When is the core nonempty? It is not enough that c be subadditive. For example, the following 3-person cost game is subadditive but has an empty core.

$$c(1) = c(2) = c(3) = 6$$

$$c(1,2) = c(1,3) = c(2,3) = 7$$

$$c(1,2,3) = 11 \quad .$$

The reason is that the inequalities on the charges

$$x_1 + x_2 \leq 7, \quad x_1 + x_3 \leq 7, \quad x_2 + x_3 \leq 7, \tag{4}$$

when summed, imply that $2(x_1+x_2+x_3) \leq 21$, which contradicts the break-even requirement $x_1+x_2+x_3 = 11$.

A natural but quite strong condition guaranteeing the nonemptiness of the core is that c be *concave* (or *submodular*)

$$c(S \cup T) + c(S \cap T) \leq c(S) + c(T) \quad \text{for all } S, T \subseteq N \quad . \tag{5}$$

For each $i \in N$ and $S \subset N$ define i's *marginal cost contribution relative to* S by

$$c^i(S) = \begin{cases} c(S) - c(S-i) & \text{if } i \in S \\ c(S+i) - c(S) & \text{if } i \notin S \end{cases} \quad . \tag{6}$$

The function $c^i(S)$ is the ***derivative*** of c with respect to i. It may be verified that c is concave if and only if for all i the derivative of c with respect to i is a monotonically decreasing function of S,

$$S \subseteq S' \text{ implies } c^i(S) \geq c^i(S') \quad . \tag{7}$$

Theorem 1 (Shapley, 1971): *If* c is concave then the core of c is nonempty.

3.4 *Costs versus benefits*

In cooperative game theory it is customary to focus on the ***gains*** each coalition of players can realize rather than on the costs directly. Given a cost function c, one way of defining the potential gain of a coalition S is the savings its members can achieve by cooperating instead of going alone:

$$v(S) = \sum_S c(i) - c(S) \quad \text{for all } S \subseteq N \quad . \tag{8}$$

$v(S)$ is called the ***value*** of S, and v is the ***characteristic function*** of the ***cost-savings game***. Note that $v(\phi) = 0$.

In some situations net profits or benefits, rather than costs, are the primary objects of allocation. For example, if $c(S)$ is the cost of a firm producing a subset of outputs $S \subseteq N$, and r_i is the revenue from product $i \in N$, then the net profits from S are given by the characteristic function

$$v(S) = \sum_S r_i - c(S) \quad . \tag{9}$$

The net profitability of different combinations of divisions is a a natural concern for corporate management interested in acquiring or disposing of certain divisions.

Whether costs or benefits are the primary focus of attention depends on the context. An allocation method φ may be applied to any characteristic function (c or v) whether it represents costs or benefits. In principle, of course, it would be desirable if the two approaches give equivalent results. In general, say that two characteristic functions u and v are ***strategically equivalent*** if for some scalar $\alpha \neq 0$ and vector $\boldsymbol{b} \in R^N$, $u = \alpha v + \boldsymbol{b}$; that is

$$u(S) = \alpha v(S) + \sum_S b_i \quad \text{for all } S \subseteq N \quad . \tag{10}$$

An allocation method φ is ***covariant*** if

$$\varphi(\alpha v + \boldsymbol{b}) = \alpha \varphi(v) + \boldsymbol{b} \quad . \tag{11}$$

Most (though not all) of the methods we will discuss are covariant. As a practical matter, in most applications it is the costs that are known. Benefits are often conjectural and subject to manipulation or distortion. Hence the main focus will be on cost functions. Methods for determining demand are treated in Section 6.

4 Methods

4.1 *The separable costs remaining benefits method*

The TVA asserted that its allocation of joint costs was not based on any one mathematical formula, but on ***judgment*** (TVA, 1938). As Ransmeier (1942) wryly observes, "there is little to recommend the pure judgment method for allocation. In many regards it resembles what Professor Lewis has called the 'trance method' of utility valuation" (p 385). Nevertheless, according to Ransmeier, the TVA did in fact use a method and merely "rounded off" the resulting allocations in the light of judgment. This method, called the "alternative justifiable expenditure method", is a variant of an earlier proposal called the "alternate cost avoided method", due to Martin Glaeser, professor of economics at the University of Wisconsin (see Ransmeier, 1942, pp 270–5). It has become, after further refinements, the principal textbook method used by civil engineers to allocate the costs of multipurpose reservoirs, and is known as the "separable costs remaining benefits method" (SCRB); see James and Lee (1971).

The ***separable cost*** of a purpose $i \in N$ is its marginal cost $s_i = c(N)-c(N-i)$. The ***alternate cost*** for i is $c(i)$, and the ***remaining benefit***[1] to i (after deducting its "separable cost") is

$$r_i = c(i)-s_i \quad .$$

The SCRB method assigns costs according to the formula

$$x_i = s_i + \frac{r_i}{\sum_N r_j}\left[c(N)-\sum_N s_j\right] \quad . \tag{12}$$

In other words, each purpose pays its separable cost and the "***nonseparable costs***" $c(N)-\sum_N s_j$ are then allocated in proportion to the remaining benefits. The implicit assumption is that all $r_i \geqq 0$, which is the case if c is subadditive.

For the TVA modified cost data of Figure 1.3, the separable costs, remaining benefits, and corresponding allocation of total costs are shown in Table 1.4.

Table 1.4 The SCRB method applied to the modified TVA data of Figure 1.3.

	n	f	p	*Total*
Alternate cost $c(i)$	163,520	140,826	250,096	554,442
Separable cost s_i	147,630	136,179	213,393	497,202
Remaining benefit, $r_i = c(i)-s_i$	15,890	4,647	36,703	57,240
Allocation of nonsep. cost	4,941	1,445	11,412	17,798
Allocation	152,571	137,624	224,805	515,000

A simple manipulation of (12) reveals that the SCRB method can be given a succinct formulation in terms of the cost-savings game v. For each agent $i \in N$ let $v^i(N) = v(N)-v(N-i)$ represent the marginal cost savings attributable to i. Given the charges $x_1, \ldots, x_n$, let $y_i = c(i)-x_i$ represent the savings imputed to i. The SCRB imputes savings according to the formula

$$y_i = \frac{v^i(N)}{\sum v^j(N)}\, v(N) \quad \text{for all } i \quad . \tag{13}$$

In other words, the SCRB allocates cost savings in proportion to i's marginal contribution to cost savings (see Straffin and Heaney, 1981). This solution has also been proposed for general games v as a means of minimizing players' "propensity to disrupt" the solution (Gately, 1974).

4.2 *The Shapley value*

Imagine that the participants in a cost allocation problem are rational agents who view the outcome as being subject to uncertainty. They might reason about their prospects as follows. Everyone is thought of as "signing up", or committing themselves, in some random order. At each stage of the sign-up the allocation rule is myopic: each player must pay the incremental cost of being included at the moment of signing. The assessments will therefore depend on the particular *order* in which the players join.

Instead of actually proceeding in this way, rational agents might simply evaluate their prospects from the comfort of their armchairs by calculating their expected payoffs from such a scheme. Assume that all orderings are *a priori* equally likely. The "expected" cost assessment for i is then

$$x_i = \sum_{\substack{S \subseteq N \\ i \in S}} \frac{|S-i|!\,|N-S|!}{|N|!} c^i(S) \quad , \tag{14}$$

where $c^i(S)$ is the marginal cost of i relative to S, and the sum is over all subsets S containing i.

Formula (14) is known as the *Shapley value* of the characteristic function c (Shapley, 1953). It may be interpreted as the *average* marginal contribution each player would make to the grand coalition if it were to form one player at a time.

The Shapley value for the revised TVA cost game is:

$$x_n = 151{,}967\text{-}2/3\,,\quad x_f = 134{,}895\text{-}1/6\,,\quad x_p = 228{,}137\text{-}1/6 \quad .$$

This allocation is not in the core (see Figure 1.4) since

$$x_n + x_p = 380{,}104\text{-}5/6 > 378{,}821 = c(n,p) \quad .$$

It can be shown (Shapley, 1971) that if c is concave, then the Shapley value is in Core (c).

4.3 *The nucleolus and its relatives*

If the core conditions are considered of primary importance (as in public utility pricing, for example), the Shapley value may not do. An allocation method φ is a *core* allocation method if $\varphi(c) \in$ Core (c) whenever Core $(c) \neq \phi$. What constitutes a reasonable and consistent way of selecting a unique point from the (nonempty) core of a game?

A standard answer to this question is to select an allocation that makes the least-well-off coalition as well-off as possible. The problem is to agree on a meaning of "well-off". One tack is to say that coalition S is ***better off*** than T, relative to an allocation $\boldsymbol{x}$, if

$$c(S) - \sum_S x_i > c(T) - \sum_T x_i \quad . \tag{15}$$

The quantity $e(\boldsymbol{x}, S) = c(S) - \sum_S x_i$ is called the ***excess of S relative to x***.

To find an allocation $\boldsymbol{x}$ that minimizes the maximum excess $e(\boldsymbol{x}, S)$ over all proper subsets $\varphi \subset S \subset N$ is a problem in linear programming:

$$\max \varepsilon$$
$$\text{subject to } e(\boldsymbol{x}, S) \geqq \varepsilon \text{ for all } S \neq \phi, N \tag{16}$$
$$\text{and } \sum_N x_i = c(N)$$

If there is a unique optimal solution $\boldsymbol{x}^*$ to (16), this is the nucleolus of c. If not, use the following tie-breaking rule. Order the excesses $e(\boldsymbol{x}, S)$, $\phi \subset S \subset N$, from lowest to highest, and denote this $(2^n - 2)$ vector by $e(\boldsymbol{x})$.

The ***nucleolus*** [2] (Schmeidler, 1969) is the vector $\boldsymbol{x}$ that maximizes $e(\boldsymbol{x})$ lexicographically, i.e., for which the value of the smallest excess is as large as possible and is attained on as few sets as possible, the next smallest excess is as large as possible, and is attained on as few sets as possible, etc. It can be demonstrated (Schmeidler, 1969) that there is a unique $\boldsymbol{x}$ that minimizes $e(\boldsymbol{x})$. The proof is based on the observation that if $\boldsymbol{x}'$ and $\boldsymbol{x}''$ maximize $e(\boldsymbol{x})$, then $e[(\boldsymbol{x}' + \boldsymbol{x}'')/2]$ is strictly larger lexicographically than either $e(\boldsymbol{x}')$ or $e(\boldsymbol{x}'')$ unless $\boldsymbol{x}' = \boldsymbol{x}''$.

The idea of the nucleolus is to find a solution in the core that is "central" in the sense of being as far away from the boundaries as possible (see Figure 1.4). There is some arbitrariness in the definition of the metric, however. A reasonable variant of the nucleolus is to define the excess of a coalition on a per capita basis: $e(\boldsymbol{x}, S) = [c(S) - \sum_S x_i] / |S|$. The above construction then defines the ***per capita nucleolus***; (or "normalized" nucleolus; Grotte, 1970). As will be seen in Section 5 below, the nucleolus is probably superior to the per capita nucleolus from an axiomatic perspective.

Solutions for the Swedish municipal cost-sharing game by various methods are compared in Table 1.5.

Table 1.5 Cost allocations for the Swedish municipal cost-sharing game by the SCRB, Shapley value, nucleolus, and per capita nucleolus.

	A	H	K	L	M	T	*Total*
SCRB	19.54	13.28	5.62	10.90	16.66	17.82	83.82
Shapley	20.01	10.71	6.61	10.37	16.94	19.18	83.82
Nuc.	20.35	12.06	5.00	8.61	18.32	19.49	83.82
PC. Nuc.	20.03	12.52	3.94	9.07	18.54	19.71	83.82

5 Principles of Cost Allocation

Faced with a host of competing methods, what basis is there for choosing among them? Each is computationally seductive; each has a certain mathematical charm. The question remains: what are the fundamental ***properties*** that an allocation method should enjoy? We focus on three general types of allocative principles that apply not only to cost sharing but to other allocation problems as well. These are: (1) ***additivity***: if an allocation problem decomposes naturally into subproblems, can their solutions be added?; (2) ***monotonicity***: as the data of the problem change, do solutions change in parallel fashion?; (3) ***consistency***: are solutions invariant when restricted to subgroups of agents? Not surprisingly, these properties cannot all be satisfied simultaneously. Nevertheless, taken in various combinations and strengths, they provide a framework for determining which methods are most appropriate to the situation at hand.

Throughout, several basic properties of an allocation method φ will be taken for granted. These are the break-even (or "efficiency") constraint: $\Sigma\varphi_i(c) = c(N)$ for all c and N. A second and very significant assumption is that all of the data relevant to the allocation problem are contained in the cost function c. This leads to the natural requirement that, if c is symmetric with respect to some two players i and j (i.e., if interchanging i and j leaves c invariant), then $\varphi_i(c) = \varphi_j(c)$. This ***symmetry*** assumption rules out "biased" methods that allocate everything to player number 1, for example. It also rules out methods that allocate on the basis of information contained in c ***and*** some other criterion (such as size, usage rate, etc.). For a discussion of asymmetric solutions, see Shapley (1981).

5.1 *Additivity*

For accounting purposes it is often convenient to assign costs to different "cost categories" such as operations, plant maintenance, interest expense, and marketing. In theory, each such category (denoted $k = 1, \ldots, m$) gives rise to a different joint cost function $c_k(S)$ on the given set N of activities or products. The sum of these cost functions

$$\sum_{k=1}^{m} c_k(S) = c(S)$$

represents the total joint cost function. From an accounting standpoint it would be desirable if the allocation process could be carried out separately for each of the cost categories. The total allocation of costs would be the sum of the cost assignments in each category, and we would like this total allocation to be independent of the particular way the costs are categorized.

A cost allocation method φ is ***additive*** if for any joint cost functions c and c' on N, $\varphi(c+c') = \varphi(c) + \varphi(c')$, where $c+c'$ is defined by $(c+c')(S) = c(S) + c'(S)$ for all $S \subseteq N$.

In any cost game c on N, a player i is a ***dummy*** if i contributes nothing to any coalition; i.e., if $c^i(S) = 0$ for all $S \subseteq N$. The ***dummy axiom*** states that if i is a dummy in c, then $\varphi_i(c) = 0$.

Theorem 2 (Shapley, 1953): *There is a unique allocation method that satisfies the dummy and additivity axioms, namely the Shapley value.*

The proof of this result relies on the following lemma, which shows that any cost function can be decomposed into a linear combination of "primitive" ones. For each nonempty subset $R \subseteq N$ define the *primitive* cost function c_R as follows:

$$c_R(S) = \begin{cases} 1 & \text{if } S \supseteq R \\ 0 & \text{if } S \not\supseteq R \end{cases} .$$

Lemma (Shapley, 1953): *Any cost function c can be expressed as a weighted combination of primitive ones:*

$$c = \sum_{\phi \neq R \subseteq N} w_R c_R \quad . \tag{17}$$

The proof is obtained by letting

$$w_R = \sum_{T \subseteq R} (-1)^{r-t} c(T)$$

and verifying that (17) is an identity.

To prove Theorem 2, let φ be a method that satisfies dummy and additivity. For any nonempty $R \subseteq N$, the dummy axiom implies that $\varphi_i(c_R) = 0$ for all $i \notin R$. Since c_R is symmetric with respect to all $i \in R$ and since $\sum_N \varphi_i(c_R) = 1$, it follows that

$$\varphi_i(c_R) = \begin{cases} 1/\,|R| & \text{if } i \in R \\ 0 & \text{if } i \notin R \end{cases} \quad . \tag{18}$$

In other words, φ is uniquely determined on all primitive cost functions. By (17) and additivity, φ is uniquely determined on all games. Since the Shapley value obviously has the claimed properties, it follows that φ is the Shapley value.

For the airport game (Section 2, Example 3), the Shapley value has a particularly simple interpretation and is easy to compute (Littlechild and Owen, 1973). All aircraft of type k are charged the same amount x_k where

$$x_k = \sum_{l=1}^{k} \left[(c_l - c_{l-1}) \Big/ \sum_{j=l}^{m} n_j \right] \quad .$$

The cost of serving the smallest types of aircraft is divided equally among the aircraft of all types, then the incremental cost of serving the second smallest type is divided equally among all aircraft *except* those of the smallest type, and so forth. The proof follows by observing that for all S,

$$c(S) = \sum_{k=1}^{m} c_k(S) \quad ,$$

where each cost function $c_k(S)$ has the form $c_k(S) = c_k - c_{k-1}$ if S contains a player of type $j \geq k$, and $c_k(S) = 0$ otherwise. The Shapley solution for Birmingham airport is shown in Table 1.3.

5.2 *Monotonicity*

Stated broadly, monotonicity means that if some player's contribution to all coalitions to which he belongs increases or stays fixed, then that player's allocation should not decrease. This concept is relevant to situations in which an allocation is not a one-shot affair, but is reassessed periodically as new information emerges. For example, operations and maintenance costs will typically be allocated annually or even quarterly; investment costs may have to be reassessed *ex post* if there is a cost overrun.

In its most elementary form, monotonicity says that an increase in $c(N)$ alone does not cause any player's allocation to decrease. We say that φ is ***monotonic in the aggregate*** if for all c, $\bar{c}$ and N

$$c(N) \geq \bar{c}(N) \text{ and } c(S) = \bar{c}(S) \quad \text{for all } S \subsetneq N \tag{19}$$
$$\text{implies } \varphi_i(c) \geq \varphi_i(\bar{c}) \quad \text{for all } i \in N \quad .$$

This version of monotonicity was first stated for cooperative games by Megiddo (1974). It is analogous to "house monotonicity" in apportionment (Balinski and Young, 1982) and to monotonicity in bargaining problems (Kalai, 1977).

It is a simple exercise to show that both the Shapley value and the per capita nucleolus are monotonic in the aggregate and that they allocate any change in $c(N)$ *equally* among the players.

Both the SCRB and the nucleolus are egregiously non-monotonic. To illustrate, consider the SCRB allocation of 515,000 in total costs given in Table 1.4. Suppose that total costs increase by 3,000 due to a cost overrun. Then the separable cost of each player increases by 3,000, and the total nonseparable costs decrease by 6,000. The new allocation is $x_n = 153{,}781$, $x_f = 139{,}581$, $x_p = 224{,}636$. Thus the overrun is allocated as $\Delta x_n = 1{,}210$, $\Delta x_f = 1{,}957$, and $\Delta x_p = -169$. In other words, p enjoys a *rebate* from a cost overrun.

The nucleolus suffers from the same defect. In the Swedish municipal cost-sharing game, if total costs increase from 83.82 to 87.82, the nucleolus allocates the cost overrun in the amounts:

A	H	K	L	M	T
0.41	1.19	-0.49	1.19	0.84	0.84

Hence K reaps a gain out of the group's loss.

Monotonicity in the aggregate can be generalized to changes in the value of any coalition. A method φ is *monotonic* if an increase in the cost of a particular coalition implies, *ceteris paribus*, no decrease in the allocation to any member of that coalition. That is, for all c, $\bar{c}$ and $T \subseteq N$,

$$c(T) \geq \bar{c}(T) \text{ and } c(S) = \bar{c}(S) \quad \text{for all } S \neq T \tag{20}$$
$$\text{implies } \varphi_i(c) \geq \varphi_i(\bar{c}) \quad \text{for all } i \in T \quad .$$

It is readily verified that (20) is equivalent to the following definition: φ is monotonic if for all c, $\bar{c}$, and every fixed $i \in N$,

$$c(S) \geq \bar{c}(S) \text{ whenever } i \in S, \text{ and } c(S) = \bar{c}(S) \text{ whenever } i \notin S$$
$$\text{implies } \varphi_i(c) \geq \varphi_i(\bar{c}) \quad .$$

The following "impossibility" theorem shows that monotonicity is fundamentally incompatible with staying in the core.

Theorem 3 (Young, 1985a): *For* $|N| \geq 5$ *there exists no monotonic core allocation method.*

The proof is by example. Consider the cost function c defined on $N = \{1,2,3,4,5\}$ as follows:

$c(S_1)$	=	$c(3,5)$	=	3	$c(S_2)$	=	$c(1,2,3)$	=	3
$c(S_3)$	=	$c(1,3,4)$	=	9	$c(S_4)$	=	$c(2,4,5)$	=	9
$c(S_5)$	=	$c(1,2,4,5)$	=	9	$c(S_6)$	=	$c(1,2,3,4,5)$	=	11

For $S \neq S_1, \ldots, S_5, S_6, \phi$, define $c(S) = \min_{S_k \supset S} c(S_k)$, and let $c(\phi) = 0$.

If $\boldsymbol{x}$ is in the core of c, then

$$\sum_{S_k} x_i \leq c(S_k) \quad \text{for } 1 \leq k \leq 5 \quad ;$$

adding these five relations we find that $3\sum_N x_i \leq 33$ whence $\sum_N x_i \leq 11$. But $\sum_N x_i = 11$ by definition, so all inequalities $\sum_{S_k} x_i \leq c(S_k)$ must be equalities. These have a unique solution, $\boldsymbol{x} = (0,1,2,7,1)$, which constitutes the core of c.

Compare the game $\bar{c}$, which is identical to c except that $\bar{c}(S_5) = \bar{c}(S_6) = 12$. A similar argument shows that the unique core element is now $\bar{\boldsymbol{x}} = (3,0,0,6,3)$. Thus the allocation to both 2 and 4 *decreases* when the value of some of the sets containing them monotonically *increases*. This shows that no core allocation procedure is monotonic for $|N| = 5$, and by extension for $|N| \geq 5$.

Monotonicity refers to monotonic changes in the value of a single coalition, or in the value of coalitions containing a single player. It is more likely in practice that, over a period of time, some coalitions will increase in value and others will decrease. These changes may occur by chance, or they may be under the control of the players themselves.

A classical instance of a cost allocation problem in which the players can manipulate the value of the coalitions is the apportionment of overhead costs among the divisions of a firm (Shubik, 1962). Corporate management may employ an internal cost accounting scheme to provide incentives for divisions to perform more efficiently. Yet non-monotonic allocation methods have just the opposite effect: by providing perverse incentives they reward inefficiency and sloth.

To illustrate, let N now denote a set of divisions (or product lines) in a firm, each represented by a manager. For each $S \subseteq N$, $c(S)$ represents the costs of those divisions doing business without benefit of the others. The numbers under the (partial) control of a division manager i are the costs of the coalitions i belongs to. Suppose that manager i improves efficiency by lowering the costs of *all* sets S of which he is a member. Simultaneously, other product managers may be engaged in decreasing certain of the numbers $c(T)$, or through mismanagement increasing them. This adjustment process yields a new cost function cost c on N. We can unequivocally say that i's *contribution to costs* is lower in $\bar{c}$ than in c if

$$c^i(S) \geqq \bar{c}^{\,i}(S) \quad \text{for all } S \subseteq N \quad . \tag{21}$$

A method φ is *strongly monotonic* (does not create perverse incentives) if whenever (21) holds for c, $\bar{c}$, and i, then $\varphi_i(c) \geqq \varphi_i(\bar{c})$. Strong monotonicity obviously implies monotonicity.

Theorem 4 (Young, 1985a): *The Shapley value is the unique allocation method that is strongly monotonic.*

Note that the Shapley value is not the only ***monotonic*** allocation method, since equal division [$x_i = c(N) / |N|$ for all i] has this property and is additive too.

5.3 *Consistency*

If attention is restricted to concave cost functions, then the Shapley value is in the core, is additive, and strongly monotonic. In short, it is a uniquely desirable method for such problems. If we consider cost functions that are not concave and the allocation problem is one-shot, it may be more desirable to choose a method that guarantees a solution in the core whenever the core is nonempty.

The core has important applications to the pricing policies of regulated monopolies (Faulhaber, 1975; Zajac, 1978; Sharkey, 1982). Let $(r_1, r_2, \ldots, r_n)$ denote the revenues from products $\{1,2, \ldots, n\} = N$ of a public utility. If these violate the core conditions for some subset S of products, $\sum_S r_i > c(S)$, then a competitor might be able to enter the market, undercut the prices of the regulated firm and steal a portion of his business (namely, the products in S) while making a positive profit. Moreover, if c is subadditive, the regulated firm would be left in the unenviable position of producing the set $N-S$ at a *loss*, since $c(N) - c(S) \leqq c(N-S)$ and $\sum_S r_i > c(S)$ imply $\sum_{N-S} r_i < c(N-S)$.

For such problems the nucleolus has a strong claim to being the preferred method. The argument for the nucleolus is based on the following general allocative principle. A globally valid, acceptable, or "fair" allocation should also be seen as valid, acceptable, or "fair" when viewed by any subgroup of the claimants. In effect, no subgroup should want to "re-contract". This concept is known variously in the literature as ***consistency*** or ***stability***. It has been applied to a wide variety of allocation problems, including the apportionment of representation (under the name "uniformity", Balinski and Young, 1982); taxation rules, Young, 1985c), bankruptcy allocation methods (Aumann and Maschler, 1985), cooperative games (Sobolev, 1975), surplus sharing (Moulin, 1985), and bargaining problems (Harsanyi, 1959; Lensberg, 1982, 1983).

To illustrate the consistency concept for cost games, consider the nucleolus $\bar{x}$ of the revised TVA cost game as pictured in Figure 1.4. Players n and p might reason about their situation as follows. If f's allocation is granted to be $\bar{x}_f = 138{,}502.5$, this leaves a total of 376,497.5 to be divided between n and p. The range of possible divisions (holding $\bar{x}_f$ fixed) is represented by the dotted line segment through $\bar{x}$ labeled L. In effect, L is the core of a smaller (or "reduced") game on the two-player set $\{n, p\}$. The question is: does $\bar{x}_n, \bar{x}_p$ represent a fair division ***relative to this reduced game***? The answer is ***yes*** if we accept the nucleolus as a fair division concept. The reason is that the nucleolus of any two-person game with a nonempty core (which is simply a line segment) is the ***midpoint*** of the core. As shown in Figure 1.4, the nucleolus bisects each line segment through it in which the coordinate of one player is held fixed. In other words, the nucleolus is still the nucleolus when viewed by any pair of players. The same conclusion holds for ***every*** subgroup of players.

This leads to the following general definition. Let c be any cost game on a set N (with or without a core) and let $\boldsymbol{x}$ be an allocation of c, $\sum_N x_i = c(N)$. For any proper subset $T \subseteq N (T \neq \phi)$, the ***reduced cost game for T relative to x*** is the game $c_{T,\boldsymbol{x}}$ defined on all subsets S of T as follows:

$$c_{T,\boldsymbol{x}}(S) = \min_{S' \subseteq N-T} \{ c(S \cup S') - \sum_{S'} x_i \}, \quad \text{for } \phi \subset S \subset T \quad , \tag{22}$$

$$c_{T,\boldsymbol{x}}(T) = \sum_{T} x_i$$

$$c_{T,\boldsymbol{x}}(\phi) = 0 \quad .$$

An allocation method φ is *consistent* if for every set of players N, every cost game c on N, and every proper subset T of N, $\varphi(c) = \bar{\boldsymbol{x}}$ implies $\varphi(c_{T,\bar{\boldsymbol{x}}}) = \bar{\boldsymbol{x}}_T$. Note that this definition applies to all cost games c, whether or not they have a nonempty core.[3] It is a rather remarkable fact (and straightforward to check) that the nucleolus is consistent.

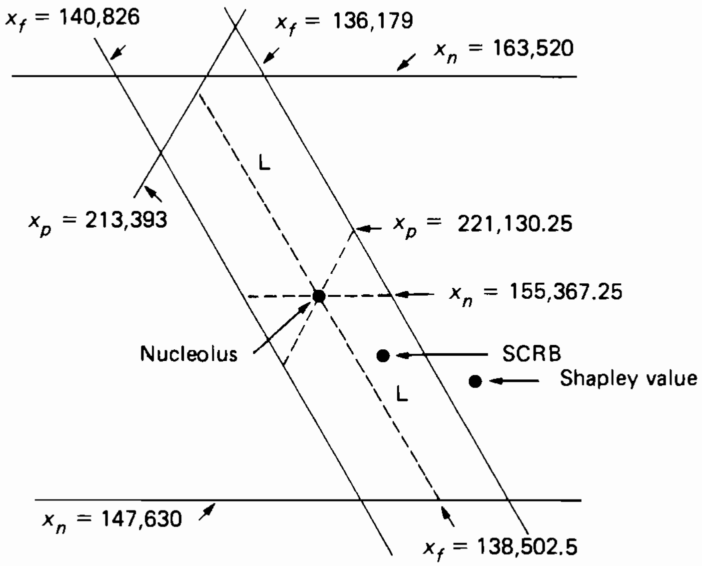

Figure 1.4 Solutions of the TVA modified cost game.

Theorem 5 (Sobolev, 1975): *The nucleolus is the unique allocation method that is covariant and consistent.*

Recall that a method φ is covariant if $\varphi(\alpha c + \boldsymbol{b}) = \alpha\varphi(c) + \boldsymbol{b}$ for every non-zero scalar α and every n-vector $\boldsymbol{b}$. In other words, changing the units in which costs are measured, or translating the baselines from which they are measured, does not materially affect the outcome. The nucleolus is clearly covariant. The property is an important one from a practical standpoint, because there is often a question as to which cost components ought to be included in the joint cost function $c(S)$, and which should be attributed directly to particular players. In the mathematician's lecture tour, for example, one might include in c the costs of local travel (e.g. taxi fares), or attribute them directly to particular lectures and leave them out of c. In the Swedish cost-sharing problem one could either attribute local pumping costs to individual municipalities ***separately*** or else include them in the joint cost function. The covariance property says that, insofar as such costs are additive and separable, the resulting allocations should be equivalent.

6 Economic Efficiency and Demand Revelation

Formulating the cost allocation problem as a cost game sweeps several important issues under the rug. In particular, the cost game says nothing about agents' willingness to pay. If a cost allocation method assigns charges that exceed the benefits received, the agents might refuse to pay. There is a further problem: if benefits are not incorporated into the analysis, how is one to choose the optimal set of customers to serve? In specifying the set N, the cost game takes the optimal set for granted. Assuming that benefits are known, how should they be incorporated into the allocation process?

To answer this question, consider a cost game $c(S)$, $S \subseteq N$. Suppose that each agent $i \in N$ has a known utility of being served, u_i. To keep matters simple, think of u_i as i's willingness to pay for receiving a fixed level of service (e.g., for having a telephone installed). The *net benefit* of serving the set S of customers can be defined as

$$\beta(S) = \sum_S u_i - c(S)$$

An economically ***efficient*** set is one that maximizes $\beta(S)$, say S^*. The ***value*** of a coalition $S \subseteq N$ can be defined as the maximum net benefit obtainable by serving some, or all, of its members:

$$v(S) \stackrel{'}{=} \max_{R \subseteq S} \beta(R) \tag{23}$$

v is called the ***benefit game***. Note that by definition $v(S) \geq v(T) \geq v(\phi) = 0$, whenever $S \supseteq T$; also $v(i) = \max\{u_i - c(i), 0\}$. If benefits are known, v can be computed and an allocation $\boldsymbol{y} = (y_1, \ldots, y_n)$ defined, where $\sum_N y_i = v(N)$. For example, $\boldsymbol{y}$ might be obtained by applying the Shapley value, the SCRB method, the nucleolus, or any other method directly to the game v. For every allocation of benefits $\boldsymbol{y}$, and every choice of efficient set $S^* \subseteq N$, there is a concomitant allocation of costs, namely

$$x_i = \begin{cases} u_i - y_i & \text{if } i \in S^* \\ 0 & \text{if } i \notin S^* \end{cases} \tag{24}$$

Unfortunately, benefits do not have the same aura of reliability and objectivity that costs do. They are too easily subject to manipulation: agents will have an incentive to misreport their true benefits if this strategy results in lower assessed costs.

Suppose, then, that each agent's willingness to pay is known only to himself, but that the cost function is common knowledge. Is there a bidding procedure or incentive system that implements an efficient decision ***and*** allocates costs exactly? Essentially the answer is no. For any general allocation mechanism that exactly covers costs there is always some situation in which it pays some individual to distort the result by sending a false signal.

Nevertheless positive results can be obtained by relaxing these requirements and exploiting the special structure of cost allocation problems. Namely, it is possible to design a straightforward demand revelation mechanism that *covers* costs and selects an economically efficient set S^*. Moreover, for certain important classes of cost functions, one can do this without generating any surplus, i.e., so that assessments exactly equal total costs.

Given a subadditive cost game c, the idea is to let each agent announce his willingness to pay – not necessarily the true one – and to select an "apparently" efficient set on this basis. (If there are ties the auctioneer has discretion in which efficient set to select.)[4] Each member in the selected set pays his bid; the others pay nothing and are not served. This is called the ***demand revelation game***. There always exists a set of bids such that *no* set of agents can deviate from these bids and ***each*** improve his gain. At such an ***equilibrium*** (called a ***strong equilibrium***) the set of bids accepted (i.e., the set of agents served) will be efficient relative to the ***true*** demands, and costs will be covered (though not necessarily exactly). The ***payoff*** to agent i is $y_i = u_i - b_i$ if i is served and b_i is his bid, and $y_i = 0$ if i is not served.

We can illustrate this with an example. Consider again the TVA cost data of Table 1.1, and suppose that the benefit to each of the three purposes is $u_n = 125{,}000$, $u_f = 150{,}000$, $u_p = 275{,}000$. The costs $c(S)$, net benefits $\beta(S)$, and values $v(S)$ are shown in Table 1.6.

The "true" bids $b_n = 125{,}000$, $b_f = 150{,}000$, and $b_p = 275{,}000$ are not an equilibrium: at these bids the set $\{n, f, p\}$ uniquely maximizes bids less costs, hence all bids will be accepted; but each of the players can lower his bid (say by 10,000) and thereby raise his payoff by 10,000, because even at these lower bids the set $\{n, f, p\}$ still uniquely maximizes bids less costs, so all bids will be accepted.

Is there a set of bids, and a selection of bids to accept, from which there is no incentive to move? The answer is yes. For example, suppose that n lowers his bid from 125,000 to 45,214, the others staying put. Then both of the sets $\{n, f, p\}$

Table 1.6 A benefit game using TVA data.

	$c(S)$	$\beta(S)$	$v(S)$
ϕ	0	0	0
$\{n\}$	163,520	−38,520	0
$\{f\}$	140,826	9,174	9,174
$\{p\}$	250,096	24,904	24,904
$\{n, f\}$	301,607	−26,607	9,174
$\{n, p\}$	378,821	21,179	24,904
$\{f, p\}$	367,370	57,630	57,630
$\{n, f, p\}$	412,584	137,416	137,416

and $\{f, p\}$ maximize bids net of costs at 57,630. Next suppose that f lowers his bid from 150,000 to 117,274, and p lowers from 275,000 to 250,096. Then total bids equal total costs. Moreover, no agent can lower his bid further without being excluded and receiving a payoff of zero. If any group of agents should lower their bids, at least one of them will receive a payoff of zero. Hence the bids $b_n = 45{,}214$, $b_f = 117{,}274$, $b_p = 250{,}096$, with all bids accepted, constitutes a strong equilibrium with payoffs

y_n	=	125,000 − 45,214	=	79,786
y_f	=	150,000 − 117,274	=	32,726
y_p	=	275,000 − 250,096	=	24,904

This allocation $\boldsymbol{y}$ is in the core of the benefit game (see Table 1.6) in the sense that $\sum_S y_i \geq v(S)$ for all $S \subseteq N$. This is a special case of the following result.

Theorem 6 (Young, 1980): *Let c be a subadditive cost game on N, and $\boldsymbol{u} \in \mathbf{R}^N$ a vector of benefits. The associated demand revelation game has a strong equilibrium, and for any such equilibrium the set selected is efficient and the total of accepted bids covers total costs. Moreover, if the benefit game v defined as in* (23) *has a nonempty core, then there exist strong equilibria at which the total of accepted bids exactly equals total costs, and the payoff vectors from such equilibria correspond one-to-one with the allocations in Core* (v).

7 Aumann–Shapley and Ramsey Pricing

An alternative formulation of the cost allocation problem familiar to economists involves a firm producing n homogeneous products in quantities $q_1, \ldots, q_n \geq 0$. Costs are given by a *production function* $F: \mathbf{R}_+^n \to \mathbf{R}$, where $F(\boldsymbol{q})$ is the joint cost of producing the "bundle" $\boldsymbol{q} = (q_1, \ldots, q_n)$. Assume that F has continuous first partial derivatives everywhere on the domain $\mathbf{R}_+^n$ (one-sided on the boundary) and that $F(0) = 0$ Given a target level of production $\boldsymbol{q}^* > 0$, the goal is to find *unit prices* $\boldsymbol{p} = (p_1, \ldots, p_n)$ for the n goods such that costs are exactly covered:

$$\Sigma p_i q_i^* = F(\boldsymbol{q}^*) \quad . \tag{25}$$

This is the ***break-even constraint***. By suitably defining F, one can include in this formulation the cost of capital, or a percentage mark-up over cost.

An n-vector $\boldsymbol{p}$ satisfying (25) is a ***solution*** of the ***problem*** $(F, \boldsymbol{q}^*)$. A ***method*** is a function ψ defined on all problems $(F, \boldsymbol{q}^*)$, and all $n \geq 1$, such that $\psi(F, \boldsymbol{q}^*) = \boldsymbol{p}$ is a solution of $(F, \boldsymbol{q}^*)$.

Most of the concepts introduced previously have their analogues in this setting. We briefly mention two of them. F may decompose naturally into a sum of different cost categories

$$F(\boldsymbol{q}) = F_1(\boldsymbol{q}) + F_2(\boldsymbol{q}) \text{ for all } \boldsymbol{q} \geq 0 \quad . \tag{26}$$

In this case it would be desirable if the cost allocation process could be similarly decomposed. ψ is ***additive*** if (26) implies that for all $\boldsymbol{q}^* > 0$,

$$\psi(F, \boldsymbol{q}^*) = \psi(F_1, \boldsymbol{q}^*) + \psi(F_2, \boldsymbol{q}^*) \tag{27}$$

As shown by Billera and Heath (1982) and Mirman and Tauman (1982) (see also Billera, Heath, and Verrecchia, 1981), additivity, together with several additional regularity conditions, uniquely characterizes the following pricing method, which is based on the Aumann–Shapley value for non-atomic games (Aumann and Shapley, 1974):

$$p_i = \int_0^1 [\partial F(t\mathbf{q}^*) / \partial q_i] \, dt \quad \text{for all } i \quad . \tag{28}$$

These *Aumann–Shapley prices* satisfy the break-even constraint, and are additive. Intuitively, p_i represents the average marginal cost of i along the ray from $\mathbf{0}$ to the target $\mathbf{q}^*$.

Aumann–Shapley prices have an important incentive property analogous to strong monotonicity in cost games. Imagine that each product line i is supervised by a division manager. The manager's concern is that his division looks profitable – indeed his bonus may depend upon it. Whether it *is* profitable depends on how costs are imputed to his division, and hence on the cost allocation formula. From the point of view of corporate headquarters, the purpose of an accounting scheme is to encourage product managers to innovate and reduce costs. In short, its interest is in a method that rewards efficiency. How can increases in efficiency be measured with respect to individual products? A natural test is to say that i's *contribution to costs decreases* if $\partial F(\mathbf{q}) / \partial q_i$ decreases for *all* levels of output $\mathbf{q}$.

A method ψ is *strongly monotonic* if for all production functions F, G and for every fixed i,

$$\begin{gathered} \partial F(\mathbf{q}) / \partial q_i \geq \partial G(\mathbf{q}) / \partial q_i \quad \text{for all } \mathbf{q} \geqq \mathbf{0} \\ \text{implies } \psi_i(F, \mathbf{q}^*) \geq \psi_i(G, \mathbf{q}^*) \quad \text{for all } \mathbf{q}^* > \mathbf{0} \quad . \end{gathered} \tag{29}$$

Together with the analog of symmetry, strong monotonicity characterizes the Aumann–Shapley method uniquely (Young, 1985b).

Aumann–Shapley prices may be computed from the piecewise linear production function that arises from solving a cost-minimizing linear program.

Another method of cost allocation, which is popular in the economics literature, is "Ramsey pricing". As above, assume that there are n homogeneous divisible goods produced in the quantities $\mathbf{q} = (q_1, \ldots, q_n)$ and a joint production function $F(\mathbf{q})$ having continuous first partial derivatives on the domain $\mathbf{q} \geq \mathbf{0}$. In Ramsey pricing, the demands for the products are factored into the calculation. To simplify matters, assume that demands for the products are independent: $q_i = u_i(p_i)$ is the amount demanded of product i when prices are set at p_i. u_i is assumed to be strictly monotone decreasing in p_i and continuously differentiable. The target level of production, $\mathbf{q}^*$, is not given *a priori* as in the Aumann-Shapley setting, but is derived from the demand and cost information. The idea is to set prices and production levels so as to maximize benefits net of costs ("net social product"). One measure of net social product is total consumers' surplus over all products, minus the costs of production:

$$\text{maximize } S(\mathbf{q}) = \left[\sum_{i=1}^{n} \int_0^{q_i} u_i^{-1}(t) \, dt \right] - F(\mathbf{q}) \tag{30}$$

There are two constraints on this maximum problem. The first is implicit in the formulation: namely, the same price p_i is charged to all consumers of i (i.e., pricing is *nondiscriminatory*). The second is the break-even requirement

$$\Sigma p_i q_i - F(\mathbf{q}) = 0 \quad . \tag{31}$$

Solving (30) subject to (31) is a standard exercise in constrained optimization. Form the Lagrangian

$$L(q) = S(q) + \lambda [\Sigma q_i u_i^{-1}(q_i) - F(q)] \quad .$$

A necessary condition for the optimal target q is that for all i

$$u_i^{-1}(q_i) - \frac{\partial F}{\partial q_i}(q) + \lambda \left[u_i^{-1}(q_i) + q_i \frac{\partial u_i^{-1}}{\partial q_i}(q_i) - \frac{\partial F}{\partial q_i}(q) \right] = 0 \quad . \tag{32}$$

Let $p_i = u_i^{-1}(q_i)$, $c_i = \partial F / \partial q_i(q)$, which is the marginal cost of producing i, and $\eta_i = (p_i / q_i)(\partial q_i / \partial p_i)$, which is the elasticity of demand for i, and is always negative since u_i is decreasing. Then (32) has the simpler expression

$$(c_i - p_i) / p_i = \frac{\lambda}{1+\lambda} \frac{1}{\eta_i} \quad . \tag{33}$$

In other words, a necessary condition for optimality is that, for each i, the percentage by which marginal cost differs from price is inversely proportional to the elasticity of demand for i.

Prices satisfying (33) are known as *Ramsey prices*, after Frank Ramsey (1927) who applied similar reasoning to determining optimal methods of taxation. They have been discussed extensively in the economics literature (see Manne, 1952; Baumol and Bradford, 1970; Boiteux, 1971). Their essential property, in industries with declining average costs, is that the percentage mark-up over marginal cost is greater the more inelastic the demand for the good is. As W. Arthur Lewis (1949) put it:

> The principle is ... that those who cannot escape must make the largest contribution to indivisible cost, and those to whom the commodity does not matter much may escape. The man who has to cross Dupuit's bridge to see his dying father is mulcted thoroughly; the man who wishes only to see the scenery on the other side gets off lightly. (p 21)

The inherent fairness of such a criterion may strike the reader as debatable. For example, should low-income families who need telephones for emergencies subsidize long-distance usage by businesses? The Ramsey principle might well lead to such a result. Certainly there is no guarantee that Ramsey prices are subsidy-free, i.e., are in the core of a suitably defined cost game. From a practical point of view, Ramsey prices may be difficult to estimate, since they rely on demand elasticities that may not be known. For these and other reasons Ramsey pricing, while interesting in theory, has not thus far found much practical acceptance.

8 Conclusion

There is no shortage of plausible methods for cost allocation. The essence of the problem, however, lies not in defining methods, but in formulating principles and standards that should govern allocations, and then determining which methods satisfy them.

Minimal data of the problem are the total costs to be allocated, and the objects to which costs are to be assigned. In specifying the latter one must be clear about when two cost objects are comparable, i.e., would be assigned equal costs in the absence of other information. The cost objects could be divisions of a firm, dollars of revenue from different products, individual consumers, classes of consumers, stops on a trip, lectures given on a trip ..., etc. Their identity varies greatly from one situation to the next.

The second element of the problem is estimating the cost associated with each subset of cost objects. The specification of these costs for all subsets defines a *cost game* on the cost objects, which are called "players". Computing a cost for each of 2^n subsets of n cost objects is daunting if n is larger than 4 or 5. In practice the structure of the problem often allows simplifications. For example, airport-type cost games, in which costs depend only on certain critical thresholds, are manageable even for very large n. Other situations may allow the grouping of players and allocating costs hierarchically, first among groups and then within each group. (There is no guarantee, however, that the outcome will be the same as if costs had been allocated in one fell swoop.)

A third element of the problem is the anticipated *benefit* from the project. Economic efficiency suggests that the optimal set of cost objects is the one that yields the greatest benefits net of costs. In principle, the allocation of benefits can be carried out using the same methods that apply to the allocation of costs. If demands are not known, however, then serious difficulties arise in trying to design a cost allocation scheme that implements an efficient decision. One reasonable approach is to employ a competitive bidding scheme that generates an efficient decision and at least *covers* all costs.

From a normative standpoint, four general principles stand out as important criteria for judging cost allocation methods. These are: monotonicity, additivity, consistency, and staying in the core. The last two seem most compelling for one-shot investment problems, or in public utility pricing where cross-subsidization is a major issue. In this case the nucleolus seems to be the best choice, although it should be noted that another core solution – the per capita nucleolus – has the advantage of being monotonic in the aggregate, which the nucleolus is not.

An impossibility theorem states that *no* core solution method is fully monotonic, and only the Shapley value is monotonic in the strongest sense. The Shapley value can also be characterized as the unique method that is additive and allocates no costs to "dummy" players. On the other hand, the Shapley value is not necessarily in the core. The Shapley value seems well suited to situations in which costs are allocated in parts, or are reallocated periodically, but it is not satisfactory when core solutions are required.

In sum, there is no all-embracing solution to the cost allocation problem. Which method suits best depends on the context, the computational resources, and the amount of cost and benefit information available.

Acknowledgments

This work was supported in part by the National Science Foundation, Grant SES 8319530. The author wishes to thank M.L. Balinski, H. Moulin, L. Mirman, A. Roth, and W.W. Sharkey for constructive comments on the manuscript.

Notes

1 If the benefit u_i from purpose i is known, then the remaining benefits are defined by $r_i = \min\{u_i, c(i)\} - s_i$.

2 This is sometimes called the "pre-nucleolus", since Schmeidler's definition required in addition that $x_i \leqq c(i)$ for all i. If c is subadditive these constraints are automatically satisfied.

3. The definition of the reduced game (and hence of consistency) is inconsistent *unless* c has a nonempty core and $x \in \text{core}(c)$. The reason is that, by analogy with the definition of $c_{T,x}$ on the other coalitions, we should have

$$C_{T,x}(T) = \min_{S' \subseteq N-T} \{c(T \cup S') - \sum_{S'} x_i\} \quad ,$$

but this is strictly *less* than $\sum_T x_i$ (for some T) unless $x \in \text{core}(c)$. Thus Sobolev's result seems most persuasive when applied only to the class of games c having nonempty cores. It seems reasonable to conjecture that the proof of Theorem 5 can be carried out on this restricted class.

4. The auctioneer is constrained to choose a set that maximizes bids net of costs. In case of ties, some choices may yield an equilibrium, others may not. The claim is that some bids exist, and some way of breaking ties exists, that is a strong equilibrium. Thus in a very limited sense the auctioneer could be considered a player too, although his only function is to break ties efficiently. In practice, ties rarely, if ever, occur; instead some stopping rule is imposed that allows bids to converge. Experiments conducted at IIASA with the six-person Swedish municipal cost-sharing game show that this convergence is remarkably rapid: a solution in the core or nearly in the core was usually found within ten bids.

References

Aumann, R.J. and Maschler, M. (1985) Game theoretic analysis of a bankruptcy problem from the Talmud. *Journal of Economic Theory* (forthcoming).

Aumann, R.J. and Shapley, L.S. (1974) *Values of Non-Atomic Games* (Princeton, NJ: Princeton University Press).

Balinski, M.L. and Young, H.P. (1982) *Fair Representation* (New Haven, CT: Yale University Press).

Baumol, W. and Bradford, D. (1970) Optimal departures from marginal cost pricing. *American Economic Review* 60: 265–83.

Billera, L.J. and Heath, D.C. (1982) Allocation of shared costs: a set of axioms yielding a unique procedure. *Mathematics of Operations Research* 7: 32–39.

Billera, L.J., Heath, D.C. and Verrecchia, R.E. (1981) A unique procedure for allocating common costs from a production process. *Journal of Accounting Research* 19: 185–96.

Boiteux, M. (1971) On the management of public monopolies subject to budgetary constraints. *Journal of Economic Theory* 3: 219–40.

Faulhaber, G. (1975) Cross-subsidization: Pricing in public enterprises. *American Economic Review* 65: 966–77.

Gately, D. (1974) Sharing the gains from regional cooperation: a game theoretic application to planning investment in electric power. *International Economic Review* 15: 195–208.

Gillies, D.B. (1953) *Some theorems on n-person games*. PhD Thesis, Department of Mathematics, Princeton University.

Grotte, J.H. (1970) *Computation of and Observations on the Nucleolus and the Central Games*. MSc Thesis, Cornell University.

Harsanyi, J.C. (1959) A bargaining model for the cooperative n-person game, in A.W. Tucker and D. Luce (Eds) *Contributions to the Theory of Games*, Vol. IV, Annals of Mathematics Studies, No. 40 (Princeton, NJ: Princeton University Press), 325–355.

James, L.D. and Lee, R.R. (1971) *Economics of Water Resources Planning* (New York: McGraw-Hill).

Kalai, E. (1977) Proportional solutions to bargaining situations: interpersonal utility comparisons. *Econometrica* 45: 1623–30.

Lensberg, T. (1982) *Stability and the Nash Solution* (Bergen, Norway: Norwegian School of Economics and Business Administration).

Lensberg, T. (1983) *Stability and Collective Rationality* (Bergen, Norway: Norwegian School of Economics and Business Administration).

Lewis, W.A. (1949) *Overhead Costs* (London: Allen & Unwin)

Littlechild, S.C. and Owen, G. (1973) A simple expression for the Shapley value in a special case. *Management Science* 20: 370–2.

Littlechild, S.C. and Thompson, G.F. (1977) Aircraft landing fees: a game theory approach. *Bell Journal of Economics* 8: 186–204.

Manne, A. (1952) Multi-purpose public enterprises – criteria for pricing. *Economica*, N.S.19: 322–6.

Megiddo, N. (1974) On the nonmonotonicity of the bargaining set, the kernel, and the nucleolus of a game. *SIAM Journal on Applied Mathematics* 27: 355–8.

Mirman, L.J. and Tauman, Y. (1981) Valeur de Shapley et repartition equitable des coûts de production. *Les Cahiers du Seminaire d'Econometrie*, No. 23.

Mirman, L.J. and Tauman, Y. (1982) Demand compatible equitable cost sharing prices. *Mathematics of Operations Research* 7: 40–56.

Moriarity, S. (Ed) (1981) *Joint Cost Allocations* (Norman, Oklahoma: University of Oklahoma Press).

Moulin, H. (1985) Equal or Proportional Division of a Surplus, and Other Methods. Working Paper E85-05-02 (Blacksburg, VA: Virginia Polytechnic Institute and State University).

Parker, T. (1943) Allocation of the Tennessee Valley Authority projects. *Transactions of the American Society of Civil Engineers* 108: 174–87.

Ramsey, F. (1927) A contribution to the theory of taxation. *Economic Journal* 37: 47–61.

Ransmeier, J.S. (1942) *The Tennessee Valley Authority: A Case Study in the Economics of Multiple Purpose Stream Planning* (Nashville, TN: Vanderbilt University Press).

Schmeidler, D. (1969) The nucleolus of a characteristic function game. *SIAM Journal on Applied Mathematics* 17: 1163–70.

Shapley, L.S. (1953) A value for n-person games, in H.W. Kuhn and A.W. Tucker (Eds) *Contributions to the Theory of Games*, Vol. II, Annals of Mathematics Studies No. 28 (Princeton, NJ: Princeton University Press), pp 307–317.

Shapley, L.S. (1971) Cores of convex games. *International Journal of Game Theory* 1: 11–26.

Shapley, L.S. (1981) Discussant's comments: "Equity considerations in traditional full cost allocation practices: an axiomatic perspective", in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press).

Sharkey, W.W. (1982) *The Theory of Natural Monopoly* (London: Cambridge University Press).

Shubik, M. (1962) Incentives, decentralized control, the assignment of joint costs and internal pricing. *Management Science* 8: 325–43.

Sobolev, A.I. (1975) Characterization of the principle of optimality for cooperative games through functional equations, in N.N. Voroby'ev (Ed), *Mathematical Methods in the Social Sciences*,, Vipusk 6, Vilnius, USSR, 92–151.

Straffin, P. and Heaney, J.P. (1981) Game theory and the Tennessee Valley Authority. *International Journal of Game Theory* 10: 35–43.

Tennessee Valley Authority (TVA) (1938) *Allocation of Investment in Norris, Wheeler, and Wilson Projects.* US House of Representatives, Document No 709, Congress Third Session (Washington, DC: US Government Printing Office).

Von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press).

Young, H.P. (1980) *Cost Allocation and Demand Revelation in Public Enterprises.* Working Paper WP-80-130 (Laxenburg, Austria: International Institute for Applied Systems Analysis).

Young, H.P. (1985a) Monotonic solutions of cooperative games. *International Journal of Game Theory* 14.

Young, H.P. (1985b) Producer incentives in cost allocation. *Econometrica* 53.

Young, H.P. (1985c) Taxation and bankruptcy, in H.P. Young (Ed), *Fair Allocation*, Proceedings of Symposia in Applied Mathematics, Vol. 33 (Providence, RI: The American Mathematical Society).

Young, H.P., Okada, N., and Hashimoto, N. (1982) Cost allocation in water resources development. *Water Resources Research* 18: 463–75.

Zajac, E. (1978) *Fairness or Efficiency: An Introduction to Public Utility Pricing* (Cambridge, MA: Ballinger).

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 2

COMMON COST ALLOCATION IN THE FIRM

Gary C. Biddle and Richard Steinberg

1 Introduction

Twenty years ago, Martin Shubik (1962) suggested that the Shapley value of a game be used to allocate accounting costs. While that suggestion has spawned a number of cost allocation proposals based on game-theoretic constructs, these studies comprise only one of several major streams of cost allocation research. Because these various approaches have developed independently, they often employ differing assumptions, definitions, and methodologies. The resulting difficulties in comparing alternative proposals may help to explain their limited acceptance.[1]

This study provides a framework in which the major streams of cost allocation research are evaluated and compared, with a special emphasis on those areas which relate to common cost allocation in the firm. This paper is a condensed version of "Allocations of Joint and Common Costs" by Biddle and Steinberg (1984). Section 2 clarifies cost allocation concepts and explores some initial attempts to explain why cost allocations are made. Section 3 examines proposals for allocating common costs. Section 4 separately examines those common cost allocation proposals which employ game-theoretic concepts. Section 5 suggests some directions for future research.

2 The Practice of Allocating Costs

Allocating costs is perhaps the most ubiquitous of accounting activities. Costs are allocated among firms' divisions, products, and accounting periods for use not only in the preparation of external financial reports, but also in a variety of managerial decisions. Uncomfortable with the *ad hoc* nature of many traditional cost allocation practices, researchers have proposed a number of "improved" cost allocation schemes. Before these proposals can be compared, however, their diversity requires the development of a common set of concepts and definitions.

2.1 *Alternative Cost Concepts*

To an economist, "the ultimate economic cost of any activity is to be interpreted as the alternative opportunities foregone" (Hirschleifer, 1976, p 299). Applied to a traditional comparative statics analysis, this opportunity cost concept implies that the marginal, average variable, and average cost curves include returns to all factors of production including capital (see Turvey, 1969).[2] These curves, which are often used to illustrate optimal output decisions, are actually simplifications of more complex relationships. As Alchian (1971) has observed, economic costs depend not only on output but also on the output rate, the time until the first unit is produced, and the length of the production run. Analyses relating costs and output levels implicitly hold these other factors constant. Economists also employ a distinction between short and long run, which recognizes that for certain decisions, factors like production technology and plant size cannot be altered, giving rise in the short run to related fixed costs. Through these concepts economists have emphasized the importance of identifying relevant costs for a given decision: choosing short-run production levels by equating marginal costs and revenues is a well known example.

Certain practical considerations have led accountants and managers to develop alternative cost concepts. Potential legal liabilities associated with financial statement and tax disclosures have produced an accounting emphasis on objectivity and verifiability. This precludes the use of opportunity costs in accounting statements which are based on historical costs instead. Historical costs are easy to measure since they are the acquisition costs paid by a firm for goods or services in past arm's-length transactions. Historical costs and estimates of future historical costs are also often used in managerial decision-making. The need for an objective basis for managerial performance evaluation and contracting provides one explanation. In addition, historical costs may provide cost-effective approximations of some of the cost concepts recommended by economists. For example, historical direct labor costs and materials costs plus variable overheads are often assigned to production units on a constant per unit basis. When an additional per unit amount representing a fixed manufacturing overhead is added to these variable costs, full absorption costs are obtained. These costs are not unlike the economist's concepts of average variable and average manufacturing costs, respectively.

2.2 *Cost Allocations Defined*

A cost allocation is defined here as the efficient partitioning of a cost among a set of cost objects.[3] Borrowing a more descriptive term from Demski (1981), a cost allocation is required to be "tidy", meaning that all of a cost is allocated, no more and no less. This definition in no way assures that allocated costs will be useful. Usefulness depends jointly on the nature of the cost being allocated, the allocation method selected, and the decisions to be based on the allocated costs.

Generally accepted accounting principles and tax regulations require cost allocations. While their use in managerial decision-making is more controversial, available evidence suggests that many managerial decisions also depend on allocated costs (Hughes and Scheiner, 1980; Thomas, 1980; Fremgen and Liao, 1981). The role of cost allocations in providing information for managerial decisions suggests that the cost allocation process should begin with a determination of the appropriate cost concept for the decision being made. For example, output decisions should depend on marginal rather than average costs. The allocation process is completed by identifying the costs to be allocated and then applying an allocation method.[4] Subsequent sections emphasize the importance of linking the selection of a method with the decisions to be made.[5]

Because jointness is a pervasive cost characteristic, the term joint has been applied to a wide range of cost allocation settings. Jointness occurs because a firm finds it less costly to incur costs relating to two or more cost objects simultaneously than to incur costs individually for each. A distinction is drawn between two major classes of cost allocations based, in part, on the sources of these cost savings. Using terms that have often been used interchangeably, this study distinguishes between joint and common costs.[6]

Joint cost applies to a setting in which production costs are a nonseparable function of the outputs of two or more products. The nonseparability of the cost function and the joint production of the products reflect cost savings due to what Baumol *et al.* (1982) have termed economies of scope. Economies of scope arise when it is less costly to jointly produce (given amounts of) a set of products than to separately produce subsets of the products. While economies of scale for the separate (subsets of) products is sufficient to produce economies of scope, it is not necessary. Economies of scope can exist for a set of products even though there are diseconomies of scale for all subsets. Baumol *et al.* (1982) show that in markets accessible to potential entrants, economies of scope is both a necessary and sufficient condition for the existence of multiproduct firms.[7]

The focus in joint cost settings is the allocation of the joint production costs to the joint products and the uses (and usefulness) of the allocations in output decisions. The classic example of a joint cost setting is where a packing house allocates the cost of a steer between its beef and hide. This joint production provides obvious economies of scope. In this case, and in similar situations found in agriculture, manufacturing, chemical production and mining, the products are produced in (more or less) fixed proportions. In other joint cost settings, such as petroleum refining, it is possible to vary the product mix (i.e., within some range it is possible to produce more gasoline and less kerosene from each barrel of crude oil).

Common cost applies to a setting in which production costs are defined on a single intermediate product or service that is used by two or more users. Common production is undertaken due to cost savings related to economies of scale.[8] An example is the common provision of computer services to two or more divisions of a multidivision firm. (The term division is used for units ranging from individuals to affiliated companies in decentralized firms.) Because the cost objects in common cost settings are often cost or profit centers assigned to individual managers, allocations of common costs are often the subject of intense discussion and negotiation. In the absence of outside markets for intermediate goods, division and service department managers may have incentives to behave like monopolists or imperfect competitors to the detriment of firm-wide profits. Thus, whereas joint cost allocations emphasize output decision incentives, common cost allocations emphasize incentives to potential users to participate in the common provision of a product or service.[9]

2.3 *Motives for Allocating Costs*

Economists have long considered certain cost allocations unnecessary and possibly misleading since according to their models optimal decisions should depend on unallocated costs.[10] Stigler (1966, p 165), for example, has warned that "***any*** allocation of common costs to the product is irrational if it affects the amount of the product produced, for the firm should produce the product if its price is at least equal to its minimum marginal cost." Because few managers have heeded calls to modify or terminate their cost allocation practices, researchers have recently begun to examine positive explanations for allocations of costs. [11] Zimmerman (1979, p 504), for example, asked "why rational, maximizing individuals would want to allocate costs." He addressed this question by suggesting two settings in which roles for allocations may arise. Drawing on the work of Williamson (1966), Zimmerman first proposed that the imposition of a lump-sum tax in the form of a fixed overhead allocation would cause division managers to reduce spending on perquisites.[12] However, and as Zimmerman observed, this result relies on a wealth effect – the managers consume less of all normal goods because they are, in effect, poorer. There is no reason to expect them to accept such a tax as long as other employment opportunities are available.[13] Moreover, the effect of Zimmerman's tax in no way depends on there being a cost to allocate. Nor is it linked to the size of the cost (see Baiman, 1981). Zimmerman did not explore the possibility that allocations of variable costs might be used to reduce perquisite consumption through their effects on marginal wage rates.[14]

A second setting examined by Zimmerman was the use of fixed cost allocations in providing long-distance telephone services [a Wide-Area Telephone Service (WATS) line] for a university's faculty members. Zimmerman observed that one justification for charging some allocation of the fixed lump-sum costs is that these allocations will proxy for the degraded service, delay, and capacity costs associated with the use of a common resource by users. A more formal approach to this problem is presented in Billera and Heath (1982).

Demski (1981) specifically identified three cost allocation motives: decomposition, motivation, and coordination. He then asked whether the cost allocation mechanisms derived from these motives could be represented as game-theoretic solutions. In contrast to normative studies, the desired properties of cost allocations were made endogenous by deriving them from the settings in which demands for allocation arise. This representational approach need not be confined to a consideration of game-theoretic allocations, but could be used to similarly examine the consistency between other cost allocation motives and methods. If a proposed cost allocation method is found consistent with a given motive, then we might feel more sympathetic toward normative arguments advocating its use in settings where that motive is likely to arise. Alternatively, if we are considering methods widely used in practice,a representational strategy may allow us to identify likely motives for this use. (Normative implications are less clear if a method is found inconsistent with a given set of demand settings since there is no guarantee that an exhaustive set of demands has been identified.)

Amershi (1981) examined the demands for allocations which arise from Demski's motivation motive in a Nash-equilibrium owner–manager (principal–agent) relationship. He found that these settings often evoke demands for simple schemes involving proportional allocations of overhead costs. Balachandran ***et al.***

(1982) proposed that a firm's fixed cost allocations may represent capacity costs "intended to apprise the operating divisions of the long run variable costs of their service department usage" (p 2). However, when they examined a situation in which both capacity and usage decisions could be made and where operating departments provided the information used in the capacity decision, they found that standard allocations of fixed costs seldom achieved both short- and long-term efficiency.[15]

Although preliminary, the Zimmerman (1979), Demski (1981), Amershi (1981), and Balachandran *et al.* (1982) studies make important contributions by emphasizing the underlying demands for cost allocations and by suggesting that allocations may represent cost-effective means of addressing certain practical managerial problems. As Dopuch (1981, p 6) has observed, "Too much cost allocation has been carried out in practice to lead us to believe that it is all bad."

3 Allocating Common Costs

Whereas in joint cost settings the fundamental nonseparability is across products, in common cost settings costs are ultimately allocated across rational evaluative decision-makers capable of (semi-) autonomous actions. As a result, the focus is on ensuring their cooperation. This can be illustrated in the context of a firm which can realize scale economies through the common provision of computer services to two or more divisions. Depending on the level of autonomy granted division managers, allocations may be required to induce them to incur fixed investment costs. Alternatively, their participation may depend on comparisons between proposed operating cost allocations and offers from external suppliers. Even when managers are not given the freedom to contract independently, cost allocations may influence their demands for computer services. If another manager is given profit responsibility for the computer department, allocations may affect the supply of services as well as their distribution.[16] In the absence of opportunities to contract with outside suppliers or intervention by upper management, monopolistic or monopsonistic behavior may arise. Obvious incentive problems accompany allocation schemes that rely on information supplied by the division managers.

3.1 *Common Cost Allocation Practices*

The pervasiveness of common cost allocations is documented in two surveys of corporate allocation practices (Mautz and Skousen, 1968; Fremgen and Liao, 1981). Mautz and Skousen report results from a survey by the Financial Executives Research Foundation of allocations of noninventoriable common costs (including research and development, advertising, administrative and financing cost, and taxes) by 412 firms (including 212 of the "Fortune 500"). Of the 346 usable responses, 306 indicated that these common costs were being allocated to divisions. For the 255 firms that also submitted financial disclosures, the average ratio of common costs to sales (7.83%) actually exceeded the average ratio of earnings to sales (6.29%). For 63 of these firms common costs exceeded ten percent of sales. Mautz and Skousen also found a variety of allocation methods being used. Some firms reported using simple allocations based on single allocation bases (e.g., divisional sales, assets, or investments) while others used more complex formulas (e.g., the Massachusetts Formula).[17] Some used partial and others tidy allocations. Some firms reported using different bases for allocating the same costs. More than 73% of the firms reported using more than one allocation method. When they examined in detail the allocation practices of six diversified firms, Mautz and Skousen found that by altering the methods used to allocate common costs they could significantly change measures of divisional performance, including their rank orders.

Fremgen and Liao (1981) report the results of a survey sponsored by the National Association of Accountants of allocations made by 123 large diversified firms (sales greater than $15 million) from Standard and Poor's Register of Corporations. They found that more than 80 percent allocated common costs associated with corporate management and service departments to their divisions. These costs ranged from 0.02 to 44 percent of corporate sales with over 50 percent of the firms indicating common costs to sales ratios in excess of 5 percent. When asked why they allocate common costs, most firms stated that their motive was "to remind profit center managers that [common] costs exist and that profit center earnings must be adequate to cover those costs" (p 61). The second most cited reason was to "fairly reflect each profit center's usage of essential common services." However, few firms were found to allocate fixed and variable cost components separately.

3.2 *Common Cost Allocation Proposals*

Researchers have often been critical of actual allocation practices, arguing that allocations based on activity measures, such as divisional sales, or on investment measures, such as plant assets, lead managers to depart from profit-maximizing decisions. As replacements, they have offered a series of normative allocation proposals. Moriarity (1975), for example, proposes "a new allocation method that eliminates many of the undesirable characteristics of current allocation bases" (p 795).

Moriarity (1975, 1976) observes that common cost settings arise because of cost savings.[18] Thus, "rather than allocate costs directly to a cost object, it should be possible to allocate cost savings as an offset of the cost of obtaining services independently" (p 792). Notationally, Moriarity's approach can be written as:

$$M_i = W_i - \left[\frac{W_i}{\sum_j W_j}\left(\sum_j W_j - (CC + \sum_j I_j)\right)\right] - I_i$$

where

M_i	=	common cost allocated to division i using the Moriarity method,
W_i	=	the minimum cost alternative available to division i for the common good or service
	=	$\min(Y_i, CC + I_i)$,
Y_i	=	cost at which division i could have obtained the common good or service independently,
CC	=	common cost,

The term in square brackets represents an allocation of cost savings in proportion to minimum available costs. The total costs incurred by division i under the Moriarity approach are $M_i + I_i$.

Moriarity justifies his approach by citing five "advantages of the proposed method over existing methods" (1975, p 794). First, the cost object is never charged more than the next best alternative method of providing the product or service. Thus, a division manager is encouraged to utilize a common good or service when its use results in a cost saving for the firm. As it applies to Moriarity's method, notice that this property relates to individual rationality and does not address the possibility that division managers may form coalitions to obtain greater savings than they obtain from the proposed allocation.[19] Second, the allocation process involves a comparison of the cost of providing common products or services with the next best alternative. As a result, the process of making an allocation renders the firm aware of the relative costs of common and independent production. However, this advantage also illuminates the method's information requirements; in order to allocate common costs the firm must obtain estimates of independent provisions for each of its divisions.

The third advantage cited by Moriarity is that every cost object shares in the savings resulting from the decision to incur the common cost. While Moriarity appeals to the "fairness" of this result, fairness in this case must include savings allocations in proportion to minimum independent cost.[20] The fourth advantage is that some cost is allocated to every cost object using the common product or service. This result suggests that no division using a common input is subsidized and is also motivated on fairness grounds. The fifth advantage claimed for the Moriarity method is that it provides an incentive to managers to continue to search for less costly alternative means of obtaining the common product or service. By lowering Y_i or I_i a manager may be able to lower his or her share of a common cost.

The Moriarity method is illustrated in Table 2.1 for a three-division firm with available cost savings of $600. The Moriarity allocations result in each division being assigned a cost (M_i) that is lower than its least-cost alternative (W_i). Thus, no *single* division has an incentive to reject common provision of the intermediate product or service. However, as observed by Louderback (1976), the Moriarity method may allocate a cost to a division that is less than its incremental processing cost (I_i). Notice in Table 2.1 that for division 3 the common cost allocation is negative. This violates the fourth advantage requiring that no division be subsidized. If permitted, divisions 1 and 2 would have an incentive to form a subcoalition excluding division 3.[21]

Louderback (1976) suggested a modification of the Moriarity method that preserves its advantages. It allocates common costs in proportion to an "ability to bear" measure, $Y_i - I_i$:

$$L_i = \frac{Y_i - I_i}{\sum_j (Y_j - I_j)} (CC) \quad .$$

The total costs incurred by division i are therefore $L_i + I_i$ (where the terms are as defined above). To avoid subsidies, Louderback assumed that, for each i, $Y_i \geq I_i$. Based on the information in Table 2.1, the Louderback method results in common cost allocations of $146 to division 1, $250 to division 2, and $104 to division 3 (and total costs of $196, $350, and $854, respectively). Now each division's total cost is more than its incremental processing cost and there are no incentives for divisions or subcoalitions to break away.[22]

Table 2.1 The Moriarity method applied to a three-division firm with common costs of $500.

Division, i	*Independent Costs, Y_i*	*Incremental Processing Costs, I_i*	*Minimum Alternative Costs, W_i*	*Allocated Cost Savings*[a]	*Moriarity Allocation, M_i*	*Total Costs Incurred, $M_i + I_i$*
1	400	50	400	120	230	280
2	700	100	600	180	320	420
3	1 000	750	1 000	300	−50	700
	$2 000	$900	$2 000	$600	$500	$1 400

a $\frac{W_i}{\sum_j W_j}\left[\sum_j W_j - (CC + \sum_j I_j)\right]$

Another variant, suggested by Balachandran and Ramakrishnan (1981), also preserves the advantages cited by Moriarity. It allocates common costs according to

$$BR_i = \frac{W_i - I_i}{\sum_j (W_j - I_j)} (CC) \quad .$$

(and assigns total costs $BR_i + I_i$ to division i). When applied to the information in Table 2.1, this approach allocates common costs of $159 to division 1, $227 to division 2, and $114 to division 3 (with total costs of $209, $327, and $864, respectively).

Gangolly (1981) extended the Moriarity approach to a setting in which not only the grand coalition of all divisions can benefit from the common provision of a good or service, but where subcoalitions are also permitted to form. Thus, in the previous example, divisions 1 and 2 could form a subcoalition excluding division 3. In his Independent Cost Proportional Scheme (ICPS), Gangolly preserves the Moriarity principle of "sharing cost savings in proportion to independent costs" (p 299). Hence, the allocated cost savings (within any coalition) per dollar of independent costs is the same for all divisions. Under the ICPS method, incremental processing costs are separately assigned; "If the estimated cost figures used in the allocation includes additional processing costs, then the use of ICPS allocations (prior to the disaggregation of such costs) can distort allocated costs" (p 309). In addition to the information needed to apply the Moriarity and Louderback methods, the ICPS approach requires common costs for each of the possible subcoalition of divisions. In general, the ICPS approach requires $2^n - n - 2$ additional common cost estimates where n is the number of divisions. Thus, the information requirements increase exponentially rather than linearly in n. For a firm with just two divisions the ICPS method is identical to the Moriarity method.

The Moriarity, Louderback, Balachandrian–Ramakrishnan, and Gangolly methods emphasize incentives for division managers to participate in the common provision of a good or service. Motivational benefits may be obtained even when managers are not allowed sufficient autonomy to contract independently. However, all three methods have severe informational requirements. Independent and incremental processing cost estimates must be obtained for each division as well as a common cost estimate for the grand coalition. The Gangolly approach also requires common cost estimates for all subcoalitions. If asked to supply these estimates, managers may have incentives to misreport. In addition, the Louderback and Balachandran–Ramakrishnan methods may create incentives to *increase* costs – under certain conditions, higher incremental processing costs lead to lower cost allocations.

Another notable feature of these proposals is that the common cost is assumed to be invariant with respect to the allocations. This implies that either the allocations do not influence the usage of the common good or service or, if so, its cost is insensitive to changes in usage. Ignoring wealth effects on demands for capacity, one potential application is the allocation of a fixed common cost. If applied to variable costs, completely inelastic divisional demands must be assumed.[23]

When the common cost varies with usage and divisional demands are not completely inelastic, the amounts of a common good or service supplied to each division and the allocations of its cost must be determined simultaneously. Cohen and Loeb (1982) have suggested that the degree of publicness is an important determinant of the possibility of efficient decentralization in this case. Drawing an analogy from public and private goods,[24] they define a pure common input as a common good or service that has the properties of nonexclusion (i.e., once produced it is not possible to exclude divisions) and nonrivalry (i.e., consumption by one division does not affect the amounts available to others). In contrast, a common good or service that is a pure private input exhibits both excludability and rivalry. Examples are corporate image advertising and a lot-size materials purchase, respectively. Hughes and Scheiner (1980) have shown that for pure private inputs, tidy allocation schemes "will not necessarily produce optimal demand decisions (firm-wide) under a decentralized regime where divisions are allowed to set their demands individually, subject only to the charge formula being announced in advance" (p 84). Although Cohen and Loeb (1982) show that optimal decentralized decisions are *possible* for pure common inputs, free rider problems may hinder this outcome.[25] See Chapter 5 in this volume for related results on cost allocation and incentives.

4 The Game-Theoretic Approach to Common Cost Allocations

In his pioneering paper, Martin Shubik (1962)[26,27] proposed that as a procedure for assigning common costs and revenues (and thus profits) the Shapley value possessed certain "desirable incentive and organization properties". Shubik envisioned a highly decentralized firm with profits to be allocated among semi-autonomous divisions.[28] Each division was viewed as a player of an abstract game with a measure of joint coordinated action given by the characteristic function, a superadditive set function defined on the set of possible coalitions of the various divisions.[29] In general, if S is a set of one or more divisions considered as acting in unison, the characteristic function $v(S)$ describes the profit made by the set S on the assumption that the remaining divisions have closed down (or otherwise left the firm) and the optimum alternative use made of the relinquished resources.

Because the actions taken by one division can affect the profits realized by others, Shubik suggested that an allocation scheme should provide incentives for joint action which would maximize firm-wide profits. In addition, given the level of autonomy assumed by Shubik, each division participating in the joint activity should be allocated profits at least as large as it could earn independently. Shubik observed that the characteristic function can serve as the basis of an allocation scheme which satisfies these requirements. Specifically, he proposed a restatement of the Shapley axioms in the allocation setting, leading to a solution in which each division received a profit allocation equal to its Shapley value.[30]

Shubik's restatement of Shapley's properties was as follows:[31]

Property 1 The profit allocated to a given division depends only upon the various profits which can be earned by all possible combinations of one or more divisions acting in unison.

In other words, the assignment to the divisions is based solely on the characteristic function; all other information is irrelevant. Shapley (1981) calls this the "domain" axiom.[32]

Property 2 The profit allocated to a division depends symmetrically upon all divisions of the firm.

In other words, if two firms are identical except that their divisions are called by different names, then the procedure allocates the same profit to corresponding divisions which are identical except in name. This property is most often referred to as symmetry, although the terms anonymity[33] and fairness[34] are also frequently used.

Property 3 The procedure allocates all profits earned by the firm. This tidiness property has been called efficiency by Shapley (1953)[35] and Pareto optimality by Luce and Raiffa (1957). It forms part of the definition of a cost allocation provided in Section 2.

Property 4 A division whose presence adds nothing to the profits of any coalition should be allocated no profit.

This is the dummy axiom.[36]

Property 5 If two independent allocation problems are combined into one problem, then for each division the profit allocated under the combined allocation is the sum of the allocations under the two individual problems.

This is usually referred to as additivity.[37] The usual interpretation is that various costs can be disaggregated and then allocated separately.[38]

If φ_i represents the profit to be allocated to division i, φ_i is uniquely determined by Properties 1 to 5 to equal the Shapley value:

$$\varphi_i = \frac{(s-1)!(n-s)!}{n!}[v(S) - v(S-\{i\})]$$

where the summation is over the set of all possible coalitions S, where s is the number of divisions in coalition S, n is the total number of divisions, and v is the characteristic function.

This can be seen, as follows, to be the expected marginal contribution of division i, assuming the coalitions form randomly. There are $n!$ orderings of the divisions, but only $(s-1)!(n-s)!$ orderings where division i follows all divisions contained in S and precedes all divisions not in S. Hence the expected marginal contribution of division i is the sum of its marginal contributions to each coalition S, $v(S) - v(S\{i\})$, each weighted by the proportion of the orderings in which division i completes that coalition.[39] Whereas Shubik defined these marginal contributions as profits, they can alternatively be defined as costs.

Shapley's bargaining procedure interpretation should now be clear. A set N of players agrees to play a game with characteristic function v in a grand coalition (the coalition consisting of all the players). Starting with a single member (chosen randomly), the coalition randomly adds, with equal probability, one player at a time, promising each player on his or her admission the amount by which his or her inclusion increases the value of the coalition as determined by the function v. The grand coalition then plays the game and obtains the amount $v(N)$, which is precisely the amount needed to meet all of the promises. The expectation of player i is φ_i, as shown above. Thus, as noted by Aumann and Shapley (1974), the two characterizations of the Shapley value illustrated above are complementary: the deductive, axiomatic approach and the constructive, interpretative approach tend to justify and clarify each other.[40] In addition to the desirable incentive properties identified by Shubik, Jensen (1977) has shown that Shapley allocations satisfy the properties suggested by Moriarity (1975) and Louderback (1976).

As Shapley (1953) observed, his evaluation scheme depends on several "important underlying assumptions" (p 307). One of these assumptions is that the games being evaluated are cooperative affairs in which the players evaluating prospects of their prospective allocations are rational decision-makers.[41] Thus, Shapley allocations most readily apply to common cost rather than joint cost settings. Another underlying assumption is that the characteristic function captures all of the information relevant to a given evaluation or allocation. Finally, although Shapley explicitly intended his value to be an *a priori* evaluation, Verrecchia (1981a) suggests that cost allocation schemes in the real world are often evaluated *ex post facto*.

Proposed applications of the Shapley value include allocations of: aircraft landing fees (Littlechild and Thompson, 1977; Dubey, 1982), the cost of public goods and services (Loehman and Whinston, 1971), water resource costs (Young *et al.*, 1982), and depreciation (Callen, 1978). Verrecchia (1981a,b, 1982) provided a thorough discussion of a dispute among several schemes, including the Shapley value, to allocate costs (and taxes) at the McDonnell Douglas Corporation. However, aside from these examples, little additional evidence exists of actual or proposed Shapley value allocations. In addition, not all of the proposed applications have maintained Shapley's assumption regarding players as evaluative decision makers. Callen, for example, allocated depreciation to alternative investment projects.

4.1 *Variations on Shubik's Approach*

The limited number of applications may reflect the fact that although allocations based on the Shapley value possess (and are uniquely determined by) certain desirable properties, there are also limitations. Several authors have noted that reliance on the characteristic function may ignore other information that is relevant in many allocation settings. For example, Roth and Verrecchia (1979) argued that the characteristic function ignores strategic aspects.[42] They did not find the expected marginal contribution approach to be fully compelling, since in general there is no reason to believe that the formation of coalitions should be regarded as equally likely events. Instead, they observed that allocations among autonomous divisions would depend on bargaining among division managers.

Roth and Verrecchia (1979) also observed that bargaining is a costly activity. To avoid these costs they proposed three assumptions regarding managerial preferences from which the Shapley value emerges as a costless surrogate for the allocation that would be obtained through bargaining. Thus, their idea in a sense combines Shapley's two interpretations of his value – that of a unique evaluative procedure derived from a desiderata of reasonable properties, and that of an allocation procedure obtained through bargaining – with Shubik's managerial application.

Roth and Verrecchia (1979) first assumed that a manager is indifferent between a game in which he or she does not participate and a game in which there are no benefits (costs) to be shared by a division. From this they obtained the dummy axiom. Second, they assumed ordinary risk neutrality (see Roth, 1977) or probabilistic neutrality: each manager is indifferent between being uncertain as to which cost allocation game is played or participating for certain in a game in which the benefit structure is simply the benefit structure implied by the uncertain situation. This implies the additivity axiom. Finally, they assumed strategic risk neutrality: each manager is indifferent between bargaining among r managers for an uncertain outcome or receiving $1/r$ of the benefit for certain. Loosely speaking, this means that managers consider themselves to have equal bargaining ability. This is a symmetry property. Roth and Verrecchia (1979) then showed (with the implicit assumption of the domain axiom and tidiness) that a manager's expected utility for playing in a game equals the Shapley value if and only if his or her preferences obey these assumptions.

As Roth and Verrecchia (1979) observed, their conclusion depends critically on managers' preferences obeying certain assumptions, and that otherwise Shapley values might not yield entirely appropriate proxies for bargained outcomes. Boatsman *et al.* (1981), for example, expressed concern with the third assumption, strategic risk neutrality, doubting that many allocation settings would find players viewing themselves as having equal bargaining positions.[43] However, Roth and Verrecchia (1979) did allow for this possibility with a generalization to the case where a manager is indifferent between bargaining among r managers for an uncertain outcome or receiving some fraction $f(r)$ for the benefit for certain.[44]

Roth and Verrecchia (1979) presented their result as demonstrating the Shapley value to be a fair, equitable, neutral, and costless surrogate. However, as the authors themselves admitted, the concepts of fairness, equity, and neutrality are somewhat broad. It also may be questioned why attention should be focused on only these properties when accounting theory offers many others.

Concerns regarding relative bargaining power have also motivated several other cost allocation proposals. Boatsman *et al.* (1981) suggested that bargaining power may depend on a player's resources rather than on his or her marginal contributions to a set of coalitions. They proposed the use of minimum resource theory, which predicts that players will divide a payoff in proportion to the resources contributed (see also Gamson, 1961). Balachandran and Ramakrishnan (1981) observed that in many organizational settings there are situations in which the cost or value of a resource to a coalition depends on how the other coalitions are aligned. This motivates the use of games in partition function form, as proposed by Thrall and Lucas (1963), and the corresponding Shapley value developed by Myerson (1977). Recently, Shapley (1981) proposed a modified version of his value that is designed to be applicable to situations in which the cost of a joint venture is to be allocated, not necessarily symmetrically, among interested parties. This is accomplished through the use of an exogenously determined vector of importance weights, which are independent of the cost data, and represent the proportions into which the fixed cost of a jointly used facility would be divided.

Balachandran and Ramakrishnan (1981) have argued that an attractive feature of the Shapley value is that it considers all hypothetical coalitions simultaneously and thus, in a sense, is superior to other allocation methods that do not consider all of this information. However, this also means that the calculation of Shapley-based allocations requires enormous quantities of information. The relevant costs (and revenues) applicable to a given allocation must be estimated for every possible coalition of divisions. Yet accounting records typically exist only for those coalitions which have been employed in the past. Furthermore, obvious incentive problems may arise if division managers supply the information on which the estimates are based. The Shapley value also entails formidable computational requirements. For example, the calculation of a Shapley value for a group of only ten divisions would require over a thousand characteristic function values. Littlechild and Owen (1973) and, more generally, Megiddo (1978) have shown that efficient algorithms for the computation of the Shapley value do exist in special cases. Jensen (1977) has also suggested some simplifications in the Shapley formula which reduce computational effort when the number of divisions to consider is large. Aumann and Shapley (1974) give an alternative formula for the Shapley value that is generally somewhat easier computationally. Nevertheless, formidable informational and computational requirements remain.

4.2 *The Core*

Shubik cited as a favorable incentive property of the Shapley approach the fact that the profit allocated to a division is at least as great as the profit the division could earn on its own (individual rationality).[45] Several authors have argued that an allocation scheme should satisfy an even stronger condition (group rationality). It requires that the sum of the profits (costs) allocated to any coalition of divisions be among the set of feasible outcomes that cannot be improved upon by any other coalition; that is, it must be in the *core* of the game.

Hamlen *et al.* (1977) examined the core properties of several cost allocation schemes including: the activity level approach (see Section 2), which is always in the core; the nucleolus,[46] which is always in the core; and the Moriarity scheme, which may or may not be in the core. Balachandran and Ramakrishnan (1981) showed that the Louderback scheme is always in the core, while Boatsman *et al.* (1981) showed in an appendix to their paper that the minimum resource allocation is always in the core.

The core criterion, however, has also been criticized. Hughes and Scheiner (1980) showed that an allocation scheme meeting the core criterion does not necessarily produce optimal firm-wide decisions under a decentralized system where divisions are allowed to set their demands individually (subject only to the allocation formula being announced in advance). Although Boatsman *et al.* (1981) are correct in noting that Hughes and Scheiner inappropriately allowed demands to vary without making the appropriate changes in the characteristic function, the fundamental problem remains: the core property cannot ensure that a (tidy) firm-wide allocation scheme will result in optimum divisional demands without the intervention of a central authority.[47] Nor does the core criterion necessarily imply a unique solution. As the discussion above suggests, several allocations can simultaneously satisfy the core criterion. Another compelling limitation of the core criterion is that in certain cost allocation settings the core may not exist.[48]

The use of the core criterion in connection with the Shapley value is even more problematic. Even when the core exists it might not contain the Shapley value. The superadditive set function on which a game is based implies simply that there are incentives for forming coalitions. It is only when the set function is also *convex*[49] that the core is both guaranteed to exist and to contain the Shapley value (see Shapley, 1965).[50] (Gangolly (1981) found a similar relationship for the Independent Cost Proportional Scheme discussed in Section 2.) Shapley noted that "a convex game is based on a convex set function which means that the incentives for joining a coalition increase as the coalition grows, so that one might expect a 'snowballing' or 'bandwagon' effect." Unfortunately, whereas superadditivity arises naturally, "convexity is another matter".

Hamlen *et al.* (1980) cited the uniqueness of the Shapley value as a limitation on managerial flexibility which may prevent the core criterion from being satisfied. They proposed a modified Shapley value in which the symmetry property is relaxed by assigning weights to the divisions. However, its usefulness may be limited. While Hamlen *et al.* (1980) were able to show that there is always some set of weights that result in an allocation in the core (if it exists), they were unable to produce a rule for assigning such weights for every situation.

5 Conclusions and Directions for Future Research

This study has presented a critical review and synthesis of the cost allocation literature, emphasizing common cost allocation in the firm. Cost definitions and conflicting terminologies have been clarified, allocation proposals critically examined, and allocation practices described. The resulting taxonomy of cost allocation research provides a framework in which allocation procedures can be developed and compared.

A striking aspect of the cost allocation literature to date is its normative tone. Equally striking is the limited impact it has had on cost allocation practices. Foremost among the research areas suggested by this study is a more thorough understanding of the motives for allocating costs. A positive approach is called for in explaining why allocated costs are so extensively employed by firms in internal decision making and external disclosures. As observed in Section 2, cost allocations may represent attempts to approximate economic cost concepts. Thus, a promising line of enquiry is the continued development of (theoretical and empirical) links between the allocations obtained from alternative methods and the appropriate cost concepts for the decisions in which they are employed. Greater attention could also be given to differences between the allocations appropriate for *ex post* and *ex ante* applications. The axiomatic approach offers a promising technique for deriving allocation procedures from previously specified properties. Indeed, as pointed out by Jensen (1977) it may be easier to persuade managers to agree on a set of axioms than to agree on a set of allocation schemes. Recent advances in the axiomatic approach are treated in Chapters 1 and 3, this volume.

While one explanation for the limited acceptance of many allocation proposals is their possible inconsistency with managers' motives for allocating costs, another possible explanation is their computational complexity. Further research could be directed toward the development of heuristic and other approximations to previously proposed allocation schemes. It may be possible, for example, to identify circumstances in which managers could use rules of thumb to approximate more complicated expressions. A positive stream of this research could examine the extent to which current allocation practices already approximate the results obtained from the more sophisticated approaches. In addition, while many of the methods examined in this study require extensive information sets, little attention has been given to the expense and incentive aspects of obtaining these inputs.

Cost allocations form an integral part of most managerial incentive and reward structures and financial accounting systems. Their widespread use attests to their potential impact on resource allocation decisions and to their usefulness. In spite of the extensive literature surveyed in this study, the process of explaining and ultimately improving cost allocation practices has only just begun.

Notes

1. See Kaplan (1981) for a discussion of other possible reasons for the limited interaction between managerial research and practice.
2. That marginal costs should reflect the foregone opportunities of using capital assets is easily illustrated by considering two firms: one that purchases a component from an outside supplier and one that produces the component using its own equipment. In either case, the opportunity cost of the component should reflect capital costs. Turvey (1969) terms allocations of capital replacement costs "depreciation", a meaning which differs from its typical accounting definition.
3. That a cost allocation efficiently partitions a cost does not necessarily mean that the set of costs to which allocation procedures are applied is either mutually exhaustive or mutually exclusive. Efficiency in this case is a mathematical term which simply means that the application of a cost allocation procedure to a given cost results in a tidy partitioning. Our stipulation that allocations be efficient is motivated by our desire to limit the scope of the enquiry. Moreover, we feel that partitions that are not efficient may be more fruitfully examined in the context of transfer pricing. An alternative characterization of a cost allocation procedure is the determination of a set of weights to be assigned to cost objects. From a

programming perspective, the dual of allocating costs to outputs is allocating profits to inputs. Also notice that less emphasis is placed on revenue allocations. As observed by Eckel (1976, p 774), "revenue is taken as being appropriately allocated to the period in which it falls, and accordingly, it is costs which will be manipulated" in most accounting systems.

4. Dopuch *et al.* (1982) suggest a slightly different order for the cost allocation process. They would begin by identifying a total cost to be allocated and a set of cost objects and then apply a cost allocation method. While the impact of the resulting allocations on managerial decisions is not explicitly recognized in their characterization, Dopuch *et al.* separately consider some underlying demands for cost allocations (Chapter 9).
5. A similar sentiment is expressed by Demski and Kreps (1982) in their review and critique of model-based research in managerial accounting: "We feel that the cost allocation literature would benefit by increased explicit attention to decision-focus (p 123)."
6. Davidson *et al.* (1979, p 15) also draw a distinction between joint and common costs: "Common costs result when multiple products are produced together although they could be produced separately: joint costs occur when multiple products are of necessity produced together."
7. Specifically, Baumol *et al.* (1982) show that economies of scope is a necessary and sufficient condition for the existence of multiproduct firms in perfectly contestable markets (where contestable markets are those accessible to potential entrants who can serve the same market demands and use the same production techniques as incumbent firms and who can evaluate the profitability of entry at the incumbent firms' pre-entry prices). Geometrically, economies of scope implies that the hyperplane passing through the origin and the costs of given amounts of each (subset of) products(s) produced individually lies above the cost of producing them jointly.
8. Economies of scale are typically defined as a less than proportionate increase in costs for a proportionate increase in outputs. Actually, a weaker condition, subadditivity, is sufficient to produce common cost savings. This concept is defined in Section 3. For a related discussion see Baumol *et al.* (1982, Chapter 2).
9. This emphasis is echoed by Henderson and Quandt (1980, p 92): "The case of joint products is distinguished on technical rather than organizational grounds and exists whenever the quantities of two or more products are technically interdependent."
10. Cost allocations have also been questioned on other grounds. Thomas (1971, 1974, 1978), for example, has argued that all cost allocations are arbitrary and incorrigible – this latter term meaning indefensible. His argument rests on the observation that no single allocation method is appropriate for all decision settings and that when just one method is to be chosen, the choice must therefore rest on arbitrary criteria. However, it does not necessarily follow that such a choice is arbitrary (or incorrigible). Moreover, even though a single allocation method must serve various users of external financial disclosures, managers are free to use any number of allocations for their decisions; see also Eckel (1976).
11. Positive research involves the development of a language and set of hypotheses which allow an activity to be explained and predicted; see Friedman (1953).
12. This argument employs the reasonable assumptions that division managers strive to maximize the utility rather than wealth, that perquisites increase their utilities, and that perquisites consumption is difficult for upper management (or owners) to monitor.
13. Alchian (1965) identifies some other problems with Williamson's arguments that also apply to Zimmerman. Specifically, Alchian recommends a three-dimensional alternative to Zimmerman's Figure 1 which would separately identify the utilities associated with pecuniary and nonpecuniary goods.

14. See Rees (1973, pp 22–30) for an analysis of the effects of changing marginal wage rates on the consumption of leisure.
15. They also examine the New Soviet Incentive Scheme, Incentive Compatible Allocations, and the Groves Allocation Scheme. While the Soviet Scheme is easier to implement, it achieves a lower level of efficiency than the Groves Allocation. In the New Soviet Incentive Scheme an operating division is charged a fixed amount which depends upon the difference between the actual and forecast usage of a service (zero if they are equal). The Incentive Compatibility approach is not an allocation scheme *per se*, but rather a set of conditions under which truthful reporting is an equilibrium. Thus, other allocation schemes can be evaluated according to these conditions. In a Groves Allocation Scheme the cost allocated to one operating division of a two-division firm equals the entire fixed service cost plus the budgeted level of variable cost for the other division adjusted by a subsidy that does not depend on the forecast or actual usage.
16. In a survey of the "Fortune 1000" companies, Reece and Cool (1978) found that 95.8 percent of the 620 respondents used investment or profit centers.
17. The Massachusetts Formula gives equal weight to sales, assets, and employees.
18. Although Moriarity states his allocation scheme more generally as applying to both joint and common cost settings, the advantages he suggests for his proposal apply to a situation in which the cost objects are ultimately rational, evaluative decision makers. Thus, his scheme is discussed in conjunction with common cost allocations.
19. As a result, the Moriarity method does not always satisfy the core criterion (is discussed in Section 3). See Ayers and Moriarity (1983) for a modification of the Moriarity method that does satisfy the core criterion.
20. In this case divisions 1 and 2 receive greater cost savings ($450) than they are allocated under the Moriarity method ($300).
22. Thus, the Louderback approach satisfies the core criterion.
23. A procedure widely used for allocating the common costs of water supply projects among users is the Separable Cost-Remaining Benefits (SCRB) method. It is based on the idea that common costs should be allocated essentially in proportion to the willingness of the user to pay. The method is described in detail by Young *et al.* (1982) who compare it with several other allocation schemes and conclude that it may be "one of the worst".
24. See Samuelson (1954).
25. Cohen and Loeb (1982) suggest that many common inputs, such as the provision of corporate legal staffs, WATS lines, and computer services, possess characteristics of both pure private and pure common inputs. They argue that the choice of the level of input (i.e., the numbers of lawyers, WATS lines, or size of a computer) is analogous to the choice of the amount of a pure common input and that the choice of its rate of use is analogous to the choice of amounts of a pure private input. However, notice that these example_ exhibit both excludability and rivalry, albeit at certain levels of usage. It would be an easy task, for example, to exclude a division that paid nothing for using a WATS line or computer. They also exhibit rivalry since their use by one division reduces the amounts potentially available to others at the time of use. The analogy to pure common inputs depends, therefore, on difficulties inherent in measuring demands for capacity and tracing costs associated with their use.
26. The games that Shapley examined are based on the finite theory of Von Neumann and Morgenstern (1944). See Luce and Raiffa (1957) for game theory terminology.
27. Shubik also published a slightly revised version of his paper a year later (Shubik, 1963).
28. While Shubik explicitly states his intention to illustrate a method applicable to allocations of costs, revenues, profits (and the setting of transfer prices), he limits his discussion by speaking primarily of profits; we do likewise. It should be noted, however, that the allocation of profit rather than revenues or costs is not innocuous. As Hughes and Scheiner (1980) observe, profit allocations require the

central authority to know both the separable profit functions of each user division and the cost functions of each supplier division.

29. A set function v where $v(\phi)=0$ is ***superadditive*** if $v(S \cup T) \geq v(S)+v(T)$ for all sets S, T where $S \cap T = \phi$. Thus, superadditivity captures the notion of synergy arising from joint rather than independent action. Note that in the cost allocation setting the corresponding condition is ***subadditivity***, wherein $v(S \cup T) \leq v(S)+v(T)$.
30. Shubik's emphasis on joint action and incentive effects is characteristic of common cost settings.
31. Actually, Shubik (1962) presents a slightly different set of properties based on an earlier set of axioms given by Shapley in Rand paper RM-670, which are equivalent in that they also result in the Shapley value. The original version differs in that the additivity axiom is weaker and that there is a ***homogeneity*** (or ***homogeneity of degree one***) axiom rather than a dummy axiom. In the above setting, homogeneity means that a homogeneous increase in prices (hence in costs, revenues, and profits) results in a corresponding homogeneous increase in each of the allocations. In other words, the unit of measurement is irrelevant. Homogeneity arises easily as a corollary to the set of properties given above.
32. Shapley also refers to it as the hidden axiom, since several authors have inadvertently omitted it and arrived at erroneous results from the viewpoint of his theory. See Shapley (1981) for more details, including a discussion of the interpretation of his axioms given by Luce and Raiffa (1957).
33. Shapley (1953).
34. For further discussion of the concept of fairness, see Thomas (1969, 1974), Moriarity (1975), Jensen (1977), Callen (1978), Hamlen and Hamlen (1981), Baiman (1981) and Balachandran and Ramakrishnan (1981).
35. See also Coase (1946).
36. Actually, there are two essentially equivalent dummy axioms. The version given above is the zero dummy axiom. The other is the inessential factor axiom, which in the above setting states: If a particular division contributes the same incremental profit to each coalition, then the division should be allocated precisely this incremental profit.
37. This property is called linearity in the context of nonatomic games, to be discussed in Section 4 of this chapter.
38. However, difficulty might be encountered if an attempt is made to allocate profits in place of costs and revenues separately. Shubik, for example, suggests truncating the profit distribution at zero when defining the characteristic function. Specifically, he permits a manager to be able to liquidate his division in order to guarantee himself a profit of not less than zero. Shubik allows liquidations because in his formulation the game is the firm. In order to specify it, he includes a central office as a player and defines all coalitions not including the central office to have a characteristic value of zero. However, this results in an allocation being made to the central office, which may be undesirable ***vis à vis*** the original intent of the allocation. (Shubik offers that this may be thought of as resulting from a levy paid by each of the producing departments to the central office). As an alternative to Shubik's formulation, one may wish to allow the possibility of allocating a loss (which could provide useful information), as well as disallow self-liquidation (which is perhaps more realistic).
39. Both Thomas (1980) and Hamlen ***et al.*** (1980) point out that the Shapley value evenly divides interaction effects.
40. Aumann and Shapley also report a non-probabilistic interpretation of the Shapley value due to Harsanyi (1959). Imagine each coalition S of players forming a syndicate which declares a dividend. Each syndicate S distributes evenly among its members the difference between its worth, $v(S)$, and the total dividend declarations of all its subsyndicates. Harsanyi shows that the sum of all dividends received by a player under this scheme is precisely his Shapley value. Luce and Raiffa (1957) also suggest an interpretation of the Shapley value as an arbitration

scheme. This is appealing, since certain aspects of arbitration seem to be inherent in some of the cost allocation applications, notably those that are concerned with fairness to division managers.

41. Shapley (1953) defines a game as "a set of rules with specified players in the playing positions" (p 307).
42. A related point is made by Luce and Raiffa (1957), who suggest that since the Shapley value depends only upon the characteristic function it does not adequately reflect the threat powers of the players. In a similar vein, Aumann and Shapley (1974) describe the term characteristic function as being misleading and mathematically imprecise, in that v characterizes only certain ***coalitional*** – but not ***strategic*** – aspects of the overall competitive situation. They suggest the alternative terms worth function and coalitional worth function.
43. They also question an underlying assumption of Roth and Verrecchia (1979) that each manager's utility function is approximately linear in the range of potential payoffs.
44. If the fraction in the assumption is $f(r)$ rather than $1/r$, then the utility of playing a game is given by:

$$u(v) = \sum_{S} k(s)[v(S) - v(S - \{i\})]$$

where

$$k(s) = \sum_{r=s}^{n} (-1)^{r-s} \binom{n-s}{r-s} f(r)$$

45. A second property also presented by Shubik as an easily proved theorem states that any action taken by a division which increases the value of the firm as a whole does not cause the profits allocated to that division to decrease.
46. Briefly, the nucleolus is the solution in which the minimum excess over all coalitions is maximized, where a coalition's excess is defined as the difference between the total amount allocated to its members and its characteristic value. For a formal definition of the nucleolus and derivations of some of its properties, see Schmeidler (1969).
47. As noted in Section 2, only a marginal approach can provide that result.
48. In note 38, it is pointed out that Shubik's original intention was that the characteristic function in the profit allocation setting be nonnegative. However, Telser (1978) points out, "It would be nice to have a means of converting a given characteristic function into a nonnegative characteristic function without affecting the status of the core of a given function. Unfortunately, this in general is not possible."
49. A set function v, where $v(\phi) = 0$ is *convex* if $v(S) + v(T) \leqq v(S \cup T) + v(S \cap T)$ for all sets S, T.

Bibliography

Abdel-khalik, A.R. and Lusk, E.J. (1974) Transfer pricing – a synthesis. *The Accounting Review* 49: 8–23.

Alchian, A.A. (1965) The basis of some recent advances in the theory of management of the firm. *Journal of Industrial Economics* November: 30–41.

Alchian, A.A. (1971) Costs and outputs, in W. Breit and H. Hochman (Eds) *Readings in Microeconomics* (New York: Holt, Rinehart and Winston).

Amershi, A.H. (1981) *Equilibrium Cost Allocations* Working Paper (Stanford University).

Aumann, R.J. and Shapley, L.S. (1974) *Values of Non-Atomic Games* (Princeton, NJ: Princeton University Press).

Avery, H.G. (1951) Accounting for joint costs. *The Accounting Review* 26: 232–8.

Ayers, F. and Moriarity, S. (1983) *A Comparison of the Shapley and Moriarity Allocations in the Product Mix Decision* Working Paper (University of Oklahoma).

Bailey, E.E., and White L.J. (1974) Reversals in peak and off peak prices. *The Bell Journal of Economics* 5: 75–92.

Baiman, S. (1981) The concept of fairness in the choice of joint cost allocation methods, in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press) pp 106–9.

Baiman, S. (1982) Agency research in managerial accounting: a survey. *Journal of Accounting Literature* 1: 154–213.

Baker, K.R. and Taylor, R.E. (1979) A linear programming framework for cost allocation and external acquisition when reciprocal services exist. *The Accounting Review* 54: 784–90.

Balachandran, B., Li, L., and Magee, R. (1982) *On the Allocation of the Fixed and Variable Costs from Service Departments* Working Paper (Northwestern University).

Balachandran, B. and Ramakrishnan, R. (1981) Joint cost allocation: a unified approach. *Accounting Review* 56: 85–96.

Baumol, W.J., Panzar, J.C. and Willig, R.D. (1982) *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt, Brace, Jovanovich).

Bernhard, R.N. (1968) Some problems in applying mathematical programming to opportunity costing. *Journal of Accounting Research* 6: 143–8.

Biddle, G.C. and Steinberg, R. (1984) Allocations of joint and common costs. *Journal of Accounting Literature* 3: 1–45.

Billera, L.J. and Heath, D.C. (1982) Allocation of shared costs: a set of axioms yielding a unique procedure. *Mathematics of Operations Research* 7: 32–9.

Billera, L.J., Heath, D.C., and Raanan, J. (1978) Internal telephone billing rates – a novel application of non-atomic game theory. *Operations Research* 26: 956–65.

Billera, L.J., Heath, D.C., and Verrecchia, R.E. (1981) A unique procedure for allocating common costs from a production process. *Journal of Accounting Research* 19: 185–96.

Boatsman, J., Hansen, D., and Kimbrell, J. (1981) A rationale and some evidence supporting an alternative to the simple Shapley value, in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press) pp 53–74.

Callen, J.L. (1978) Financial cost allocations: a game theoretic approach. *The Accounting Review* 53: 303–8.

Capettini, R. and Salamon, G.L. (1977) Internal versus external acquisition of services when reciprocal services exist. *The Accounting Review* 52: 690–96.

Churchill, N. (1964) Linear algebra and cost allocations: some examples. *Accounting Review* 39: 894–904.

Coase, R.H. (1946) The marginal cost controversy. *Economica* 169–82.

Cohen, S.I. and Loeb, M. (1982) Public goods, common inputs, and the efficiency of full cost allocations. *Accounting Review* 57: 336–47.

Colberg, M.R. (1941) Monopoly prices under joint costs: fixed proportions. *Journal of Political Economy* 49: 103–10.

Davidson, S., Stickney, C., and Weil, R.L. (1979) *Accounting: The Language of Business* (New York: Horton).

Davidson, S. and Weil, R.L. (1977) *Handbook of Modern Accounting* (New York: McGraw-Hill).

Dearden, J. (1962) *Cost and Budget Analysis* (Englewood Cliffs, NJ: Prentice-Hall).

Demski, J.S. (1981) Cost allocation games in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press) pp 142–73.

Demski, J.S. and Kreps, D.M. (1982) Models in managerial accounting. *Journal of Accounting Research* Supplement: 117–48.

Dopuch, N. (1981) Some perspectives on cost allocations, in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press) pp 1–7.

Dopuch, N., Birnberg, J.G., and Demski, J.S. (1982) *Cost Accounting* (New York: Harcourt, Brace, Jovanovich).

Dopuch, N. and Drake, D. (1964) Accounting implications of a mathematical programming approach to the transfer price problem. *Journal of Accounting Research* 2: 10–24.

Dubey, P. (1982) The Shapley value as aircraft landing fees – revisited. *Management Science* 28: 869–74.

Eckel, L.G. (1976) Arbitrary and incorrigible allocations. *Accounting Review* 50: 764–777.

Fremgen, J.M. and Liao, S.S. (1981) *The Allocation of Corporate Indirect Costs* (National Association of Accountants).

Friedman, M. (1953) *Essays in Positive Economics* (Chicago, IL: University of Chicago Press).

Gamson, W. (1961) A theory of coalition formation. *American Sociological Review* June: 373–82.

Gangolly, J.S. (1981) On joint cost allocation: independent cost proportional scheme (ICPS) and its properties. *Journal of Accounting Research* 19: 299–312.

Gould, J.R. (1964) Internal pricing in firms when there are costs of using an outside market. *Journal of Business* 37: 61–7.

Hadley, G. (1962) *Linear Programming* (New York: Addison–Wesley).

Hamlen, S.S. and Hamlen, W.A. (1981) The concept of fairness in the choice of joint cost allocation methods, in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press) pp 84–102.

Hamlen, S.S., Hamlen, W.A., and Tschirhart, J.T. (1977) The use of core theory in evaluating joint cost allocation schemes. *Accounting Review* 51: 616–27.

Hamlen, S.S., Hamlen, W.A., and Tschirhart, J.T. (1980) The use of the generalized Shapley value in joint cost allocation. *Accounting Review* 55:269–87.

Hartley, R.V. (1971) Decision making when joint products are involved. *The Accounting Review* 46: 746–55.

Harsanyi, J.C. (1959) A bargaining model for the cooperative n-person game, in A.W. Tucker and R.D. Luce (Eds) *Contributions to the Theory of Games*, Vol. IV (Princeton, NJ: Princeton University Press) pp 324–56.

Hass, J.E. (1968) Transfer pricing in a decentralized firm. *Management Science* 14B: 310–31.

Hayek, F.A. (1944) *The Road to Serfdom* (Chicago, IL: University of Chicago Press).

Hazelwood, A. (1951) Optimum pricing as applied to telephone service. *Review of Economic Studies* 21: 67–78.

Henderson, J.M. and Quandt, R.E. (1980) *Microeconomic Theory* (New York: McGraw-Hill).

Hirsch, W.M. and Dantzig, G.B. (1968) The fixed charge problem. *Naval Research Logistics Quarterly* 15: 413–24.

Hirschleifer, J. (1956) On the economics of transfer pricing. *Journal of Business* 29: 172–84.

Hirschleifer, J. (1958) Peak loads and efficient pricing: comment. *Quarterly Journal of Economics* 451–62.

Hirschleifer, J. (1964) Internal pricing and decentralized decisions, in C. Bonini, R. Jaedicke, and M. Wagner (Eds) *Management Controls: New Directions in Basic Research* (New York: McGraw-Hill).

Hirschleifer, J. (1976) *Price Theory and Applications* (Englewood Cliffs, NJ: Prentice-Hall).

Horngren, C. (1982) *Cost Accounting: A Managerial Emphasis* (Englewood Cliffs, NJ: Prentice-Hall).

Hughes, J.S. and Scheiner, J.H. (1980) Efficiency properties of mutually satisfactory cost allocations. *Accounting Review* 55: 85–95.

Itami, H. and Kaplan, R.S. (1980) An activity analysis approach to unit costing with multiple interactive products. *Management Science* 26: 826–39.

Jennergren, L.P. (1972) Decentralization on the basis of price schedules in linear decomposable resource-allocation problems. *Journal of Financial and Quantitative Analysis* 1407–17.

Jensen, D.L. (1974) The role of cost in pricing joint products: a case of production in fixed proportions. *Accounting Review* 49: 465–76.

Jensen, D.L. (1977) A class of mutually satisfactory allocations. *Accounting Review* 51: 842–56.

Jensen, R.E., Manes, R., and Park, S. (1981) *Parametric Programming of Capacity Costs in Reciprocal Service Departments* Working Paper (Champaign, IL: University of Illinois).

Kaplan, R.S. (1973) Variable and self-service costs in reciprocal allocation models. *Accounting Review* 48: 738–48.

Kaplan, R.S. (1981) The impact of management accounting research on policy and practice, in J. Buckley (Ed) *Proceedings of the Arthur Young Professors' Roundtable* (Council of Arthur Young Professors).

Kaplan, R.S. (1982) *Advanced Management Accounting* (New York: Prentice-Hall).

Kaplan, R.S. and Thompson, G.L. (1971) Overhead allocation via mathematical programming models. *Accounting Review* 46: 352–64

Kaplan, R.S. and Welam, U.P. (1974) Overhead allocation with imperfect markets and non-linear technology. *Accounting Review* 49: 477–814.

Kreps, T.J. (1930) Joint costs in the chemical industry. *Quarterly Journal of Economics* 46: 416–61.

Lindahl, E. (1958) Just taxation – a positive solution, in R.A. Musgrave and A.T. Peacock (Eds) *Classics in The Theory of Public Finance* (New York: Macmillan) pp 168–176.

Littlechild, S.C. (1970) Marginal cost pricing with joint costs. *The Economic Journal* 80: 323–35.

Littlechild, S.C. (1975) Common costs, fixed charges, clubs and games. *Review of Economic Studies* 42: 117–24.

Littlechild, S.C. and Owen, G. (1973) A simplistic expression for the Shapley value in a special case. *Management Science* 20: 370–2.

Littlechild, S.C. and Thompson, G.F. (1977) Aircraft landing fees: a game theory approach. *Bell Journal of Economics* 8: 186–204.

Loehman, E. and Whinston, A.B. (1971) A new theory of pricing and decision-making for public investment. *Bell Journal of Economics* 2: 606–25.

Loehman, E. and Whinston, A.B. (1974) An axiomatic approach to cost allocation for public investment. *Public Finance Quarterly* 2: 236–51.

Louderback, J.G. (1976) Another approach to allocating joint costs: a comment. *Accounting Review* 50: 683–85.

Luce, R.D. and Raiffa, H. (1957) *Games and Decisions* (New York: Wiley).

Manes, R.P. (1965) Comment on matrix theory and cost allocation. *Accounting Review* 40: 640–43.

Manes, R.P. (1982) *Market Based Joint Cost Allocations*, Working Paper (Champaign, IL: University of Illinois).

Manes, R.P., Park, S.H., and Jensen R. (1982) Relevant costs of intermediate goods and services. *Accounting Review* 52: 594–606.

Manes, R.P. and Smith V.L. (1965) Economic joint cost theory and accounting practice. *Accounting Review* 40: 31–5.

Manes, R.P. and Verrecchia, R. (1982) A new proposal for setting intra-company transfer prices. *Accounting and Business Research* 13: 77–104.

Mautz, R.K. and Skousen, K.F.(1968) Common cost allocation in diversified companies. *Financial Executive* 15–7, 19–25.

Megiddo, N. (1978) Computational complexity of the game theory approach to cost allocation for a tree. *Mathematics of Operations Research* 3: 189–95.

Minch, R. and Petri, E. (1972) Matrix models of reciprocal service cost allocation. *Accounting Review* 47: 576–80.

Mirman, L.J., Samet, D. and Tauman, Y. (1982) *An Axiomatic Approach to the Allocation of a Fixed Cost Through Prices*, Working Paper (Evanston, IL: Northwestern University).

Mirman, L.J. and Tauman, Y. (1982) Demand compatible, equitable, cost sharing prices. *Mathematics of Operations Research* 7: 40–56.

Moriarity, S. (1975) Another approach to allocating joint costs. *Accounting Review* 49: 791–5.

Moriarity, S. (1976) Another approach to allocating costs: a reply. *Accounting Review* 50: 686–7.

Myerson, R.B. (1977) Values of games in partition function form. *International Journal of Game Theory* 6: 23–31.

National Association of Accountants (1957) *Costing Joint Products* (New York: NAA).

Neuner, J.W. and Frumer, S. (1967) *Cost Accounting: Principles and Practices* (New York: Irwin).

Nurnberg, H. (1978) Joint products and by-products, in S. Davidson and R.C. Weil (Eds) *Handbook of Cost Accounting* (New York: McGraw-Hill).

Panzar, J.C. (1976) A neoclassical approach to peak load pricing. *The Bell Journal of Economics* 7: 521–30.

Pigou, A.C. (1913) Railway rates and joint cost. *Quarterly Journal of Economics* 27: 536–38.

Rappaport, A. (1978) Divisional Cost Analysis, in S. Davidson and R.L. Wei (Eds) *Handbook of Cost Accounting* (New York: McGraw-Hill).

Reece, J.S. and Cool W.R. (1978) Measuring investment center performance. *Harvard Business Review* May-June: 28–46, 174–6.

Rees, A. (1973) *The Economics of Worth and Pay* (New York: Harper and Row).

Roth, A.E. (1977) *The Shapley value as applied to cost allocation: a reinterpretation*, Working Paper (University of Illinois).

Roth, A.E. and Verrecchia, R.E. (1979) The Shapley value as applied to cost allocation: a reinterpretation. *Journal of Accounting Research* 17: 295–302.

Samuelson, P.A. (1954) The pure theory of public expenditure. *Review of Economics and Statistics* 387–9.

Schmeidler, D. (1969) The nucleolus of characteristic function game. *SIAM Journal of Applied Mathematics* 17: 1163–70.

Shapley, L.S. (1953) A value for n-person games, in H.W. Kuhn and A.W. Tucker (Eds) *Contributions to the Theory of Games*, Vol. II (Princeton, NJ: Princeton University Press) pp 307–17.

Shapley, L.S. (1965) *Notes on n-Person Games VII: Cores of Convex Games*, RM-4571 (Santa Monica, CA: Rand Corporation).

Shapley, L.S. (1981) Comments on P.D. Banker's 'Equity considerations in traditional full cost allocation practices: an axiomatic perspective', in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press).

Sharkey, W.W. (1982) Suggestions for a game-theoretic approach to public utility pricing and cost allocation. *Bell Journal of Economics* 13: 57–68.

Shilling law, G. (1982) *Managerial Cost Accounting* (New York: lrwin).

Shubik, M. (1962) Incentives, decentralized control, the assignment of joint costs and internal pricing. *Management Science* 8: 325–43.

Shubik, M. (1963) Incentives, decentralized control, the assignment of joint costs and internal pricing, in C. Bonini, R. Jaedicke, and M. Wagner (Eds) *Management Controls* (New York: McGraw-Hill).

Solomons, D. (1965) *Divisional Performance: Measurement and Control* (New York: Irwin).

Solomons, D. (1977) Divisional Reports, in S. Davidson and R.L. Weil (Eds) *Handbook of Modern Accounting* (New York: McGraw-Hill).

Staubus, G.J. (1971) *Activity Costing and Input–Output Accounting* (New York: Irwin).

Steiner, P.O. (1957) Peak loads and efficient pricing. *Quarterly Journal of Economics* 71: 585–610.

Stigler, G.J. (1966) *The Theory of Price*, 3rd edn (New York: Macmillan).

Telser, L.G. (1978) *Economic Theory and the Core* (Chicago, IL: University of Chicago Press).

Thomas, A.L. (1969) *The Allocation Problem in Financial Accounting Theory*. Studies in Accounting Research No. 3 (American Accounting Association).

Thomas, A.L. (1971) Useful arbitrary allocations. *Accounting Review* 45: 472–9.

Thomas, A.L. (1974) *The Allocation Problem: Part II*. Studies in Accounting Research No. 9 (American Accounting Association).

Thomas, A.L. (1978) Arbitrary and incorrigible allocations: a comment. *Accounting Review* 52: 263–9.

Thomas, A.L. (1980) *A Behavioral Analysis of Joint Cost Allocation and Transfer Pricing* (Champaign, IL: Stipes).

Thrall, R.M. and Lucas, W.F. (1963) *N*-person games in partition function form. *Naval Research Logistics Quarterly* 10: 281–98.

Turvey, R. (1969) Marginal cost. *Economic Journal* June: 282–99.

Verrecchia, R.E. (1981a) *The Choice Among Cost Allocation Alternatives*, Working Paper (University of Chicago).

Verrecchia, R.E. (1981b) A question of equity: use of the Shapley value to allocate state and local income and franchise taxes, in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press) pp 14–40.

Verrecchia, R.E. (1982) An analysis of two cost allocation cases. *Accounting Review* 57: 579–93.

Von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press).

Walters, A.A. (1960) The allocation of joint cost with demands as probability distributions. *American Economic Review* 50: 419–32.

Weil R.L. (1968) Allocating joint costs. *American Economic Review* 58: 1342–5.

Williamson, O.E. (1966) Peak load pricing and optimal capacity under indivisibility constraints. *American Economic Review* 56: 810–27.

Williams, T.H. Griffin, C.H. (1964) Matrix theory and cost allocation. *Accounting Review* 39: 671–78.

Wright, F.K. (1970) Dual variables in inventory measurement. *Accounting Review* 45: 129–133.

Wright, R.L. (1983) Measuring the precision of statistical cost allocations. *Journal of Business and Economic Statistics* 1: 93–100.

Young, H.P., Okada, N., and Hashimoto, T. (1982) Cost allocation in water resources development. *Water Resources Research* 18: 463–75.

Young, H.P. (1980) *Cost Allocation and Demand Revelation in Public Enterprises*. Working Paper WP-80-130. (Laxenburg, Austria: International Institute for Applied Systems Analysis)

Young, H.P.(1985) Monotonic solutions of cooperative games *International Journal of Game Theory* 14.

Young, H.P. (1985) Producer incentives in cost allocation, *Econometrica* 53.

Zimmerman, J.L. (1979) The costs and benefits of cost allocations. *Accounting Review* 54: 504–21.

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 3

ON THE USE OF GAME-THEORETIC CONCEPTS IN COST ACCOUNTING

Leonard J. Mirman, Yair Tauman, and Israel Zang

1 Introduction

The problem of allocation in cost accounting has received substantial attention in recent years; see, for example, Moriarity (1981a), Thomas (1969, 1974, 1980), and Zimmerman (1979) for analyses of various aspects of this problem in the accounting context. In this chapter game-theoretic concepts are used in what may be called the contribution approach to cost allocation problems.

The need for allocation schemes arises in many actual applications. Some are known as cost allocation problems; that is, the allocation of the cost of a joint production process resulting in the output of several goods whose costs (possibly both fixed and variable) are nonseparable. In these applications it is necessary to allocate the total (joint) cost of production among the outputs. Similarly, there is a dual problem in that profits must be allocated among the inputs contributing to these profits.

The classical cost allocation problem is that of pricing the products of a regulated industry. In this case there is a monopoly which operates on a cost, cost plus, or fixed profit basis. Output prices that allocate the total revenue among the outputs must be determined. A similar, but dual, situation arises when several firms or divisions cooperate, yielding profits that are greater than the sum of the individual profits. The problem is how to allocate these profits among the contributing firms. The need for cost or revenue allocation encompasses a substantially wider set of applications in which there is a need to know the contribution or the average cost of one or several, but not necessarily all, outputs to the total joint cost of producing all the outputs.

Such a need arises in accounting reporting when, for example, the value of inventories must be assessed. Consider, as an example, a firm producing two different intermediate goods in a joint production process, where some of the goods are used for the production of final products in the present period and the rest are kept in inventory for use in the second period. For tax reporting purposes it is necessary to associate some cost accounting figures to each of the two commodities to determine the net profit of the firm in each of the two periods. If a cost-based approach is followed, then it is necessary to know the unit production cost of the different intermediate goods. If, however, a net realizable-value approach is adopted, then knowledge of the contribution to the total second-period net revenue of each unit of the inputs in inventory is required. The need to know the contribution of some outputs to joint costs arises also in the case of a firm jointly producing both nonregulated and regulated products, or if different tax rates apply to different jointly produced goods, or in cases where user fees must be determined for some (say, government) services. Another application that may call for the knowledge of actual contributions to joint costs is the determination of transfer prices in decentralized organizations.

While the use of cost allocation models in some of the above cases is debatable (see, e.g., Dopuch, 1981; Moriarity, 1981b; Zimmerman, 1979), it is obvious that cost allocation figures are used in practice in many situations. Thus, in this chapter we avoid the problem of the appropriateness of allocation models and discuss the determination of the contribution of outputs to joint costs (or of inputs to joint profits) in cases where these figures are needed.

In the single-product case the contribution of each unit of the output to the total cost is the average cost. Thus, the problem may be viewed as the determination of the average costs of each of several jointly produced goods. This chapter is concerned mostly with cost allocation problems through a generalization of the one-dimensional notion of average costs. Allocating profits among inputs can be carried out similarly, as is demonstrated in Example 2.

In order to describe the problem and several solution concepts some notation is needed. Suppose that m goods are produced, let $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ be a vector whose ith component x_i denotes the quantity of the ith good. Also, let $F(\boldsymbol{x}) = F(x_1, x_2, \ldots, x_m)$ be the cost of producing the vector $\boldsymbol{x}$. For a particular vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ of quantities, we wish to determine the vector of prices $\boldsymbol{p} = (p_1, p_2, \ldots, p_m)$. Here p_i represents that part of the total cost $F(\alpha)$, of producing α, attributable to each unit of the ith good, namely, its (generalized) average cost. This means that the well known cost-sharing requirement holds, i.e.,

$$\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_m p_m = F(\alpha_1, \alpha_2, \ldots, \alpha_m) \quad .$$

Some or all of the p_is may be used depending upon the requirements of the particular problem. It is evident that in some applications knowledge of the price vector may affect the decision makers in determining actual market prices.

In recent years, concepts of game theory have been frequently suggested as tools in allocation problems. Special attention was devoted to the use of the *Shapley value* (Shapley, 1953), the *core* and *nucleolus* (Schmeidler, 1969). (See, for example, Balachandran and Ramakrishnan, 1981; Callen, 1978; Hamlen *et al.*, 1977, 1980; Littlechild and Thompson, 1977; Loehman and Whinston, 1971, 1976; Roth and Verrecchia, 1979; Shubik, 1964.) However, these concepts are not as appropriate for real contribution costing as the approach discussed in Section 2 of this chapter. This approach, based on cost considerations only, is due to Billera *et al.* (1978), who used concepts of the theory of nonatomic games due to Aumann and Shapley (1974) to allocate the cost of the Cornell University telephone system to its users. They suggested the use of what has come to be known as the Aumann–Shapley (hereafter AS) price mechanism for allocating joint *variable* costs among products. The derivation of this mechanism, based on game-theoretic concepts, is discussed in Section 2. As it turns out, this price mechanism can be systematically calculated in many practical situations. One of these is the case in which the minimum cost is obtained as a result of a linear programming cost minimization problem. In this case the calculation of prices involves a parametric linear programming process. An algorithm for carrying out these calculations for transportation problem cost functions was suggested by Samet *et al.* (1984). This process for the general linear programming cost function is outlined in Section 2 with some examples.

In Section 3 an axiomatic approach to the derivation of AS prices is discussed. This approach, free of game-theoretic concepts, constitutes the main justification for the use of these prices as the real contributions or average costs of each of the outputs. Accordingly, it can be argued that AS prices constitute a very natural extension of average cost prices to the case of joint production of several goods. This extension emanates from the axiomatic approach derived independently by Billera and Heath (1982) and Mirman and Tauman (1981, 1982), where it is shown that AS prices are characterized uniquely for a large family of cost problems using a set of natural requirements. Billera *et al.* (1981) provided a detailed discussion and interpretation of this axiomatic approach in the accounting context. Thus, the discussion in Section 3 focuses on justifying the use of these prices as average cost prices. The interested reader should refer for further ideas to Billera *et al.* (1981). The axiomatic approach has been extended in several directions by Mirman *et al.* (1982a,b), Samet and Tauman (1980) and Samet *et al.* (1984). Some of these extensions are reviewed in Section 3. The important complementary problem of allocating fixed costs is discussed, using the work of Mirman *et al.* (1983), in Section 4.

2 Game-Theoretic Concepts and Aumann–Shapley Prices

Consider a joint production process, described by a cost function F, yielding several outputs whose quantities are given by the vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$, and suppose we wish to find the contribution of a unit of each output to the total production cost $F(\boldsymbol{\alpha})$. Since most applications make use of this information to determine prices, these per unit contributions are frequently referred to as prices. Throughout this section it is assumed that the production process contains no fixed cost component so that only the contribution of each output to the ***variable cost*** is determined. Hence, it is assumed that F is continuous and $F(0, 0, \ldots, 0) = 0$. In Section 4 fixed costs are incorporated into the analysis.

Before beginning the analysis it should be stressed that, in general, the use of marginal costs (MC) for allocation problems reflects only the contribution to the total cost of the last unit produced. Therefore, the use of this concept as an allocation tool may introduce, as is shown in Example 2 below, some severe distortions. Yet there are cases where marginal costs do reflect the (average) contribution of a unit output to the total cost; for example, when the production technology has constant returns to scale. In this case, AS prices and MC prices coincide. On the other hand, when the objective is profit maximization it may be desirable to use MC prices since the distortions (or cross-subsidization) can be exploited to increase profits.

In order to discuss various game-theoretic solution concepts it is first necessary to introduce several game-theoretic notions.

A ***game*** in characteristic form consists of n *players* denoted by indices $1, 2, \ldots, n$ and a characteristic function v which assigns a value $v(S)$ to every subset S of the n players. For short, we refer to this game as the game v. Denote by N the set of all players. In game-theoretic terminology, S is a *coalition* while N is the ***grand coalition***. For the cost allocation problem these concepts may be interpreted in two ways. First the players may be interpreted as the goods being produced so that $v(S)$ is the cost of producing the subset of goods S and $v(N)$ is the cost of producing all the goods. Such games are called ***goods games***. Second, the players may be interpreted as the consumers who demand the goods. In this case $v(S)$ is the cost of producing those goods demanded by a subset S of the consumers and $v(N)$ is the cost of producing all the goods demanded by everyone. Such games are called ***consumer games***. An ***imputation*** of a game v is an allocation $\mathbf{z} = (z_1, \ldots, z_n)$ of $v(N)$ among the n players of the game. Thus $\mathbf{z}$ is a vector of payments such that

$$z_1 + z_2 + \cdots + z_n = v(N) \quad . \tag{1}$$

Consider a joint production process described by a cost function F of m outputs. Denote by $1, \ldots, n$ the individuals consuming these goods and assume that individual i wishes to consume the quantities $\alpha^i = (\alpha_1^i, \ldots, \alpha_m^i)$ of the m outputs. The total consumption is thus a nonnegative vector $\boldsymbol{\alpha}$,

$$\boldsymbol{\alpha} = \sum_i \alpha^i \quad .$$

As already noted, this situation may be associated with two games in characteristic form v and $\bar{v}$, respectively, where v is the consumer game in which the players are the consumers and $\bar{v}$ is the goods game in which the players are the outputs. First we consider the game v. For any subset S of consumers let $v(S)$ be the cost of producing the total consumption $\sum_{i \in S} \alpha^i$ demanded by the consumers in S, namely

$$v(S) = F\left(\sum_{i \in S} \alpha^i\right) \quad . \tag{2}$$

With this definition

$$v(N) = F(\alpha) \quad .$$

Any imputation $\mathbf{z} = (z_1, \ldots, z_n)$ of the game v satisfies $z_1 + \cdots + z_n = v(N) = F(\boldsymbol{\alpha})$ and hence can be used to allocate the total joint costs $F(\alpha)$ among the n consumers. Thus allocating the total cost among the n consumers is the same as selecting an imputation for the game v. But this is exactly the same issue as selecting a solution concept for a game in characteristic form (or cooperative game with side payments). The most common solution concepts are the core, the Shapley value, and the nucleolus, each of which has been used to allocate joint cost. Let us briefly describe the core and the Shapley value.

The *core* of the game v consists of all imputations of v such that the total cost imputed to every possible coalition S cannot exceed the cost $v(S)$ of producing the set of goods of coalition S by themselves. Otherwise there would be some coalition of consumers that would object to the imputation since this coalition can produce its own consumption $\sum_{i \in S} \alpha^i$ more cheaply by itself. Formally, an imputation $\boldsymbol{z} = (z_1, \ldots, z_n)$ is in the core of v if and only if for each $S \subseteq N$,

$$\sum_{i \in S} z_i \leq v(S) = F\left[\sum_{i \in S} \alpha^i\right] . \tag{3}$$

The core might be an empty set [i.e., there exists no $\boldsymbol{z}$ satisfying both equations (1) and (3)] or it may contain many imputations. Thus the core cannot, in general, serve as the sole criterion for allocating the total cost $F(\boldsymbol{\alpha})$, although it may sometimes be important to know whether a certain allocation is in the core (see, e.g., Mirman *et al.* 1982a,b).

Next consider the Shapley value. Unlike the core this solution concept associates with ***each*** game a ***unique*** imputation. The ***Shapley value*** $\boldsymbol{s} = (s_1, \ldots, s_n)$ is the imputation that assigns to each of the n players in the game its average contribution to a random coalition. The randomization is taken as follows. Consider the family of all coalitions that do not contain player i and partition this family in $n-1$ subfamilies having sizes $1, 2, \ldots, n-1$. Associate with each family the probability $1/n-1$ (namely each coalition size has equal probability of appearing) and assume that all coalitions within each subfamily have the same probability. Now if S is a coalition that does not contain i the probability of S occurring is

$$\frac{1}{n} \cdot \frac{1}{\begin{pmatrix} n-1 \\ |S| \end{pmatrix}}$$

where $|S|$ is the number of players in S, $1/n$ is the probability of a coalition of size $|S|$, and

$$\begin{pmatrix} n-1 \\ |S| \end{pmatrix}$$

is the number of coalitions of this size. Hence the average contribution of i is given by

$$s_i = \sum_{S \subseteq N} \frac{|S|!(n - |S| - 1)!}{n!} [v(S \cup \{i\}) - v(S)] , \tag{4}$$

which is the Shapley value of player i. The Shapley value was originally defined by a set of four axioms that turns out to result in a unique imputation given by equation (4). Note that even if the core of v is nonempty, the Shapley value, need not, in general, be an element of the core of v.

The allocation of $F(\boldsymbol{\alpha})$ among the players in the consumer game v raises some problems. In these games only coalitions that produce the total consumption of each of their members are studied. A player is (arbitrarily) excluded from joining a coalition to produce just part of his or her consumption while producing the other part with some other coalition. That is, the possible strategy of splitting a vector of demands into several parts, each part produced with a different coalition, is not taken into account in the structure of the game v and, hence, is not accounted for in the solution allocation of this game. To take these strategies into account it is necessary to allow coalitions to produce any quantity vector $(x_1, x_2, \ldots, x_m)$ that is not greater than the total demand vector $(\alpha_1, \alpha_2, \ldots, \alpha_m)$.

As is shown later, this approach produces a game with a continuum of players which comprise infinitesimal quantities of commodities. An imputation in this game results in an allocation of the cost to the goods and not to the consumers. Details of this approach are discussed in the sequel.

Notice also that the consumers game v should not be used to determine the *contribution* of outputs to cost. Once the quantity of aggregate demand determining the production cost is fixed the contribution of each output to the total cost should also be fixed. Moreover, the contributions of each good must not depend upon how the aggregate demand vector is split among the consumers. That is, as long as the same aggregate output vector is produced the contribution of each output to the total cost should be the same regardless of how this vector is obtained. Clearly, when the consumers game v is used, the resulting contributions depend crucially on how a given output vector is distributed among consumers. On the other hand, contributions resulting from the goods game (i.e., AS prices) do not change if the aggregate demand remains the same, but individual consumers are allowed to change their individual demands. Thus, if for any reason some consumers decide to cooperate by changing their individual consumption vectors, but not their total consumption vector they, as a group, do not profit.

As a first approach to a game, in characteristic form, with goods as players, consider the ***finite goods*** game $\bar{v}$ in which each good is a player, so that there are m players denoted by the indices $1, 2, \ldots, m$. The set of all players (i.e., the grand coalition) is denoted by M. The game is defined as follows: consider the aggregate consumption vector $\boldsymbol{\alpha}$ and let S be a subset of the goods. Let $\boldsymbol{\alpha}^S$ be the vector which consists of those elements of $\boldsymbol{\alpha}$ belonging to S, namely,

$$\alpha_i^S = \begin{cases} \alpha_i & \text{if } i \in S \\ 0 & \text{if } i \notin S \end{cases} .$$

Then $\bar{v}(S)$ is the cost of producing the quantity $\boldsymbol{\alpha}^S$ of goods in S, i.e.,

$$\bar{v}(S) = F(\boldsymbol{\alpha}^S)$$

and

$$\bar{v}(M) = F(\boldsymbol{\alpha}^M) = F(\boldsymbol{\alpha}) \quad .$$

Again the Shapley value of the game $\bar{v}$ may be used to find the allocation of $\bar{v}(M)$ for the various goods. That is,

$$s_i = \sum_{S \subseteq N} \frac{|S|!(m - |S| - 1)!}{m!} [\bar{v}(S \cup \{i\}) - \bar{v}(S)] \quad ,$$

is the imputation assigned to good (player) i. Consequently, the per unit contribution of good i, denoted by p_i^S, is the ratio s_i / α_i.

There is, however, a difficulty with the use of the Shapley value for the finite game $\bar{v}$. To show this and to give some insight into the suggested solution, we first discuss an example. Consider the case in which the minimum cost is obtained from the solution of a linear programming problem; that is, F is a *linear programming cost function*. Such a function has the general form,

$$F(x) = \min_{y} \{c_1 y_1 + c_2 y_2 + \cdots + c_l y_l\} \quad ,$$

subject to

$$\mathbf{A} y \geqq x$$

$$\mathbf{B}\boldsymbol{y} \leqq \boldsymbol{b}$$

and

$$\boldsymbol{y} > 0 \quad ,$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_l)$ is a decision vector whose components denote quantities of the variable inputs and $\boldsymbol{c} = (c_1, c_2, \ldots, c_l)$ is a cost vector, $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ is the output vector, $\boldsymbol{b} = (b_1, b_2, \ldots, b_q)$ is a vector whose components denote quantities of the fixed inputs, $\mathbf{A}$ is an $m \times l$ matrix determining the relationship between the variable inputs and outputs, and $\mathbf{B}$ is a $q \times l$ matrix determining the relationship between the fixed and variable inputs.

Example 1

Consider a firm or division producing two products A and B, whose quantities are denoted by x_A and x_B, respectively, using two types of machines denoted by the indices 1 and 2. There is one machine of type 1, and two machines of type 2. Each machine can operate up to 400 minutes a day and process each product separately or jointly. A unit of A requires 1 minute processing time on machine 1 or 10 minutes on machine 2. The requirements of a unit of B are 2 and 30 minutes on machines 1 and 2, respectively. The cost of processing a unit of A is \$1 on machine 1 and \$10 on 2, while for B the costs are \$3 and \$25 on machines 1 and 2, respectively. The resulting cost function is:

$$F(x_A, x_B) = \min\{y_{A1} + 10y_{A2} + 3y_{B1} + 25y_{B2}\} \quad ,$$

subject to

$$y_{A1} + y_{A2} \geqq x_A \quad ,$$

$$y_{B1} + y_{B2} \geqq x_B \quad ,$$

$$y_{A1} + 2y_{B1} \leqq 400 \quad ,$$

$$10y_{A2} + 30y_{B2} \leqq 800 \quad ,$$

$$y_{A1}, y_{A2}, y_{B1}, y_{B2} \geqq 0 \quad ,$$

where y_{Ai} is the number of units of A and y_{Bi} the number of units of B, processed on machine i, $i = 1, 2$. Consider a specific vector of outputs $\boldsymbol{x} = \boldsymbol{\alpha} = (\alpha_A, \alpha_B) = (40, 200)$. It is obvious that in this case the minimum cost is obtained when $y_{A1} = y_{B2} = 0$ and $y_{A2} = 40$, $y_{B1} = 200$; that is, all units of A are processed on machine 2 and all units of B on machine 1. The corresponding minimum cost is:

$$F(40, 200) = 40 \cdot 10 + 200 \cdot 3 = \$1000 \quad .$$

This optimal production mix, where each product is processed on a different machine, immediately suggests a direct costing approach that will impute \$400 to A and \$600 to B, or \$10 per unit of A and \$3 per unit of B. The question is, however, whether these figures reflect the real contributions of A and B to the total cost. The answer is certainly no. Product A can be processed on each of the two machines faster than B and at a unit cost which is lower than that of B. The minimum cost mix assigns A to 2 and B to 1 because the penalty paid if B is not processed on machine 1 is much higher than that paid if A is not processed on this machine. It is only the existence of a joint product B that makes it more expensive to produce A. Thus, it seems reasonable that the real contribution of a unit of output A to the total cost be smaller than the contribution of a unit of B.

Now, compute the Shapley value corresponding to the two-player (A and B) goods game, i.e., $\bar{v}(\{A\}) = F(40,0) = \min\{y_{A1} + 10y_{A2}\}$, subject to

$$\begin{aligned} y_{A1} + y_{A2} &\geqq \alpha_A = 40 \quad , \\ y_{A1} &\leqq 400 \quad , \\ 10y_{A2} &\leqq 800 \quad , \\ y_{A1}, y_{A2} &\geqq 0 \quad . \end{aligned}$$

Hence $\bar{v}(\{A\}) = \$40$. Similarly, $\bar{v}(\{B\}) = F(0,200) = \$600$, and $\bar{v}(N) = \bar{v}(\{A, B\}) = F(40,200) = \$1000$. Thus, the Shapley value (s_A, s_B) of A and B, respectively, is

$$\begin{aligned} s_A &= [\bar{v}(\{A\}) + \bar{v}(\{A, B\}) - \bar{v}(\{B\})]/2 \\ &= (40 + 1000 - 600)/2 = \$220 \quad , \end{aligned}$$

and

$$s_B = [\bar{v}(\{B\}) + \bar{v}(\{A, B\}) - \bar{v}(\{A\})]/2 = \$780 \quad .$$

Denoting by (p_A^s, p_B^s) the resulting cost sharing prices of single units of A and B respectively,

$$\begin{aligned} p_A^s &= s_A/\alpha_A = 220/40 = \$5.5/\text{unit of } A \quad , \\ p_B^s &= s_B/\alpha_B = 780/200 = \$3.9/\text{unit of } B \quad , \end{aligned}$$

Consequently, the per unit contribution of A obtained in this way is still greater than that of B, but closer to our intuitive notion of the relationship between the costs of A and B than the direct costing approach suggested above.

Hence, in this example the Shapley value for the goods game does not seem to reflect the real contribution of these goods to the total cost $F(\boldsymbol{\alpha})$. Moreover, the Shapley value fails, in general, to satisfy a desirable property, namely, "consistency". (This concept of consistency is different from that defined in Chapter 1.) Again this can be demonstrated using the above example. Suppose that good B is technically split into two goods denoted by C and D; that is, there is a new cost function G of producing the three goods A, C, and D, given by $G(\alpha_A, \alpha_C, \alpha_D) = F(\alpha_A, \alpha_C + \alpha_D) = F(\alpha_A, \alpha_B)$. Suppose that $\alpha_A = 40$, as before, $\alpha_C = 10$, and $\alpha_D = 190$. Thus, $\alpha_C + \alpha_D = 200 = \alpha_B$. One would expect that the contribution of a unit of C plus a unit of D will be equal to the contribution of a unit of B calculated before the splitting took place (that is, $p_C^s + p_D^s = p_B^s$). However, as may be verified, $s_A = \$190$, $s_C = \$90$ and $s_D = \$720$, implying that $p_A^s = \$4.75$/unit, $p_C^s = \$9$/unit, and $p_D^s = \$3.79$/unit. This compares with the previous price for B of $p_B^s = \$3.9$/unit. Note also that even p_A^s is changed from \$5.5/unit to \$4.75/unit. Thus a technical manipulation results in a different per unit contribution not only for the good that was split but, in general, for all goods.

The basic reason that the finite goods game fails to result in a satisfactory allocation is that it ignores important relevant cost information. The Shapley value of the game $\bar{v}$ is based only on the cost of producing a finite number (2^m) of the vectors α^S for all possible coalitions S in M. The cost of producing vectors $x_1, x_2, \ldots, x_m$, where $0 \leqq x \leqq \alpha$, is not considered and therefore does not affect the Shapley value. That is, if the value of $F(x)$, for some x, $0 \leqq x \leqq \alpha$, which does not coincide with an α^S, is changed, this does not alter the Shapley values. On the one hand, the method is clearly advantageous since the amount of information needed for the Shapley value is relatively small. (Actually, in many applications the cost functions are piecewise linear, in which case the production cost of only a finite number of different output vectors is sufficient to recover the entire cost function F.) On the other hand, since the resulting imputation from the Shapley value is based on limited information, distortions, like those demonstrated by the above example, do occur. As is described below, to capture all the cost information a nonatomic game structure must be used. For such a game, v, it is possible to obtain an imputation analogous to the Shapley value corresponding to the finite goods game $\bar{v}$. This imputation results in AS prices. Moreover, in many cases it is easier to calculate AS prices rather than the Shapley prices resulting from the finite goods game $\bar{v}$.

In order to capture all of the relevant cost information, we examine coalitions of players corresponding to any output vector $x = (x_1, x_2, \ldots, x_m)$ satisfying $0 \leqq x \leqq \alpha$, where α is the actual vector of quanitities produced. To this end, consider a game in which every *unit* of each good is a different player. If the goods are infinitely divisible,[1] a game with an infinite number (or a continuum) of players is implied. If each infinitesimal unit of a good makes a negligible (nonatomic) contribution to the total cost then a nonatomic game (see Aumann and Shapley, 1974) is implied. The uniform average of the marginal contributions of each infinitesimal unit of each good to the cost of all perfectly sampled coalitions is its AS price. The perfectly sampled coalitions are those coalitions represented by vectors proportional to the vector α representing the grand coalition. Each such coalition is thus of the form $(t\,\alpha_1, t\,\alpha_2, \ldots, t\,\alpha_m)$, where t is a real number with $0 \leqq t \leqq 1$. Thus the AS price of a unit of the ith good, denoted by p_i, is

$$p_i = \int_0^1 \frac{\partial F}{\partial x_i}(t\,\alpha_1, t\,\alpha_2, \ldots, t\,\alpha_m)\,dt, \quad i = 1, \ldots, m \quad . \tag{5}$$

Here $\partial F / \partial x_i(t\,\alpha)$ is the partial derivative of F with respect to x_i (or equivalently, the marginal cost of the ith output) evaluated for an output level of $t\,\alpha$. Here t varies between 0 and 1, yielding the integration path which is the diagonal $[0, \alpha]$.

For an intuitive interpretation of equation (5), assume that to obtain the output level α, we start from zero outputs and increase all quantities of outputs uniformly so that their relative proportions are constant and equal to the proportions of the vector α. In this way a homogeneous production expansion path or a ray in the products space starting from zero and terminating at α is followed. Suppose that for each incremental increase along the expansion path, the ith output is charged its marginal contribution to the cost due to that increment. Then, once α is obtained, the α_i units of commodity i are charged their total contributions to the cost, obtained by summing (integrating) these incremental charges. The result is the total contribution of the α_i units of the ith output to the total cost. Dividing this figure by the number of units α_i we obtain the average contribution p_i given in equation (5).

The formula for AS prices provides an answer to a problem raised by Thomas (1974, p 144; 1980, p 240) concerning the nonuniqueness of expansion paths that can be chosen in order to determine the contribution of the factors of production to net revenue. Thus, the use of concepts from the thoery of nonatomic games provides a unique expansion path. Moreover, AS prices and, consequently, the uniqueness of the expansion path can be derived using an axiomatic approach (the subject of the next section) that does not rely on the notions or concepts of the theory of games. Using this approach AS prices can be thought of as the natural extension of average costs to the case of joint production of several outputs. We only point out here that the cost is fully allocated by the prices in equation (5), that is,

$$p_1\alpha_1 + p_2\alpha_2 + \cdots + p_m\alpha_m = F(\alpha_1, \alpha_2, \ldots, \alpha_m) \quad .$$

Turning now to the calculation of the AS prices given in equation (5), note first that although expression (5) uses the partial derivatives of F there is no need, as far as the actual calculations are concerned, to impose the requirement that these derivatives vary continuously or, equivalently, that the cost function be smooth. In fact, it is possible to carry out a numerical integration process in equation (5) in case there is a finite number of discrete changes (jumps) in the marginal costs along the expansion path. The existence of jumps is a property of many cost and profit functions; in particular, many linear programming cost and profit functions have this property. In these cases the calculation of AS prices becomes relatively simple because the cost or profit functions are piecewise linear; that is, composed of different linear pieces combined to give a continuous function. For every such linear piece the slope or marginal cost is constant. Following the homogeneous expansion path from 0 to $\boldsymbol{\alpha}$ along $t\,\boldsymbol{\alpha}$ as t varies between 0 and 1, a finite number of linear pieces is encountered. On these linear pieces the (constant) marginal cost is calculated as well as the range of values of t (both lower and upper bounds) for which this marginal cost remains constant. Then, these constant marginal costs are weighted by the differences between the corresponding upper and lower bounds on t and summed. The result is the AS price.

The marginal costs and the values of t for which they exhibit a jump can be calculated using a specialized parametric linear programming process. The underlying theory and methodology for this is well established in the mathematical programming literature and can be found in almost every linear programming textbook (e.g., Bradley *et al.*, 1978). See also Samet *et al.* (1984) for a description of a parametric programming algorithm for computing AS prices for transportation problem cost functions. Note, finally, that the use of parametric programming for cost accounting allocation problems was suggested by Wright (1970) and Thomas (1974, pp 141–4).

Returning to Example 1 we now calculate the AS prices. To that end, we compute the marginal costs of $F(t\,40, t\,200)$ as t ranges from 0 to 1. It can easily be seen that if $t \leqq 10/11$, the capacity of machine A (the cheaper one) suffices to process both outputs. Thus the marginal costs of A and B are \$1 and \$3 per unit, respectively. That is, additional units of A and B are processed on machine 1 at a cost of \$1 and \$3, respectively. For $t > 10/11$ an additional unit of A is processed on machine 2 at a marginal cost of \$10, while an additional unit of B is processed on machine 1 and costs \$3, but releases 2 units of A which can be processed on machine 2, costing a net of $2(10) - 2(1) = \$18$. Thus, the marginal cost of B is \$21. Note that if an additional unit of B were processed on machine 2 it would cost \$25. In this sense, good A subsidizes good B since the prodution of good A shifts to the higher-priced process, minimizing the total cost. Finally, the AS prices are:

$$p_A = 1\left[\frac{10}{11}\right] + 10\left[\frac{1}{11}\right] = \frac{20}{11} = \$1.818/\text{unit, and}$$

$$p_B = 3\left[\frac{10}{11}\right] + 21\left[\frac{1}{11}\right] = \frac{51}{11} = \$4.636/\text{unit}$$

This is a considerably more acceptable allocation than obtained using the other approaches, since the per unit contribution of output A is much less than the value obtained for B. In particular, this approach yields an accurate index of the actual cost of producing each good which does not ignore the fact that good A is subsidizing good B. Notice that these values actually allocate the total cost since

$$p_A \alpha_A + p_B \alpha_B = (1.818)40 + (4.636)200 = \$1000 \quad .$$

To conclude this section we also demonstrate how the contributions of the factors of production (inputs) to profit may be determined. In this case the function $F(\boldsymbol{x})$ is the profit function and $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ is the input vector. Expression (5) then yields the per unit contribution of each input to the total profits. A *linear programming profit function* is of the form

$$F(\boldsymbol{x}) = \max_{\boldsymbol{y}} \{c_1 y_1 + c_2 y_2 + \cdots + c_l y_l\}$$

subject to

$$\mathbf{A}\boldsymbol{y} \leqq \boldsymbol{x} \quad ,$$

$$\mathbf{B}\boldsymbol{y} \leqq b \quad ,$$

$$\boldsymbol{y} \geqq 0 \quad ,$$

where $\boldsymbol{x}$ is the vector of inputs (available, say, in inventory), $\boldsymbol{y}$ is the vector of final outputs, c_i, $i = 1, \ldots, l$, denotes the net revenue from selling one unit of y_i, and **A**, **B** are the same matrices as in the definition of a linear programming cost function. The net revenues take into account all the costs involved in transforming inputs x into outputs y. The constraints $\mathbf{A}\boldsymbol{y} \leqq \boldsymbol{x}$ reflect the technological relationships between inputs and outputs, while the constraints $\mathbf{B}\boldsymbol{y} \leqq b$ may express some other restrictions, such as market size, machine capacities, etc.

Example 2

Two inputs (possibly in inventory), whose quantities are denoted by x_1 and x_2, are used to produce three outputs whose quantities are denoted by y_1, y_2, and y_3. The net profits of the three outputs are \$100, \$200, and \$10 per unit of y_1, y_2, and y_3, respectively. These profits assume constant selling prices. Each unit of y_1 requires one unit of x_1, while a unit of y_3 requires one unit of x_2. A unit of y_2 requires 1/2 unit of x_1 *and* one unit of x_2. It is impossible to sell more than 100 units of y_2. The resulting maximum profit is given by:

$$F(x_1, x_2) = \max\{100y_1 + 200y_2 + 10y_3\}$$

subject to

$$y_1 + \frac{y_2}{2} = x_1 \quad ,$$

$$y_2 + y_3 = x_2 \quad ,$$

$$y_2 \leqq 100 \quad ,$$

$$y_1, y_2, y_3 \geqq 0 \quad .$$

Suppose that $\boldsymbol{x} = \boldsymbol{\alpha} = (101, 101)$ is given and that we wish to determine the contribution of a unit of x_1 and x_2 to the total profit. For these given inputs, the optimal production mix is 51 units of y_1, 100 units of y_2, and 1 unit of y_3 yielding total profits of \$25,110.

The marginal approach for determining contributions of factors of production to profits suggests that the relative contributions of x_1 and x_2 should be measured according to the ratios of their marginal profits (which are the linear programming optimal dual variables corresponding to the constraints whose right-hand sides are x_1 and x_2). It is obvious that an extra unit of x_1 is used for y_1, hence its marginal profit is \$100, while an extra unit of x_2 is used for y_3, at a \$10 marginal profit. Thus the per unit contribution of x_1 and x_2 should be in the ratio 10:1. However, a closer examination reveals that only the last units consumed of x_1 and x_2 contribute to the profit in this proportion. For $x_2 < 100$, an extra unit of x_1 is still used for y_1 at \$100 marginal contribution. But, an extra unit of x_2 is used for y_2, at a \$200 contribution, together with an additional 1/2 unit of x_1, taken from those used for y_1 at a loss of \$50. Thus, the resulting marginal contribution of x_2 is \$150 and the ratio of the marginal contributions of x_1 and x_2 is 2:3 (as compared to the 10:1 ratio for the last units of x_1 and x_2). Consequently, since the 2:3 ratio prevails for most values of x_1 and x_2, we expect the ratio of the contributions to be very close to 2:3 and not to 10:1.

Just for comparison, consider the finite game in which each input is a player (i.e., there are two players). Omitting the calculations, the Shapley value is (\$17000, \$8010) for the two inputs, respectively. This yields the per unit contributions of (\$169.3, \$79.3) with a 2.13:1 contribution ratio which is still unsatisfactory. Now let us calculate the AS prices of x_1 and x_2 at $\boldsymbol{x} = \boldsymbol{\alpha} = (101,101)$. Consider $F(t\,101, t\,101)$ as t ranges between 0 and 1 and determine the marginal profits along the expansion path. Following the above discussion, if $0 \leqq t \leqq 100/101$, and $x_2 \leqq 100$, the marginal profits are \$100 and \$150 for x_1 and x_2, respectively. Similarly, if $100/101 \leqq t \leqq 1$ the marginal profits are \$100 for x_1 and \$10 for x_2. The resulting AS values are

$$p_1 = \left(\frac{100}{101}\right)100 + \left(\frac{1}{101}\right)100 = \$100/\text{unit},$$

$$p_2 = \left(\frac{100}{101}\right)150 + \left(\frac{1}{101}\right)10 = \$148.61/\text{unit},$$

and the ratio of these contributions is very close to 2:3. Note that, unlike the values of the final marginal costs (100,10), these values do share the profit; that is:

$$p_1\alpha_1 + p_2\alpha_2 = (100)101 + (148.61)101 = \$25110 \quad .$$

There are two points that should be made here. First, for an optimal decision it is the marginal values that should be used because they yield the correct allocation of resources for maximizing joint profits. Second, implicit in marginal valuations is a cross-subsidy that *must* be present when maximizing profits. However, if it is required to know the part of the profit of each unit of an input, the suggested method is to use the AS price mechanism. The properties characterizing this mechanism are discussed in the next section.

3 The Axiomatic Approach to AS Prices

In this section we review the axiomatic approach to the allocation of variable costs[2] through cost sharing prices due to Billera and Heath (1982) and Mirman and Tauman (1982). This approach results in AS prices.

Our viewpoint, however, is a little different and captures aspects other than the one dealt with by Billera *et al.* (1981). Consider first the special case in which the cost function F is separable. That is, the cost of producing α_i units of good i is given by $F_i(\alpha_i)$ and consequently

$$F(\alpha_1, \alpha_2, \ldots, \alpha_m) = F_1(\alpha_1) + F_2(\alpha_2) + \cdots + F_m(\alpha_m) \quad . \tag{6}$$

Clearly, in this case the ith good is responsible for the part $F_i(\alpha_i)$ of the total cost and should have the price of $F_i(\alpha_i)/\alpha_i$; that is, its average cost. Thus, average cost plays a crucial role in attributing costs to outputs if costs are separable. Unfortunately, in general, costs are not separable and it is not clear how to determine that part of the costs which are attributable to each good, or equivalently, how to define the average cost of each good.

The goal of this section is to develop an average cost pricing mechanism (or formula) applicable to the general case of jointly producing several goods. Such a mechanism, when available, will be the natural allocation tool in cost allocation problems when the real contributions of outputs to costs are needed. The general notion for these generalized average costs is based upon several fundamental properties satisfied by the average cost concept for the single-output case. Imposing these properties as requirements (or axioms) yields this generalized average cost mechanism for the multiple-output case.

First, consider the case of producing x units of a *single good* with production cost $F(x)$. The resulting average cost is denoted by $\mathrm{AC}(F,x)$. This notation takes account of the fact that average cost depends on the quantity being produced as well as on the cost function F. Clearly $\mathrm{AC}(F,x) = F(x)/x$. This single-output average costing formula (or mechanism) satisfies the following four elementary properties:

(1) *Cost Sharing.* The price $\mathrm{AC}(F,x)$ satisfies

$$x\,\mathrm{AC}(F,x) = F(x)$$

for all values of x.

(2) *Additivity.* Suppose that the cost function F can be split into two components, say F_1, the cost of production, and F_2, the cost of marketing (i.e., $F = F_1 + F_2$). Then, for a given quantity x the average cost is the sum of the average production cost and the average marketing cost; that is, for each x,

$$\mathrm{AC}(F,x) = \mathrm{AC}(F_1,x) + \mathrm{AC}(F_2,x) \quad .$$

(3) *Positivity.* If increasing production results in higher costs then the average cost is nonnegative. Namely, if F is a nondecreasing function, then for each x,

$$\mathrm{AC}(F,x) \geqq 0 \quad .$$

(4) *Rescaling.* The average cost is independent of the units of measurement. Suppose that F is the cost function of producing an output measured in kilograms. Let G be the cost function of the same product, but measured in metric tons. Clearly

$$G(x) = F(1000x) \quad .$$

Then the average cost per ton $\mathrm{AC}(G,x)$ is 1000 times the average cost per kilogram, $\mathrm{AC}(F,1000x)$. In general for any positive scaling factor λ, if $G(x) = F(\lambda x)$ then

$$\mathrm{AC}(G,x) = \lambda \mathrm{AC}(F,\lambda x) \quad .$$

These four properties of the average cost for the single-product case are properties which it is desirable for average costs to satisfy. Therefore, it seems reasonable to require that a multiproduct extension of average cost satisfy these properties. Indeed, let F be the production cost of a multiproduct firm; that is, $F(\boldsymbol{x}) = F(x_1,x_2, \ldots, x_m)$. Note that if F is separable [i.e., given by equation (6)], then these four properties are satisfied, componentwise, for each good only by the average cost $\mathrm{AC}_i(F_i, x_i) = F_i(x_i)/x_i$. Since our purpose is to define the average cost of each output in the general case, where F is not necessarily separable; namely, to obtain for each nonnegative vector $\boldsymbol{x}$ a price vector

$$\mathrm{AC}(F,\boldsymbol{x}) = [\mathrm{AC}_1(F,\boldsymbol{x}), \ldots, \mathrm{AC}_m(F,\boldsymbol{x})] \quad ,$$

where $\mathrm{AC}_j(F,\boldsymbol{x})$ measures the average cost of producing the jth output, these four properties are imposed as requirements on the generalized average cost.

Requirement 1: Cost Sharing. For each output vector $\boldsymbol{x} = (x_1,x_2, \ldots, x_m)$ average costs share the production cost; that is,

$$\boldsymbol{x}_1\mathrm{AC}_1(F,x) + \boldsymbol{x}_2\mathrm{AC}_2(F,x) + \cdots + \boldsymbol{x}_m\mathrm{AC}_m(F,x) = F(x) \quad .$$

Requirement 2: Additivity. Average costs remain invariant to technical manipulations of the production process. That is, if $F(x)$ is the total production cost which is the sum of $F_1(x)$ (assembly cost) and $F_2(x)$ (marketing cost), then

$$\mathrm{AC}_j(F,\boldsymbol{x}) = \mathrm{AC}_j(F_1,x) + \mathrm{AC}_j(F_2,\boldsymbol{x})$$

holds for $j = 1, \ldots, m$. That is, the average cost of output j is the sum of its average assembly cost and average marketing cost.

Requirement 3: Positivity. If the cost function F is increasing (i.e., it costs more to produce more), then the average costs of each good should be nonnegative:

$$\mathrm{AC}_j(F, \boldsymbol{x}) \geqq 0 \quad , \text{ for } j = 1, \ldots, m \quad .$$

Requirement 4: Rescaling. A change in the scale of measurement (or units) of a product should yield an appropriate change in its average cost. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ denote the scaling factors and define

$$G(\boldsymbol{x}) = F(\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_m x_m) \quad .$$

That is, G is the production cost corresponding to output quantities measured in the scaled units. Then,

$$\mathrm{AC}_j(G, \boldsymbol{x}) = \lambda_j \mathrm{AC}_j[F, (\lambda_1 x_1, \lambda_2 x_2, \ldots, \lambda_m x_m)] \quad \text{for } j = 1, \ldots, m \quad .$$

For the multiproduct case, however, one additional requirement connecting the single-product and the multiproduct case is needed. That is, two (or more) commodities which are the same should have the same average cost. Since average cost prices depend only on the cost function it is clear that being the same commodity means playing the same role in the cost function. As an illustration consider the production of red and blue cars. The cost of producing x_1 red cars and x_2 blue cars can be represented as a two-variable function $F(x_1, x_2)$. But, in fact, the cost of producing a red car is the same as the cost of producing a blue car. This can be formulated as follows. There is a one-variable function C for which $C(x)$ is the cost of producing a total of x cars (red ones, blue ones, or both) and

$$F(x_1, x_2) = C(x_1 + x_2) \quad .$$

In this case we require that the average cost $\mathrm{AC}_1[F, (x_1, x_2)]$ of a red car be the same as the average cost $\mathrm{AC}_2[F, (x_1, x_2)]$ of a blue car which is the average cost $\mathrm{AC}(C, x)$ of a car, where $x = (x_1 + x_2)$.

Requirement 5: Consistency. If all goods are the same, that is, if there exists a cost function C such that

$$F(x_1, x_2, \ldots, x_m) = C(x_1 + x_2 + \cdots + x_m) \quad ,$$

then

$$\mathrm{AC}_1(F, \boldsymbol{x}) = \mathrm{AC}_2(F, \boldsymbol{x}) = \cdots = \mathrm{AC}_m(F, \boldsymbol{x}) = \mathrm{AC}(C, x_1 + x_2 + \cdots + x_m) \quad .$$

Bearing in mind that the desired multiproduct extension of average cost should satisfy the above five requirements, a pricing mechanism is sought which satisfies these axioms universally for as many cost functions as possible. This leads to the following theorem.

Theorem 1. *The expression*

$$\mathrm{AC}_i(F,\boldsymbol{x}) = \int_0^1 \frac{\partial F(tx_1, tx_2, \ldots, tx_m)}{\partial x_i}\,\mathrm{d}t \tag{7}$$

is the one and only one satisfying Requirements 1–5 for all continuously differentiable cost functions F with $F(0) = 0$.

Thus, AS prices introduced in Section 2 have been obtained as the *only* prices satisfying the five requirements universally for a considerable family of cost problems. It follows that AS prices can be viewed as a natural extension of the average cost concept. It is worth mentioning that AS prices (or AS average costs) coincide with the standard average costs in the case where $F(x_1,x_2, \ldots, x_m)$ is separable and with the marginal costs if the production technology exhibits constant returns to scale; that is, if $F(\lambda\boldsymbol{x}) = \lambda F(\boldsymbol{x})$.

Several extensions and applications of the above approach should be noted. A recent different axiomatic derivation of the AS average cost prices was suggested by Young (1985; see also Chapter 1). The universality of these prices was further established by Samet *et al.* (1984) where it is shown that AS average costs are the only prices satisfying Requirements 1–5, even for certain families of cost functions that are not necessarily differentiable and including some piecewise linear ones resulting from linear programming cost problems. Samet and Tauman (1982) discuss the effect of relaxing the cost allocation requirement (Requirement 1), so that they could study various pricing mechanisms. In particular, they characterized marginal cost prices using a set of requirements that are similar, although somewhat different, to Requirements 2–5. They also discuss other intuitive requirements, in particular, replacing the additivity requirement with a separability requirement. Economic equilibrium models where AS average costs are shown to be compatible with consumers' demand are discussed by Mirman and Tauman (1982) and Samet *et al.* (1984). The importance of AS average costs in certain merging situations and in the determination of equilibrium industry structure is emphasized in Mirman *et al.* (1982a,b). Finally, extending the AS average costing formula to handle fixed costs as well as variable costs is the subject of the paper by Mirman *et al.* (1983). Owing to the importance of this issue we discuss it in detail in the next section.

4 The Allocation of Fixed and Variable Costs

This section is concerned with finding the contributions (or average costs) of outputs produced jointly using a technology containing a fixed cost component. Consider the production of quantities $\alpha_1, \ldots, \alpha_m$ of m goods with the total production cost given by $G(\alpha_1, \alpha_2, \ldots, \alpha_m)$ where

$$G(x_1, x_2, \ldots, x_m) = C + F(x_1, x_2, \ldots, x_m) \quad . \tag{8}$$

The fixed cost C is a positive constant, independent of the output level and the variable cost. Here the variable cost is given by F; that is, $F(0, 0, \ldots, 0) = 0$.

If expression (7) is applied to the function G given by equation (8) and the output vector $\boldsymbol{\alpha} = (\alpha_1,\alpha_2, \ldots, \alpha_m)$; that is,

$$\mathrm{AC}_i(G,\alpha) = \int_0^1 \frac{\partial G(t\,\alpha_1,\, t\,\alpha_2,\, \ldots,\, t\,\alpha_m)}{\partial x_i}\,\mathrm{d}t \quad ,$$

then, since the marginal costs of G are exactly those of F, the resulting prices only cover the variable portion $F(\alpha)$ of the cost. It is thus clear that a major difficulty with the approach of the previous sections is that it is applicable only to cost functions having no fixed cost component. In fact, an average costing formula [for cost functions of the type (8)] that satisfies Requirements 1–5 of the last section implies a contradiction. Hence, no formula simultaneously satisfies all five requirements. To illustrate the difficulty, we consider the following example.

Example 3

Consider the production of two goods, $\alpha_1 = \alpha_2 = 1$, with the cost function

$$G(x_1,x_2) = x_1 + x_2 + C \quad .$$

The consistency requirement immediately implies that the two goods should have equal average costs, and hence

$$\mathrm{AC}_1[G,(1,1)] = \mathrm{AC}_2[G,(1,1)] = 1 + \frac{C}{2} \quad . \tag{9}$$

Letting $G(x_1,x_2) = G_1(x_1,x_2) + G_2(x_1,x_2)$, where $G_1(x_1,x_2) = x_1$ and $G_2(x_1,x_2) = C + x_2$, then the average costs for G_1 and G_2, respectively, are

$$\mathrm{AC}_1[G_1,(1,1)] = 1 \qquad \mathrm{AC}_1[G_2,(1,1)] = 0$$

$$\mathrm{AC}_2[G_1,(1,1)] = 0 \qquad \mathrm{AC}_2[G_2,(1,1)] = 1 + C \quad .$$

Then to satisfy the additivity requirement

$$\mathrm{AC}_1[G,(1,1)] = \mathrm{AC}_1[G_1,(1,1)] + \mathrm{AC}_1[G_2,(1,1)] = 1$$

$$\mathrm{AC}_2[G,(1,1)] = \mathrm{AC}_2[G_1,(1,1)] + \mathrm{AC}_2[G_2,(1,1)] = 1 + C \quad .$$

This, however, contradicts equation (9) since it implies that the price of the second commodity covers all the fixed cost.

In this section we analyze and suggest solutions to various cost allocation problems involving fixed costs. However, we first note that a distinction must be made between the long-run cost function and the short-run cost function. The long-run cost function is the lower envelope of the short-run cost function. In other words, the long-run cost function $H(x_1, x_2, \ldots, x_m)$ is the minimum cost, over all possible technologies, of producing a given vector $\boldsymbol{\alpha}$ of outputs. The short-run cost function $G(x_1, x_2, \ldots, x_m)$ given by equation (8) thus coincides with the long-run cost function at a certain level (or levels) of output. This short-run cost function is efficient for producing these quantities, but is not efficient for producing all output levels. The relationship between the long-run and short-run cost function is illustrated for the one good case in Figure 3.1. This figure demonstrates the well known convention that long-run cost functions, in general, do not have a fixed cost component. But the implicit optimal short-run technology for producing a particular vector of outputs has, in general, a fixed cost component, which in some circumstances must be allocated among the outputs. Moreover, there are even some long-run cost functions that do have legitimate fixed cost components (e.g., transporting Moon dust to Earth).

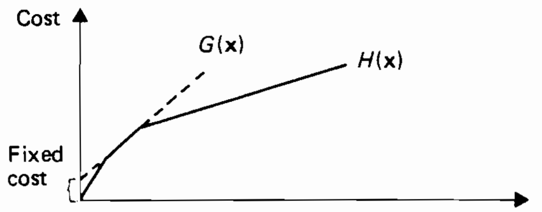

Figure 3.1

Accordingly, two possible cases are considered. The first concerns production of $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_m)$ using the best possible technology for producing this output level. This implies that the long-run cost function is used; hence the cost is given by $H(\alpha_1, ..., \alpha_m)$. When this long-run cost function does not include a fixed cost component, the AS average costs formula (7) can be applied to allocate these costs. On the other hand, it is the short-run cost function (8) that is actually used by the firm to calculate the cost of producing $\boldsymbol{\alpha}$. If the vector $\boldsymbol{\alpha}$ is optimally produced by the firm this short-run cost function, which, in general, contains a fixed cost component, coincides with the long-run cost function $\boldsymbol{\alpha}$ (or perhaps at a neighborhood of $\boldsymbol{\alpha}$). That is,

$$H(\alpha_1, ..., \alpha_m) = G(\alpha_1, ..., \alpha_m) = F(\alpha_1, ..., \alpha_m) + C \quad . \tag{10}$$

Hence the AS average costs which allocate long-run costs also indirectly allocate the short-run costs including the fixed cost of producing $\boldsymbol{\alpha}$. We now show what part of the AS prices associated with the long-run technology H at $\boldsymbol{\alpha}$ is allocated to the fixed cost C and what part is allocated to the variable cost F of the short-run technology at the output vector $\boldsymbol{\alpha}$. To that end, expression (7) is applied to the function H, which yields the cost sharing prices $p_1, p_2, ..., p_m$ given by

$$p_i = \int_0^1 \frac{\partial H(t\,\alpha_1, t\,\alpha_2, ..., t\,\alpha_m)}{\partial x_i} \mathrm{d}t \quad . \tag{11}$$

Since $H(\boldsymbol{\alpha}) = F(\boldsymbol{\alpha}) + C$, at $\boldsymbol{\alpha}$, these prices cover both the variable and fixed costs of the short-run cost function; that is,

$$p_1\alpha_1 + \cdots + p_m\alpha_m = F(\alpha_1, ..., \alpha_m) + C \quad . \tag{12}$$

Consider the set of prices $\hat{p}_1, \hat{p}_2, ..., \hat{p}_m$ that are due only to the variable cost function F

$$\hat{p}_i = \int_0^1 \frac{\partial F(t\,\alpha_1, t\,\alpha_2, ..., t\,\alpha_m)}{\partial x_i} \mathrm{d}t \quad . \tag{13}$$

These prices cover the variable cost $F(\boldsymbol{\alpha})$ of producing α, that is,

$$\hat{p}_1\alpha_1 + \cdots + \hat{p}_m\alpha_m = F(\alpha_1, ..., \alpha_m) \quad . \tag{14}$$

Obviously, from equations (12) and (14)

$$(p_1 - \hat{p}_1)\alpha_1 + \cdots + (p_m - \hat{p}_m)\alpha_m = H(\boldsymbol{\alpha}) - F(\boldsymbol{\alpha}) = C \quad .$$

This means that $p_i - \hat{p}_i$ is that part of the price p_i which may be thought of as covering the fixed cost C. This allocation of the fixed costs is illustrated by the following example.

Example 4

Consider the production of two goods and assume that there are only two available technologies for producing these goods. Let G_1 and G_2 be the cost functions associated with the two technologies, and assume that

$$G_1(x_1,x_2) = 3x_1 + 2x_2 = F_1(x_1,x_2)$$

and

$$G_2(x_1,x_2) = x_1 + x_2 + 9 = F_2(x_1,x_2) + 9 \quad .$$

Since $G_1 \leqq G_2$ whenever $2x_1 + x_2 \leqq 9$ the long-run minimum cost $H(x_1,x_2)$ of producing the output level (x_1,x_2) is given by

$$H(x_1,x_2) = \begin{cases} 3x_1 + 2x_2 \ , & \text{if } 2x_1 + x_2 \leqq 9 \\ x_1 + x_2 + 9 \ , & \text{otherwise} \ . \end{cases}$$

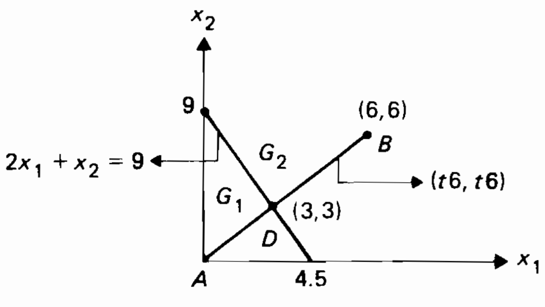

Figure 3.2

Consider now the production of 6 units of each commodity. Namely, α = (6,6). The total production cost is then $H(6,6) = 6 + 6 + 9 = 21$. For this output level the short-run technology used is G_2. Following the preceding discussion, consider the allocation of the total cost of 21 among the outputs, in particular, the allocation of the short-run fixed cost of 9. First allocate the long-run cost by calculating expression (11). Note that $H(x_1,x_2)$ is piecewise linear and thus, the calculation of AS average costs resembles the calculation in Examples 1 and 2. For each value of t, $0 \leqq t \leqq 1$, $(t\,\alpha_1, t\,\alpha_2) = (t6, t6)$ denotes an output combination on the line AB in Figure 3.2. Note also, from Figure 3.2, as long as $t \leqq 0.5$, $(t\,\alpha_1, t\,\alpha_2) = (t6, t6) \leqq (3,3)$; that is, only outputs on the line AD are considered. For these outputs only values in the region where the long-run costs H coincide with G_1 are needed. Hence the marginal costs of H (or its partial derivatives) are those of G_1. Since G_1 is linear, those marginal costs are 3 and 2 for the first and second outputs, respectively. In the same way, for $t \geqq 0.5$ only outputs on the line DB (i.e., where H coincides with G_2) need be considered. Consequently, the marginal costs of H are 1 in this region for both the first and second outputs. Now, for each output, equation (11) implies that the marginal costs are added, corresponding to the various regions and weighted by the differences of the t values corresponding to these regions. This yields

$$p_1 = (0.5 - 0)3 + (1 - 0.5)1 = 2 \ , \text{ and}$$

$$p_2 = (0.5 - 0)2 + (1 - 0.5)1 = 1.5$$

Note that equation (12) holds,

$$2.6 + (1.5)6 = 21 = H(6,6) \quad ,$$

and hence long-run costs are allocated. To determine how the fixed cost of 9 is actually allocated, we evaluate equation (13). Note that the output level of (6,6) is in the region where long-run costs coincide with G_2. Hence the relevant short-run variable cost function $F_2(x_1,x_2) = x_1 + x_2$ is the one to be substituted for F in equation (13). Since F_2 is linear,

$$\hat{p}_1 = \frac{\partial F_2}{\partial x_1} = 1 \quad ,$$

$$\hat{p}_2 = \frac{\partial F_2}{\partial x_2} = 1 \quad .$$

Hence the parts of p_1 and p_2 that are used to cover the fixed cost of 9 units are

$$p_1 - \hat{p}_1 = 2 - 1 = 1$$

$$p_2 - \hat{p}_2 = 1.5 - 1 = 0.5$$

Thus, while the price vector (2,1.5) covers the total cost of producing the output level (6,6), one-half of the price of the first good and one-third of the second good may be attributed to covering the fixed cost, that is, $[1(6) + 0.5(6) = 9]$.

We now consider situations that require a different approach. The first situation is when the long-run cost function has a fixed cost component and thus expression (7) cannot be applied. Another situation requiring a different approach is when the producer is unable to use the optimal short-run technology for the actual production level. In this case the short-run cost deviates from the long-run cost and therefore the allocation of the long-run cost is irrelevant. Thus, the AS prices that allocate long-run costs are not applicable in this case. Moreover, AS prices cannot be used to allocate the short-run costs directly since, in general, the short-run cost function contains a fixed cost component. In both cases there is a need to allocate both fixed and variable costs. Hence, consider the problem of allocating costs given by equation (8) when $\alpha_1, \ldots, \alpha_m$ are produced. As has already been shown with Example 3, it is impossible to allocate such costs in a way that satisfies Requirements 1–5 simultaneously. It turns out that it is possible to modify the additivity requirement (Requirement 2) so that it becomes compatible with the other requirements. To motivate this change, consider first two intuitive formulas for allocating fixed costs. According to the first one every unit of output bears the same share in the fixed cost. The share is

$$C / (\alpha_1 + \alpha_2 + \cdots + \alpha_m)$$

It is easily seen that this expression strongly depends on the definition of a unit of a commodity. If, for example, the units of the first good are scaled the amount α_1 changes, and hence the total number of units is altered. Consequently a new per unit charge results. This, however, means that the above formula violates the rescaling requirement. To overcome this problem consider the following charge to a unit of good i:

$$C/(m\alpha_i) \quad ;$$

that is, each good bears the same share in the fixed cost and this share is equally divided among the units of that good. This allocation satisfies the rescaling requirement, but fails to satisfy the consistency requirement. This happens since one can manipulate the above charges by splitting a commodity into two irrelevant classifications. For example, consider again the production of red cars and blue cars in amounts α_1 and α_2 where $\alpha_1 \neq \alpha_2$. By the consistency requirement the charges to red cars and blue cars must be the same. But the last formula treats these two types of cars unequally by imposing a higher proportion of the fixed costs on the cars that are produced in the smaller amount.

In fact it can easily be shown that *each* price mechanism which allocates the fixed cost independently of its variable cost violates either the rescaling or the consistency requirements. We now show how to modify the additivity requirement so that the resulting allocation of the fixed cost depends on the variable cost. Suppose that the cost function $G(\alpha_1, \ldots, \alpha_m)$ is given by equation (8) where $F(\alpha_1, \ldots, \alpha_m)$ is the variable cost and C is the fixed cost. Also assume that the variable cost is given by the sum of $F_1(\alpha_1, \ldots, \alpha_m)$ and $F_2(\alpha_1, \ldots, \alpha_m)$, that is

$$F(\alpha_1, \ldots, \alpha_m) = F_2(\alpha_1, \ldots, \alpha_m) + F_2(\alpha_1, \ldots, \alpha_m) \quad .$$

It remains to attribute the fixed cost to the two additive components of the variable cost. Actually, it is not necessary to specify, *a priori, how* the fixed cost is split between the two variable cost components. It is sufficient to require only that there *exists* a way to do it! Let us require that the fixed cost be split into two parts C_1 and C_2 such that $C_1 + C_2 = C$, where the larger portion of the fixed cost is attributed to the larger variable cost component, namely

$$C_2 \geqq C_1 \quad \text{if} \quad F_2(\alpha_1, \ldots, \alpha_m) \geqq F_1(\alpha_1, \ldots, \alpha_m) \quad . \tag{15}$$

Moreover, let the additivity property hold. Then if the total cost is split into two components $F_1 + C_1$ (assembly) and $F_2 + C_2$ (marketing), each including a portion of the fixed cost, the additivity requirement specifies that the average cost of good i should be the sum of its average assembly cost and average marketing cost, that is

$$\mathrm{AC}_i(F + C, \alpha) = \mathrm{AC}_i(F_1 + C_1, \alpha) + \mathrm{AC}_i(F_2 + C_2, \alpha) \quad . \tag{16}$$

Note that there is no need to determine C_1 and C_2. It is only required that once the variable portion of F is additive, then there is a way of splitting the total fixed cost among the additive portions of the variable cost such that:
(2*) Inequality (15) holds and the average cost formula satisfies equation (16).

Theorem 2. (Mirman *et al.*, 1983). *The following formula is the one and only one satisfying properties 1, 2*, 3, 4, and 5 for all continuously differentiable cost functions of the form* $F(x_1, x_2, \ldots, x_m) + C$:

$$\mathrm{AC}_i[F + C, \alpha_1, \ldots, \alpha_m] = \left[1 + \frac{C}{F(\alpha_1, \ldots, \alpha_m)}\right] \int_0^1 \frac{\partial F(t\,\alpha_1, t\,\alpha_2, \ldots, t\,\alpha_m)}{\partial x_i} dt \quad (17)$$

Note that this equation can be written equivalently as

$$\mathrm{AC}_i[F + C, \alpha_1, \ldots, \alpha_m] = \int_0^1 \frac{\partial F(t\,\alpha_1, \ldots, t\,\alpha_m)}{\partial x_i} dt$$

$$\int_0^1 \frac{\partial F}{\partial x_i} + \frac{C}{F(\alpha_1, \ldots, \alpha_m)} \int_0^1 \frac{\partial F(t\,\alpha_1, \ldots, t\,\alpha_m)}{\partial x_i} dt$$

Accordingly, the first term in equation (17) is the AS average cost price corresponding to the variable cost $F(\alpha_1, \alpha_2, \ldots, \alpha_m)$. The second term allocates the fixed cost proportionally to the AS average cost.

Example 5

Consider the linear variable cost function

$$F(x_1, x_2, \ldots, x_m) = b_1 x_1 + b_2 x_2 + \cdots + b_m x_m \quad .$$

Let all α_i and C be positive. Then

$$\mathrm{AC}_i[F + C, \alpha_1, \ldots, \alpha_m] = b_i + \frac{C b_i}{F(\alpha_1, \ldots, \alpha_m)} \quad ,$$

where the second term is the part of AC_i covering the fixed cost, namely

$$\sum_{i=1}^{m} \alpha_i \frac{C b_i}{F(\alpha_1, \ldots, \alpha_m)} = C \frac{\sum_{i=1}^{m} \alpha_i b_i}{F(\alpha_1, \ldots, \alpha_m)} = C \quad .$$

Notes

1. The analysis carried out here may serve, however, as an approximation for practical situations where m indivisible goods are produced in relatively large quantities.
2. Throughout this section it is assumed that $F(0, 0, \ldots, 0) = 0$.

References

Aumann, R.J. and Shapley, L.S. (1974) *Values of Non-Atomic Games* (Princeton, NJ: Princeton University Press)

Balachandran, B.V. and Ramakrishnan, R.T.J. (1981) Joint cost allocation: a unified approach. *Accounting Review* 56: 85–96.

Billera, L.J. and Heath, D.C. (1982) Allocation of shared costs: a set of axioms yielding a unique procedure. *Mathematics of Operations Research* 7: 32–9.

Billera, L.J., Heath, D.C., and Raanan, J. (1978) Internal telephone billing rates – a novel application of non-atomic game theory. *Operations Research* 26: 956–65.

Billera, L.J., Heath, D.C., and Verrecchia, R.E. (1981) A unique procedure for allocating common costs from a production process. *Journal of Accounting Research* 19: 185–96.

Callen, J.L. (1978) Financial cost allocation: a game theoretic approach. *Accounting Review* 53: 303–8.

Dopuch, N. (1981) Some perspectives on cost allocations, in S. Moriarity (Ed) *Joint Cost Allocations*. (Norman, OK: University of Oklahoma Press) pp 1–7.

Hamlen, S., Hamlen, W. and Tschirhart, J. (1977) The use of core theory in evaluating joint cost allocation schemes. *Accounting Review* 51: 616–27.

Hamlen, S., Hamlen, W. and Tschirhart, J. (1980) The use of the generalized Shapley allocation in joint cost allocation. *Accounting Review* 55: 269–87.

Littlechild, J.C. and Thompson, G.F. (1977) Aircraft landing fees: a game theory approach. *Bell Journal of Economics* 8: 186–204.

Loehman, E. and Whinston, A. (1971) A new theory of pricing and decision making for public investment. *Bell Journal of Economics* 2: 606–25.

Loehman, E. and Whinston, A. (1976) A generalized cost allocation scheme, in J. A. Y. Lin (Ed) *Theory and Measurement of Economic Externalities* (New York: Academic Press) pp 87–101.

Mirman, L.J., Samet, D., and Tauman, Y. (1983) An axiomatic approach to the allocation of a fixed cost through prices. *Bell Journal of Economics* 14: 139–151.

Mirman, L.J. and Tauman, Y. (1981) Valeur de Shapley et repartition equitable des couts de production, in E. Malinvaud (Ed) *Les Cahiers du Seminaire d'Econometrie*, pp 121–51.

Mirman, L.J. and Tauman, Y. (1982) Demand compatible, equitable, cost sharing prices. *Mathematics of Operations Research* 7: 40–56.

Mirman, L.J., Tauman, Y. and Zang, I. (1982a) *Issues in the Theory of Contestable Markets*, mimeograph.

Mirman, L.J., Tauman, Y. and Zang, I. (1982b) *Cooperative Behavior in a Competitive Market*. Working Paper 826–84 (Tel-Aviv: Israel Institute of Business Research).

Moriarity, S. (Ed) (1981a) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press).

Moriarity, S. (1981b) Some rationales for cost allocation, in S. Moriarity (Ed) *Joint Cost Allocations* (Norman, OK: University of Oklahoma Press).

Roth, A.E. and Verrecchia, R.E. (1979) The Shapley value as applied to cost allocation: a reinterpretation. *Journal of Accounting Research* 17: 295–303.

Samet, D. and Tauman, Y. (1982) The determination of marginal-cost prices under a set of axioms. *Econometrica* 50: 895–909.

Samet, D., Tauman, Y. and Zang, I. (1984) An application of the Aumann–Shapley prices for cost allocation in transportation problems. *Mathematics of Operations Research* 9: 25–42.

Schmeidler, D. (1969) The nucleolus of a characteristic function game. *SIAM Journal of Applied Mathematics* 17: 1163–70.

Shapley, L.S. (1953) A value for n-person games, in A.W. Kuhn and A.W. Tucker (Eds) *Contributions to the Theory of Games* Vol. II (Princeton, NJ: Princeton University Press) pp 307–17.

Shubik, M. (1964) Incentives, decentralized control, the assignment of joint costs and internal prices, in C.B. Bonini, R.K. Jaedicke, and H.M. Wagner (Eds) *Management Controls: New Directions in Basic Research* (New York: McGraw-Hill).

Thomas, A.L. (1969) *The Allocation Problem in Financial Accounting*, Studies in Accounting Research No. 3 (American Accounting Association).

Thomas, A.L. (1974) *The Allocation Problem: Part II*, Studies in Accounting Research No. 9 (American Accounting Association).

Thomas, A.L. (1980) *A Behavioral Analysis of Joint-Cost Allocation and Transfer Pricing* (Urbana, IL: Stipes).

Wright, F.K. (1970) Dual variables in inventory measurement. *Accounting Review* 45: 129–33.

Young, H.P. (1985) Producer incentives in cost allocation. *Econometrica* 53.

Zimmerman, J.L. (1979) The costs and benefits of cost allocation. *Accounting Review* 54: 504–21.

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 4

THE COOPERATIVE FORM, THE VALUE, AND THE ALLOCATION OF JOINT COSTS AND BENEFITS

Martin Shubik

1 Introduction

In 1953 Lloyd Shapley published his elegant paper on the value solution of an n-person game in characteristic function form. In 1962 I suggested that Shapley's axioms could be reinterpreted in terms of accounting conventions and could be used to provide a means for devising incentive-compatible cost assignments and internal pricing in a firm with decentralized decision making (Shubik, 1962).

The problem of the assignment of joint costs (and benefits) is one that has bedevilled accountants for many years. A reaction by a microeconomist oriented towards marginal analysis may be: why bother to assign overheads or joint costs at all? (see Stigler's textbook (1966, p 165) for an example). The reason for the different attitudes and perceived needs by accountants, economists, regulators, production managers, tax collectors, divisional vice-presidents and others is that they are all looking at the same institutional entity from different viewpoints.

The accountant wants the books to balance. He wants all costs allocated. Benefits, unless they can be translated directly into money, pose difficult problems; the convention of conservatism more or less dictates that if you know that an item is of positive worth but that you cannot quantify its value, carry it out at zero or at a symbolic sum such as $1 for good will.

A tax accountant looks towards minimizing a specific evaluation, namely his clients' tax bill; a tax collector may try to maximize tax revenues collected. An economist advising on profit maximization from a given plant producing a joint product wants to make sure that the arbitrary assignment of joint costs or profits does not distort the profit maximization. If there are joint products, can an internal control system be designed that enables him to delegate decisions to others who use only the information they are sent? Or is it desirable to have the decision centralized?

Even to this day microeconomic theory is disturbingly vague about what constitutes a long-term or short-term decision. In the corporation, marketing, pricing, production, product development, minor capacity change, major investment and innovation all have different time scales. One individual's decision variables are another individual's parameters.

Accountants must produce systems that are viable, acceptable and operational – taking into account the pressures and problems of management, custom, law, economics, and the tax collector. Generally Accepted Accounting Principles do not form a rigid, monolithic set of rules to be obeyed in the same way that chess players obey the rules of chess. But they are presented as a set of guidelines for the responsible businessman and others.

The problems in cost accounting are important and applied, but simpler versions of many of the problems encountered in the study of externalities and public goods. The accounting profession does not attempt to solve all of the problems of welfare that may occur to the economist, but rather tries to provide an allocation scheme of operational worth to the institutions being served.

This paper is written from the viewpoint of economics and game theory rather than that of the professional accountant. It may be that various nuances of importance to the accounting profession are overlooked or treated in a somewhat different language. (In Chapter 2, Biddle and Steinberg address accountants' concerns more explicitly.) The thrust of this investigation is in terms of cost and revenue allocations as control mechanisms of interest to the economist, accountant, and business or public executive.

My initial interest in the joint cost problem came about in the context of consulting work with Harlan Mills for a chemical company. The first example in my paper (Shubik, 1962, p 336) was based upon the experience of what can happen if fixed overheads are allocated by several acceptable accounting methods, yet there are independent profit centers that can take action based on this information. The application was by example and was tutorial, aimed at preventing error, making clear the dangers in the assignment of overhead costs and pointing out that if you wanted to have tidy accounts there was a way that could avoid the error for overhead assignment. The same level of advice was also given to a major oil company in terms of understanding why underestimation of demand was prevalent in the reporting of independent divisions.

Any abstraction results in a distortion of reality from some point of view. Thus there is no universal all-purpose accounting scheme that can always satisfy the needs of a variety of individuals utilizing accounting schemes for different purposes. Cost and control accounting, stockholder financial reporting, and tax accounting have different constituencies and purposes. Whitman and Shubik (1979) discuss the different motivations and problems that occur just at the level of financial accounting for stockholders, bondholders, other creditors, and managers.

Once the full diversity of interested parties and their different purposes is recognized – even restricting ourselves to the allocation of costs and revenues for purposes of control and leaving aside problems involving equity or taxes – we would still need to differentiate different parties, purposes and problems. These call for special considerations in both modeling and the selection of solutions.

In the subsequent sections our particular concern will be with the uses and limitations of the characteristic function of an n-person game. This also involves an excursion into the problems posed by limits on information and by threats. Given a well defined problem and a model that is regarded as a satisfactory representation of the phenomenon being studied, a solution concept must be selected. Among the candidates are the value, the core, and the nucleolus if the problem is modeled by a characteristic function; or some variant of the noncooperative equilibrium if the problem is modeled in strategic or extensive form.

2 Game-Theoretic Modeling: The Cooperative Form

Three conceptually different forms of modeling have been suggested for interactive decision making in situations which can be described as games of finite length. They are the extensive, strategic, and cooperative forms of a game. Each can be used to describe the same game, but at a different level of detail. Thus each representation is best suited to a different class of questions and poses different levels of difficulty in mathematical analysis and in computation. Moreover, if our purpose is to apply game-theoretic analysis to answer questions concerning operational problems, the ability to actually compute solutions becomes important.

Much of the work in the application of game-theoretic methods to cost problems has been based upon the cooperative form of the game (see Shubik, 1962; Littlechild and Owen, 1973; Young ***et al.*** 1982; Hamlen ***et al.*** 1977; Billera ***et al.*** 1978). Yet, as is argued here, far more attention needs to be paid to the ***ad hoc*** aspects of the cooperative form prior to the application of a solution concept.

von Neumann and Morgenstern (1944) based their cooperative solution theory upon the characteristic function of an n-person game. This function v is a superadditive set function defined on 2^n coalitions. ***Pro forma*** we may define the worth of the empty coalition ϕ to be zero, or $v(\phi)=0$. Consider two coalitions S and T with no members in common. Superadditivity calls for

$$v(S \cup T) > v(S) + v(T) \text{ where } S \cap T = \phi \quad .$$

This is merely an economic incentive condition which assumes that there are no overall losses from cooperation. There can, in general, be gains. Even this is not an innocent assumption. If the joining together of two coalitions has any concrete institutional meaning the mechanics of coalition formation may be expensive and superadditivity is not a foregone conclusion.

The casual reader of von Neumann's and Morgenstern's classic might be surprised to see the elaborate apparatus they erected in order to calculate and justify the characteristic function. They invented an $(n+1)$-person game with a fictitious player, "Nature", who loses the amount that all the real players gain. This game is converted into a constant-sum game by introducing "Nature" as a "dummy" player. In all constant-sum games the strategic problem faced by a coalition S and its complement $\bar{S}$ is one of pure opposition. A gain by S is reflected by an offsetting loss by $\bar{S}$.

The device of inventing the extra player was introduced to try to avoid the unpleasant modeling problems of describing the threat conditions that may exist when coalition S confronts $\bar{S}$ in a nonconstant-sum game. Table 4.1a shows a simple 2×2 matrix game. If we calculate the characteristic function by assuming that the opposition to S plays in a way to minimize S's gain we obtain the symmetric function shown in Table 4.1b.

Table 4.1a

		Player 2 1	Player 2 2
Player 1	1	10,−1	−1 000,0
	2	0,10	0,10

Table 4.1b

$v(\bar{1}) = 0, \quad v(\bar{2}) = 0$

$v(\overline{12}) = 10$

This calculation completely masks the underlying nonsymmetry of the game, however. In order for player 1 to hold player 2 to zero he must be willing to suffer a loss of −1 000, whereas if player 2 uses his second strategy it is in player 1's self-interest to accept zero while player 2 obtains 10.

Shapley and Shubik (see Shubik, 1982, Chapter 6) have suggested the concept of a ***c*−game**, which means a game or strategic situation that is adequately represented by a characteristic function. A simple but important example of a ***c*−game** is the cooperative version of an exchange economy. Any set of traders S can trade among its members, but the coalition $\bar{S}$ has no threat beyond not trading. In the language of the economist the economic (or other) activity of S or $\bar{S}$ generates no externalities to the other.

When, as in the example in Table 4.1a, the cost of carrying out a threat is important we would like this fact to be reflected in the cooperative representation of the game that is used. In the development of his value solution, Harsanyi (1959) suggested a way to evaluate threats, which for situations involving monetary side payments can be described as

$$h(S) + h(\bar{S}) = v(N) \quad \text{when} \quad S \cup \bar{S} = N$$

$$h(S) - h(\bar{S}) = \text{maxmin [payoff to } S - \text{payoff to } \bar{S}] \quad .$$

The first of these two linear equations states that the coalitions S and $\bar{S}$ when cooperating will obtain everything. When threatening each other they will try to maximize the difference between their scores. This defines a damage exchange rate where both the damage to the other and the cost of inflicting the damage are taken into account.

The distinction between the characteristic function $v(S)$ and the Harsanyi function $h(S)$ is in the modeling argument concerning how threats are treated. One could use other arguments to decide upon the joint product obtainable by a group S of firms, individuals, departments, or machines. A different way of looking at the characteristic function is to regard it merely as a production function with values defined only on sets of resources.

In essence the joint cost, revenue, and externality problems involve a finite set of profit centers whose activities influence each other. When interests are either independent (no externalities) or completely opposed it is clear how to calculate $v(S)$. ($v(S)$ and $h(S)$ will coincide.) When this is not so we may need to consider the special properties of the problem at hand.

Although for large n the number 2^n becomes considerable (for $n = 20$, $2^n = 1{,}048{,}576$); in actual application many combinations can be ruled out quickly. Even though the characteristic function provides many degrees of freedom in modeling the economic environment it may still not provide enough. There are at least two larger representations which merit consideration. They are the characteristic function supplemented by a list of weights indicating the relative importance of the players; and the game represented in partition form.

Instead of describing the game merely by the set N of players and the characteristic function $v(\bullet)$, we add a vector of weights to the description, one for each player. Shapley (1981) has suggested an application of the weighted game to the payment of expenses; this is discussed further in Section 3. Shubik and Weber (1981) have considered the adding of weights to players in the allocation of costs to a defense system.

The characteristic function and its variants are based upon considering a coalition S and a counter-coalition $\bar{S}$. It is possible however that the amount a coalition S can obtain depends upon the configuration of subgroups formed by the remaining players. Thrall and Lucas (1963) formulated the concept of games in partition function form and investigated an extension of the von Neumann and Morgenstern solution to such games. Let $N = \{1,2, \ldots, n\}$ be the set of players and $P = \{P_1, P_2, \ldots, P_r\}$ be a partition of N into r coalitions. Let π denote the set of all partitions and R the real numbers. Then for each partition P there is an outcome function $F_P \colon P \to R$ which assigns an outcome $F_P(P_i)$ to coalition P_i given the partition P. The function F that assigns to each partition its outcome function is called the ***partition function*** of the game. An n-person game in partition function form is characterized by (N,F).

Given any partition function F we can obtain ***upper bounding*** and ***lower bounding*** characteristic functions $\bar{v}$ and v^1 as follows: for each subset S of N

$$\bar{v}(S) = \max_{\{P \mid S \in P\}} F_P(S) \quad ,$$

$$v^1(S) = \min_{\{P : S \in P\}} F_P(S) \quad .$$

In the first instance a coalition S is given the maximum worth it can obtain as a coset in any partition to which it belongs. In the second instance it is awarded the minimum. These functions will not necessarily be superadditive. The partition function picks up the possibility that the yield to a group depends not merely upon the set of the remaining players but on the specifics of its organization.

It is easy to see that without externalities (as in an exchange economy) the partition function and characteristic function forms coincide.

The cooperative representation of an n-person game or multidivision corporation or other institution is in general far more parsimonious than either the extensive or strategic form, which we discuss further in Section 4. The cooperative representation suppresses institutional and process detail. The danger in utilizing the characteristic function is that the threat structure present in some situations may not be adequately represented. For many cost allocation and revenue assignment problems the cooperative representation appears to be the simplest and easiest to work with if it can be established that it provides an adequate model for the problem at hand.

3 Solutions to Cooperative Games

The principal candidates for solution concepts to games in the cooperative form are the core, value, nucleolus, kernel, bargaining set, stable set, and several others. It is suggested here that the two cooperative solution concepts most appropriate to the study of allocation and incentive problems are the value and the nucleolus. For games in extensive and strategic form the most appropriate solution is the noncooperative equilibrium (see Section 4).

The kernel, bargaining set and stable set (see Shubik, 1982, for definitions and discussion) are more appropriate for bargaining and sociological analysis than for allocation problems.

3.1 *The Core*

When there is the appropriate economic structure to the problem at hand the core can be defined for either a characteristic function or partition function description. A considerable body of literature exists on the relationship between the core of an economic system modeled as a game and the efficient price system (see Shubik, 1959; Debreu and Scarf, 1963; and for a survey Shubik, 1984, Part III). Yet, when the somewhat special assumptions concerning technological independence of production and lack of externalities do not hold, there is no guarantee that the core will exist. Although Hamlen *et al.* (1977) suggest the use of core theory in evaluating joint cost allocation, the only typical situations for which the core will probably exist is mergers, joint ventures, and cartel arrangements. The existence of the core when joint production, joint revenues, externalities and voting procedures are present is not guaranteed. Thus, for example, the relationship betweens the core and Lindahl prices (see Musgrave and Peacock, 1958) is tenuous at best, as the core depends on the characteristic function, which in turn depends upon technological and institutional restrictions on the behavior of groups, whereas the Lindahl prices do not depend upon coalition structure.

A key feature in determining the possibility for decentralization by prices is whether the economic entity being studied can be represented in cooperative form by a totally balanced game (see Shapley and Shubik, 1968, or Shubik, 1982, Chapter 6). Thus one way of considering limitations on individual strategic independence is to see whether it is consistent with finding a characteristic function that defines a totally balanced game.

An interesting possibility that does not appear to have been investigated in any detail is to use a partition function representation of the game, to evaluate threats. As described above in Section 2, one can define upper and lower bounding characteristic functions by giving each coalition S its largest (respectively smallest) payoff in any partition of which is a coset. Then $\bar{v}(S) \geq v^1(S)$ for all S, where $\bar{v}(S)$ is the upper and $v^1(S)$ the lower characteristic function. The lower characteristic function will have the larger core. There is no guarantee that either will be superadditive. However, if the upper characteristic function game is totally balanced then the lower characteristic function game will also be balanced. Thus there will be at least one and possibly two sets of shadow prices which permit decentralized decision making, depending upon whether the corporate rules limit the payoffs of S to $\bar{v}(S)$ or $v^1(S)$. It is possible that the upper characteristic function game has no core, yet the lower characteristic function game is totally balanced. In this case a unique set of shadow prices will permit decentralized decision making.

The point of these observations is to stress that the design of decentralized systems has two components – the laws of physics and the laws of organizations. If the technological facts of life are sufficiently bad it may not be possible to maintain a given degree of decentralization.

For the three-person game, if the lower game is balanced so is the upper game. Suppose that

$$F_{\{1,2,3\}}(1) = a\,,\ F_{\{1,2,3\}}(2) = b\,,\ F_{\{1,2,3\}}(3) = c\,,$$
$$F_{\{1,23\}}(1) = 0\,,\ F_{\{1,23\}}(23) = f \geqq 0$$
$$F_{\{2,13\}}(2) = 0\,,\ F_{\{2,13\}}(13) = e \geqq 0$$
$$F_{\{3,12\}}(3) = 0\,,\ F_{\{3,12\}}(12) = d \geqq 0$$
$$F_{123}(123) = 1 \quad .$$

Then for the lower game

$$v^{1\prime}(\phi) = 0$$
$$v^1(1) = v^1(2) = v^1(3) = 0$$
$$v^1(12) = d\,,\ v^1(13) = e\,,\ v^1(23) = f\,,\ d, e, f \geq 0$$
$$v^1(123) = 1 \quad .$$

Total balance requires that $(d + e + f)/2 \leq 1$. For the upper game

$$\bar{v}(\phi) \doteq 0$$
$$\bar{v}(1) = a\,,\ \bar{v}(2) = b\,,\ \bar{v}(3) = c$$
$$\bar{v}(12) = d\,,\ \bar{v}(13) = e\,,\ \bar{v}(23) = f$$
$$\bar{v}(123) = 1 \quad .$$

Renormalizing so that all one-person coalitions have a value of zero then checking the total balance we obtain the necessary and sufficient condition

$$0.5[d + e + f - 2(a + b + c)] < 1 - a - b - c \quad ,$$

which is equivalent to $(d + e + f)/2 \leqq 1$. The possibility for differences in total balance for the upper and lower games starts with $n = 4$.

3.2 *The Value*

In my estimation the value and possibly the nucleolus are the two most important solution concepts for the allocation of joint costs and revenues. We already know how to apply the value to the characteristic function, the player-weighted characteristic function, and the partition function. It is my opinion that the outer reaches of generality in the application of value theory to a coalitional structure would be to a partition function form with weighted importance to players, but there still remains much to be done before this extra complication is explored.

The Shapley value was originally based on the von Neumann–Morgenstern characteristic function. Shapley (1951, 1953) gave two versions of an axiom system for the value. Shubik (1962) used the first to suggest accounting desiderata. Since then, supported by the work of Aumann and Shapley (1974), various formulations and modifications of the axioms for the value with a continuum of players have been offered. Myerson (1977) has extended the axiom system for the value to games in partition function form and Shapley (1981) has extended his axiom system to include players with different weights. Every one of these systems merits consideration in terms of accounting desiderata.

The value is the natural extension of the type of thinking in economics that made the use of marginal analysis so fruitful. In essence the value is the combinatoric version of marginal analysis. Instead of evaluating at a single point the marginal contribution is evaluated over all combinations, where all permutations of the players are deemed to be equally likely.

It is possible to calculate the value by assuming a continuum of players as was illustrated by Billera *et al.* (1978) for many small players; still, when numbers are few but bigger than five or six the calculation of the value is laborious unless use can be made of special properties. Littlechild and Owen (1973) provide an example of a simple calculation (see Shapley and Shubik, 1969, and Mann and Shapley, 1964, for some relatively large calculations).

3.3 *The Nucleolus*

We define the *excess* of a coalition S at a particular payoff vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ by

$$e(S,\boldsymbol{x}) = v(S) - \sum_{i \in S} x_i \quad .$$

It provides a measure for how much more (or less) a coalition obtains in an imputation than it could obtain by acting alone. Any imputation in the core has an excess less than or equal to zero.

The *nucleolus* (Schmeidler, 1969) is a single-point solution that always exists and minimizes the dissatisfaction of the most dissatisfied coalition. More formally we define the *ε-core* of a game v to be the set of all imputations that would be in the core if each coalition were given a subsidy of ε. Now vary ε until we find the smallest nonempty ε-core. This in known as the *near core*. This is the set of imputations at which the maximum excess over all coalitions has been minimized. If we were to continue to vary ε for all coalitions we would wipe out the near core. Instead we consider only those coalitions whose value is not constant on the near core and, by varying ε, minimize their maximum excess. We repeat this procedure until only a single point remains. Littlechild (1974) and Littlechild and Thompson (1977) have used the nucleolus as a costing allocation device.

Sobolev (1975) has produced a set of axioms from which the nucleolus can be derived (which have not yet been published in English). An outline of these ideas and their potential application to cost allocation is given in Chapter 1 (Section 5.3). The attractive feature of the nucleolus is that even without the axioms the idea of an ε-adjustment has a direct interpretation either in terms of subsidies and taxes, or the minimization of claims of inequality among coalitions.

4 The Extensive and Strategic Forms: The Noncooperative Equilibrium

The most realistic model of an organism is the organism itself. Any representation is an abstraction which removes or distorts information. If one is trying to answer a specific question concerning the behavior of the organism, a model judiciously selected may portray the features of the organism which are relevant and simplify the analysis by obliterating detail irrelevant to the question at hand. Von Neumann and Morgenstern (1944) provide a description of games in extensive form, with each player having a finite set of strategies. The details of the game tree description are well known (for an exposition by Shapley and Shubik, see Shubik, 1982). Our concern is with the modeling implications of what is included and excluded in the extensive form.

The extensive form provides a total contingent planning or historical process view of a game. The use of the game tree is explicitly historical. The same position on a chess board arrived at by different sequencing of moves will lead to a different node on a game tree for every sequence. The game tree is *process-oriented*. Any path from the initial node or root of a game tree to a final or terminal node provides a move-by-move description of a play with information conditions indicated on the game tree.

A description of matching pennies as a game in extensive form is easy; the full game tree for tic-tac-toe is large but could be displayed. A full game tree for chess, though logically feasible, is technologically infeasible and operationally of little value. The difference in complexity between the detailed description of process in the play of a chess game and behavior in a corporation in a market is enormous – both quantitatively and qualitatively. In particular it is easy to specify more or less unambiguously the rules of the game for chess and identify what is meant by a move. Two-person, zero-sum games are games of pure opposition. There is nothing to be gained by the players talking to or bargaining with each other. Although in fact chess players may try ploys and various forms of psychological warfare they are forbidden by the rules to do so. As a reasonable first-order approximation, cheating can be ignored in the description of most chess games.

When we try to portray the relatively simple three- or four-person game of Poker or Monopoly, even though the formal rules of the game are given, much of the dynamics of play hinges upon language and informal communication not specified in the rules of the game, and yet not clearly or expressly ruled out by the formal rules. When are words merely words extraneous to deed and when are words and informal communication a critical part of the process? Bargaining, negotiation, contracting, the giving of promises, the offering of one's word of honor, are all examples of the importance of verbal communication as vital parts of the game. The perceptive book of Raiffa (1982) on negotiation serves as an important example of the difficulties in trying to match the formal decision structure of a game in extensive form with the squishy and poorly articulated realities of human communication systems in quasi-cooperative and quasi-competitive situations.

Any attempt to model much of the activities of the corporation as a game in extensive form must confront two new and critical sets of difficulties in contrast with modeling the game of chess. Language and informal communication count and the rules of the game cannot be easily formulated. Implicit contract, implicit collusion, reputation, trust and social or institutional customs all play a role, yet they are extremely difficult to formalize.

The existence of a vast body of law complete with intricate documents such as the commercial code and the law of contract testify both to the attempts and to the incompleteness of the attempts to formalize rules of commercial behavior.

Ideally we would like to have a parsimonious description of the corporation and to discover broad general rules of institutional control – be they accounting measures, bonus systems, or reporting routines. Yet experience teaches us that in the twilight of institutional complexity that characterizes large private corporations, public bureaucracies, or state-owned industries, there are tax consultants, lawyers, fixers, millionaires, and commissars beating the system by utilizing details of the mechanisms overlooked by their designers (or possibly left vague on purpose to provide loopholes for allies in a difficult game with a hidden agenda).

In sum, the extensive form places a laudable stress on process, but in general the modeling difficulties encountered in trying to provide a full process description are overwhelming. The sheer complexity of amount of detail combined with difficulties in characterizing rules limits the value of the extensive form as a satisfactory basis from which to start an analysis, except for highly stripped down and simplified representations.

5 Problems and Prospects

5.1 *What is an Application*?

An unkind caricature of much of operations research and management science is that they consist of a set of techniques looking for a problem. The manager undoubtedly will be somewhat institutionally oriented. He has a specific organization to run. The economist, management scientist, game theorist and programmer often think in terms of their specialized bed of Procrustes. Thus a manager and his accountants may view large calculations of the value or nucleolus more as an exercise in the employment of surplus PhDs than a serious new way to allocate costs and revenues. I do not subscribe to either a belief in this caricature, or to the view that game theory models provide full answers to all of the problems in cost accounting. My view is mildly optimistic and the remainder of this paper is devoted to the question of what constitutes a worthwhile application.

I suggest that there are at least five levels of meaning as to what constitutes an application of a methodology to an applied problem. In particular concerning game-theoretic analysis as applied to cost and revenue allocation its uses are as follows:

(1) To stop errors and challenge the basis for practices.
(2) To suggest good questions, formalized operationally.
(3) To suggest simple, better alternatives.
(4) To provide new, formal accounting control systems.
(5) To provide the calculations for specific answers to specific questions.

The first three applications are more at the level of high-level advice and criticism rather than an explicit formal program for accounting. Experience as a consultant, professional knowledge, and enough technological and institutional background can lead to identifying bad practices and to raising important questions concerning current procedures. On occasion an immediate *ad hoc* improvement may be spotted that requires little institutional adjustment. When the climate is right, the use of modeling and computers has become more or less accepted. When enough managers and accountants are receptive, there is a possibility for a conceptual and institutional reorganization of accounting practices.

5.2 *What are the Questions*?

There are many different questions concerning the assignment of joint costs and revenues and the possibility of answering them depends heavily upon *ad hoc* institutional and technological facts that determine how well the game can be formulated in coalitional or other forms.

Among the more important questions are how to design a system where it does not pay individuals to lie to the central office, the tax collector, and to whomever else they submit their reports. In general the cooperative form is not adequate to study problems of auditing, enforcement, and agency relationships under incomplete information. There is a bargaining literature based heavily upon strategic or simple extensive form models (see for example Groves and Ledyard, 1977; Shubik, 1970). Items such as the cost of spot checks and random audits call for strategic analysis. Our hope is that there are worthwhile problems where at least to a good first approximation truth revelation is incentive compatible with the accounting scheme. Thus we will not need to worry about the strategic structure in detail as the appropriate incentive design has removed the need for information distortion as part of planned strategy.

In a large corporation the choice is not between centralization or decentralization, but the degree of decentralization needed for viability and the level of decentralization that results in optimal performance. The virtues of the price system as an efficient decentralizing device are well known. But when externalities are present the price system in general is not efficient.

A corporate central office has considerable power in deciding upon the nature of the structure of the firm. Among the factors determining the nature of decentralization are geographical location, nature of products, differentiation of functions, and frequency with which some functions are needed, jointness or separability of production processes, interlinkage of marketing of products, and sharing of common facilities. Given that the central management has decided upon a structure for the firm it decides upon the freedom of decision making for its executives, the management information system, and the incentive system under which they will operate.

In a previous paper (Shubik, 1962, p 331) I suggested a partial list of relevant decisions that might be aided by an appropriate allocation procedure. They were:

(1) Decision on major investment.
(2) Liquidation of a department.
(3) Abolition of a product line.
(4) Introduction of a new product.

(5) Other innovations (such as a change in distribution).
(6) The merger of several departments.
(7) The splitting of a department into several entities.
(8) Pricing, purchase of raw materials, and sales of final products.

All of these can be described as internal corporate decisions (or inner directed decisions) in as much as the decisions are all internal to one bureaucratic structure. There are other classes of decisions where the emphasis among incentives, power, and fair division is somewhat different, but which are also amenable to game-theoretic methods. Four classes of problems are suggested. They vary considerably in terms of differences in modeling required to arrive at an adequate description of the game.

5.2.1 *Internal corporate incentive systems*

Internal to the corporation there is an intermix of geography, technology, economics, politics, accounting, legal and cultural conventions which limit the divisional structure of the firm.

Items which may appear to be trivial to the academic economist may be of paramount importance to those concerned with corporate control. For example, should one keep at least three sets of books and should one openly admit to keeping the three sets of books! One may want one set for the tax collector where the operational consideration is to minimize the tax bill. Another set may be required for stockholders and creditors and a third set for internal control and incentives. In some countries both legal and societal pressures may leave the corporation open to political and populist attack when large differences are found in different sets of books. A decentralized multinational corporation such as an oil company is faced with designing cost and revenue allocations that minimize taxes over dozens of countries, yet which do not destroy the morale of regional directors. For example, some years ago it would have been difficult for even the most incompetent manager of a Venezuelan branch of an oil company to fail to report enormous profits whereas a competent manager in Great Britain would have been derelict in his duties had he not reported losses.

If incentive pay or bonuses depend upon local performance and must be justified to the stockholders then the corporation must explain why one set of books is used for one purpose and another for the other.

If the firm is modeled as an n-person game in coalitional form, it is the general management which has the opportunity to decide upon the number of players and the constraints on their strategy sets. For example what limits are placed upon the amount of money that a general manager can invest without having to see his divisional or group vice-president? Who is permitted to generate the suggestion to merge two departments, or to split a department into two?

The frequency of the need for special services, the costs of record-keeping, accounting, calculating, auditing and communicating all enter into the decisions to choose among markets, divisions and hierarchies. Williamson's (1975) perceptive book spells out many of the detailed factors. These and other factors must already have been adequately reflected in the characteristic function description of the corporation. Furthermore, any merger of departments or institutional change unlike the costless coalition formation in much theory may have important administrative costs attached to it. This needs to be accounted for. Shapley and Shubik (1966) have suggested an approximate way to charge all coalitions an organizational cost.

In short, prior to even discussing what solution concept to imply in any serious application much of the work involves providing a sufficiently relevant description of the firm in cooperative form. Possibly one of the major contributions of game theory is to provide a conceptual framework and a strategic audit (see Shubik, 1983) which provides a guideline to problem formulation and data gathering rather than a rigid accounting system for the design of incentive-compatible allocation schemes.

5.2.2 *External corporate incentive systems*

All of the problems of measurement noted for internal corporate control systems hold for external corporate problems, which can be considered as a cooperative game. These include:

(1) Mergers and acquisitions.
(2) Cartel arrangements (where legal).
(3) The splitting of costs and revenues in joint ventures.

The differences here are that the players and their strategy sets are more naturally identified with independent decision-making groups. Furthermore, the number of players is usually below ten and often two or three.

5.2.3 *Public goods and externalities*

The literature on public utilities, public goods in general and externalities is enormous and has been in existence for some time (see for example the collection of Musgrave and Peacock, 1958). The treatment of these topics here is not intended except to note the key modeling differences among public goods problems, externalities and corporate allocation problems. A major aspect of public goods provision is that at some part of the process direct or indirect political bargaining and voting is involved. A major aspect of the structure of most situations involving externalities is that they involve a high component of legal as well as political process. In contrast to both of these most corporate problems have a far higher economic content combined with much better defined recording and reporting procedures. Furthermore, for-profit operations tend to have fewer measurement problems than public services. The work of Klevorick and Kramer (1973) on the *Genossenschaften* and the work of Young *et al.* (1982) provide examples of public goods and externality problems at the level of municipal finance.

5.2.4 *Technological, taxation and legal fair division problems*

There is class of problems where the technological component may be high and the cooperative form natural and relatively easy to define. The purpose at hand is well defined and the players and their strategy sets are reasonably easy to define parsimoniously. This class includes the telephone system, time-sharing computers, aircraft loading, the sharing of joint services, peak-load pricing, and expense account allocation.

The problems have in common the features that moral hazard and reporting distortion are clearly defined, minimal or nonexistent. In many of the network or joint service problems the strategy of the users amounts to using or not using the facility. The individuals need not be looked at as players or can be regarded as extremely small without individual power. The characteristic function can be looked at more as a production function than as a game of strategy.

5.3 *Prospects*

It is my belief that, at least in economics, a true conceptual breakthrough shows its ultimate importance in application when it serves as the basis for a new accounting scheme. The three accounting schemes of signal importance to the development of the modern economy were

(1) Double-entry book-keeping, which vastly increased the possibility for individual trade and enterprise.
(2) National income accounting, which provided a conceptual basis for much for the economic accounting control structure of the modern nation state.
(3) The input-output system accounting schemes, which provided major links in accounting for production, derived demand, and final demand.

In each instance the accounting system generated data that were not previously available. In the process of devising routines and having many professionals consider the new accounting schemes, new problems, conceptual difficulties and gaps in information gathering were discovered. The interplay between economic theory and practice has to be sufficiently two-way if benefits are to be derived. The next accounting revolution will occur in the combinatorics of joint costs and revenues. It has already begun.

Acknowledgments

The research described in this paper was partially supported by a grant from ONR and was previously an IIASA Collaborative Paper.

References

Aumann, R.J. and Shapley, L.S. (1974) *Values of Non-atomic Games* (Princeton, NJ: Princeton University Press).

Billera, L.J., Heath, D.C., and Raanan, J. (1978) Internal telephone billing rates: a novel application of monotonic game theory. *Operations Research* 26: 956–965.

Debreu, G. and Scarf, H.E. (1963) A limit theorem on the core of an economy. *International Economic Review* 4: 235–46

Groves, T. and Ledyard, J. (1977) Optimal allocation of public goods: a solution to the free rider problem. *Econometrica* 45: 785–809.

Hamlen, S.S., Hamlen, W.A., and Tschirhart, J.T. (1977) The use of core theory in evaluating joint cost allocation schemes. *Accounting Review* 52: 616–27.

Harsanyi, J.C. (1959) A bargaining model for the cooperative n-person game, in A.W. Tucker and R.D. Luce (Eds) ***Contributions to the Theory of Games 4***, Vol. IV (Princeton, NJ: Princeton University Press) pp 324–56.

Klevorick, A.K. and Kramer, G.H. (1973) Social choice on pollution management: The Genossenschaften. ***Journal of Public Economics*** 2: 101–46.

Littlechild, S.C. (1974) A simplistic expression for the nucleolus in a special case. ***International Journal of Game Theory*** 3: 21–29.

Littlechild, S.C. and Owen, G. (1973) A simple expression for the Shapley value in a special case. ***Management Science*** 20: 370–2.

Littlechild, S.C. and Thompson, G.F. (1977) Aircraft landing fees: a game theory approach. ***Bell Journal of Economics*** 8: 186–204.

Mann, I. and L.S. Shapley (1964). The *a priori* voting strength of the electoral college, in M. Shubik (Ed) ***Game Theory and Related Approaches to Social Behavior*** (New York: Wiley) pp 151–164.

Musgrave, R.A. and Peacock, P.B. (Eds) (1958) ***Classics in the Theory of Public Finance*** (New York: McGraw-Hill).

Myerson, R. (1977) Values of games in partition function form. ***International Journal of Game Theory*** 6: 23–31.

Raiffa, H. (1982) ***The Art and Science of Negotiation*** (Cambridge, MA: Harvard University Press).

Schmeidler, D. (1969) The nucleolus of a characteristic function game. ***SIAM Journal on Applied Mathematics*** 17: 1163–1170.

Shapley, L.S. (1951) Notes on the n-person game, II: The value of an n-person game, Rand RM-641 (Santa Monica, CA: Rand Corporation).

Shapley, L.S. (1953) A value for n-person games, in H. Kuhn and A.W. Tucker (Eds) ***Contributions to the Theory of Games***, Vol. II (Princeton, NJ: Princeton University Press) pp 307–317.

Shapley, L.S. (1981) Comments on P.D. Banker's 'Equity considerations in traditional full cost allocation practices: an axiomatic perspective', in S. Moriarity (Ed) ***Joint Cost Allocations*** (Tulsa, OK: University of Oklahoma Press).

Shapley, L.S. and Shubik, M. (1966) Quasi-cores in a monetary economy with non-convex preferences. ***Econometrica*** 39: 805–827.

Shapley, L.S. and Shubik, M. (1968) On market games. ***Journal of Economic Theory*** 1: 9–25.

Shapley, L.S. and Shubik, M. (1969) Pure competition, coalitional power, and fair division. ***International Economic Review*** 10: 337–362.

Shubik, M. (1959) Edgeworth market games, in R.D. Luce and A.W. Tucker (Eds) ***Contributions to the Theory of Games***, Vol. IV. (Princeton, NJ: Princeton University Press).

Shubik, M. (1962) Incentives, decentralized control, the assignment of joint costs and internal pricing. ***Management Science*** 8: 325–343.

Shubik, M. (1970) Budgets in a decentralized organization with incomplete information, (Santa Monica, CA: Rand Corporation).

Shubik, M. (1982) ***Game Theory in the Social Sciences***, Vol. I (Cambridge, MA: MIT Press).

Shubik, M. (1983) The strategic audit: a game theoretic approach to corporate competitive strategy. ***Managerial and Decision Economics*** 4: 3.

Shubik, M. (1984) ***Game Theory in the Social Sciences***, Vol. II (Cambridge, MA: MIT Press).

Shubik, M. and Weber, R. (1981) Systems defense games: Colonel Blotto command and control. ***Naval Research Logistics Quarterly*** 28.

Sobolev, A.I. (1975) Characterization of the principle of optimality for cooperative games through functional equations, in N.N. Voroby'ev (Ed), ***Mathematical Methods in the Social Sciences***, Vipusk 6, Vilnius, USSR, 92–151.

Stigler, G. (1966) ***The Theory of Price***, 3rd edn (New York: MacMillan).

Thrall, R.M. and Lucas, W.F. (1963) N-person games in partition function form. ***Naval Research Logistics Quarterly*** 10: 281–98.

von Neumann, S. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press).

Whitman, M.J. and Shubik, M. (1979) *The Aggressive Conservative Inventor* (New York: Random House).

Williamson, O.E. (1975) *Markets and Hierarchies – Analysis and Antitrust Implications* (New York: Free Press).

Young, H.P., Okada, N. and Hashimoto, T. (1982) Cost allocation in water resources development. *Water Resources Research* 18(3): 463–75.

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 5

THE IMPOSSIBILITY OF INCENTIVE-COMPATIBLE AND EFFICIENT FULL COST ALLOCATION SCHEMES

Theodore Groves

1 Introduction

Consider the following situation: A firm consists of n divisions or "profit centers", $i = 1, \ldots, n$, that are, for accounting or control purposes, treated as separate units but are also connected through some firm-wide decisions. For concreteness, consider a single firm-wide decision such as the provision of some service (e.g., computing facilities) that is made available to all divisions. We suppose that the total quantity of the good is available to each division and that its use by one division does not diminish the availability or usefulness of the good to any other division (thus, the good is a "public good" to the divisions).

Two basic questions confront the firm with respect to this decision. First, the firm must decide the quantity of the good to provide and second it must decide how to finance its provision. Concerning the decision of the optimal quantity to purchase, it is reasonable to suppose that the value to the firm of varying quantities of the resource is given by the sum of the values of the resource to each division. Thus, in order to determine an optimal quantity of this centrally provided resource it is necessary to know its value to the various divisions as well as its cost.

Now, as long as the resource is valuable to a division, the division's manager will be interested in having more of it made available. Yet there is, of course, a quantity beyond which, although still valuable to the individual divisions, the resource is not sufficiently valuable to cover its marginal cost. If the center or agent responsible for choosing the quantity of this service/resource for the firm does not know (with sufficient precision) the value of the resource to the divisions, then it must obtain this information from the divisions themselves. Thus, it is essential for optimal decisions that the divisions report accurately or "honestly" to the center and a nontrivial question is what incentives they have to do so. Since we are interested only in procedures that make optimal decisions when provided with accurate information we confine attention to such decision rules and ask how divisional managers might be induced to send correct information.

We suppose that the managers of the divisions are evaluated on the basis of their division's performance as measured by divisional net revenues adjusted for any costs assessed them for the centrally provided resource. A ***full cost allocation*** scheme is one that divides the total cost fully among the divisions. For a centrally provided public good, standard accounting practice suggests that each division be assessed in proportion to its utilization of the resource or, perhaps, if divisional use is not easily measured, in proportion to net divisional revenues. Regardless of the particular scheme adopted, a full cost allocation scheme has two features of importance for our concern. First, the cost assigned to any division will depend jointly on all the information provided the center by all the divisions, if the firm attempts to make an optimal decision. For example, even under an equal sharing of costs, the total quantity of the resource purchased by the firm will depend on all the divisions' information to the center if the center attempts to purchase the optimal amount for the firm as a whole. The second feature of any full cost allocation scheme is that the sum of adjustments to all divisions equals the total cost of the resource.

Suppose that a full cost allocation scheme is specified for our firm. Every division manager is then evaluated on the basis of his divisional revenues less the adjustment (or cost). His evaluation, which we assume he wants to maximize, depends on the information he provides the center in two ways. First, his information affects the amount of the resource purchased by the center and hence his divisional revenues. Second, his information affects his share of the cost of the resource. Thus, in choosing what information to send, we assume he will take into account exactly how his information will affect his evaluation. We call the full cost allocation scheme ***incentive compatible*** if (and only if) it is always in every division manager's interest to send correct information, regardless of the information sent by other divisional managers. That is, if the full cost allocation scheme is incentive compatible the "correct" information must be the information that will always maximize the divisional manager's evaluation.

The main result I wish to communicate here is: there is no incentive-compatible full cost allocation rule if the decision rule specified must also pick optimal decisions when supplied with correct information. Thus, either the full cost feature, incentive compatibility, or optimality of decisions must be sacrificed. If making optimal decisions is the main task of the procedure and managers are trusted to follow the incentives defined by the firm, then it appears that full cost allocation schemes are inappropriate for these purposes, regardless how useful they might be for purposes other than making optimal decisions.

This result follows from several other results that are interesting in their own right. The first of these is a result, due to Groves (1973) that exhibited a family or class of incentive-compatible cost allocation rules, but which were not full cost allocation rules. The other key result, established by Green and Laffont (1977), showed that any incentive-compatible procedure for making optimal decisions was equivalent to a member of the family of incentive-compatible cost allocation rules defined by Groves.

2 The Formal Model

Formally, these results can be modeled as follows. Let $i = 0,1, \ldots, n$ denote n divisions and a center ($i = 0$). Suppose profits or payoffs to the firm as a whole are separable in all decisions d_i, $i = 1, \ldots, n$, excepting the center's decisions $d_0 = z$. For example, d_i might be division i's choice of technique or the amount of a locally available resource, etc., while d_0, the center's decision, might be the amount of a public or common input made available to all divisions or an allocation to each division of a centrally available resource. Let profits be denoted by

$$\tilde{\pi}(d_1, \ldots, d_n ; z) = \sum_{i=1}^{n} \tilde{\pi}_i (d_i, z) - C(z) \quad ,$$

where $C(z)$ might be interpreted as the full cost to the firm of the center's decision z, and $\tilde{\pi}_i(\cdot)$ are the divisional net revenue functions.

Assuming the local divisional decisions d_i are to be taken *after* the center's decision is made and announced, we can focus exclusively on the center's decision by optimizing independently with respect to the local decisions. Thus, let

$$\pi_i(z) \equiv \max_{d_i} \tilde{\pi}_i (d_i, z) \quad ,$$

and define the firm's payoff to be

$$\pi(z) = \sum_{i=1}^{n} \pi_i(z) - C(z) \quad ,$$

where $\pi_i(z)$ denotes the (maximized) net payoff of division i. The firm's decision problem is then to find the center's decision z to maximize $\pi(\cdot)$.

Now, since information is assumed to be initially dispersed, say, each division manager i knows only his own $\pi_i(\cdot)$ and the center knows the cost function $C(\cdot)$, in order to solve the firm's problem the center must discover a sufficient amount about $\pi_i(\cdot)$. Suppose, therefore, that the center asks each divisional manager i to report his function $\pi_i(\cdot)$. The divisional manager's incentives to report "truthfully" or "honestly" depend on how their information will be used to evaluate their performance. I assume divisional managers are evaluated on the basis of their own divisional net revenues $\tilde{\pi}_i(d_i, z)$ *less* an adjustment $c_i, \tilde{\pi}_i(d_i, z) - c_i$. Since the adjustment c_i is independent of the local decisions d_i, an evaluation-maximizing divisional manager will choose d_i to maximize $\pi_i(d_i, z)$ and thus, I write the evaluation as $\pi_i(z) - c_i$.

Suppose that $C_1(z), \ldots, C_n(z)$ is any ***full cost allocation*** scheme, i.e.,

$$\sum_{i=1}^{n} C_i(z) = C(z) \quad \text{for all } z \quad ,$$

and let the adjustments c_i to the divisional evaluation measures be given by this scheme: $c_i = C_i(z)$. Then, division manager i will be evaluated on the basis of the measure $\pi_i(z) - C_i(z)$, where the function $C_i(\cdot)$, can depend only on the information acquired by the center.

Since I do not assume each divisional manager necessarily reports the "truth", i.e., sends his true net payoff function $\pi_i(\cdot)$ to the center, I let $m_i(\cdot)$ denote the ith manager's ***reported*** net payoff function. If, for some reason, the divisions ***do*** send their true net payoff functions, i.e., $m_i(\cdot) = \pi_i(\cdot)$, then the center's optimal decision z^* would be the solution to the program:

$$\max_z \sum_{i=1}^{n} m_i(z) - C(z) \quad .$$

Thus, I define the center's *decision rule* $z^*(m_1, \ldots, m_n)$ for choosing z given any messages $m_1(\cdot), \ldots, m_n(\cdot)$ from the divisions by:

$$z = z^*(m_1, \ldots, m_n) \equiv \operatorname{argmax} \sum_{i=1}^{n} m_i(z) - C(z) \quad .$$

Now, given the rule for choosing the decision z and the full cost allocation scheme $C_i(\cdot)$, $i = 1, \ldots, n$, each division manager i faces the problem of choosing the best report $m_i(\cdot)$ to send the center: choose $m_i^*(\cdot)$ to maximize with respect to $m_i(\cdot)$:

$$\pi_i[z(\tilde{m} / m_i)] - C_i[z(\tilde{m} / m_i); \tilde{m} / m_i] \quad , \tag{1}$$

where $\tilde{m} / m_i \equiv (\tilde{m}_1, \ldots, m_i, \ldots, \tilde{m}_n)$, and where C_i is allowed to depend directly on all the messages m_j, $j = 1, \ldots, n$, as well as on the decision $z(m)$.

We say the full cost allocation scheme $(C_1, \ldots, C_n)$ is *incentive compatible* if and only if π_i solves the division's optimization problem; i.e., if $\pi_i(\cdot)$ maximizes (1) for any $m_j(\cdot)$, $j \neq i$.

I now state the main theorem of this chapter:

Theorem 1 (Green and Laffont, 1977; Hurwicz, 1975, 1981; Walker, 1978): *There is no incentive-compatible full cost efficient allocation scheme. More formally, there is no decision rule–cost allocation scheme* $[z(\cdot),(C_1(\cdot), \ldots, C_n(\cdot))]$ *such that*

$$\sum_{i=1}^{n} C_i[z(m), m] = C[z(m)]: \text{ full cost allocation} \tag{2}$$

$$\pi_i(\cdot) = \operatorname*{argmax}_{m_i} \pi_i[z(m / m_i)] - C_i[z(m / m_i); m / m_i] \tag{3}$$

$$\text{for all } m: \quad \text{incentive compatibility}$$

$$z(\pi_1, \ldots, \pi_n) = \operatorname{argmax} \sum_{i=1}^{n} \pi_i(z) - C(z): \text{ efficiency} \tag{4}$$

Note that in the formal statement of the theorem it is not assumed that the messages or reports of the divisions m_j necessarily be functions of the center's decision z, i.e., be reported net divisional payoff functions. In principle, the messages (reports) might be anything. Also, note that the specification of the full cost allocation scheme allows in the special case of a centrally allocated (private) good for the cost to a division to depend on more than just its own allocation. That is, if

$$z = (z_1, \ldots, z_n), \; C(z) = \tilde{C}\left(\sum_{i=1}^{n} z_i\right) \quad , \quad \text{and } \pi_i(z) = \tilde{\pi}_i(z_i) \quad ,$$

then $C_i[z(m), m]$ may depend on a whole lot more than just $z_i(m)$.

The class of mechanisms (i.e., decision rule–cost allocation scheme pair) $[z(\cdot),(C_1(\cdot), \ldots, C_n(\cdot))]$ discovered by Groves can be defined by

$$z(m) \equiv \operatorname{argmax} \sum_{i=1}^{n} m_i(z) - C(z) \tag{5}$$

$$C_i[z(m),m] \equiv -\sum_{j \neq i} m_j[z(m)] + C[z(m)] + A_i(m \setminus m_i) \tag{6}$$

where $A_i(m \setminus m_i)$ is a function of all reports *except* that of division i; $m \setminus m_i \equiv (m_1, \ldots, m_{i-1}, m_{i+1}, \ldots, m_n)$. The relevance and interest in this class of mechanisms for the discussion here rests on the two propositions:

Theorem 2 (Groves, 1973): *Any member of the class defined by* (5) *and* (6) *above has the pair of properties that it is incentive compatible and results in efficient decisions being chosen.*

Theorem 3 (Green and Laffont, 1977): *Any mechanism (decision rule–cost allocation scheme pair) that is both incentive compatible and results in efficient decisions is equivalent to a member of the class discovered by Groves.*

In fact, the main result follows from showing that no member of this class is a full cost allocation scheme. However, one member of this family that has received some interest is defined by

$$A_i(m \setminus m_i) = \max_z \sum_{j \neq i} m_j(z) - C(z) \quad .$$

This member of the class has the property that, while it is does not allocate full cost exactly, that is,

$$\sum_{i=1}^{n} C_i[z(m),m] \neq C[z(m)] \quad \text{for all } m \quad ,$$

it always allocates at least full cost, that is,

$$\sum_{i=1}^{n} C_i[z(m),m] \geqq C[z(m)] \quad \text{for all } m \quad .$$

Other members of this family have different relations to full cost. For example, given a distribution on the space of possible net divisional payoff functions $\pi_i(\cdot)$, one member of the family is an expected full cost efficient allocating scheme, that is, one such that in expectation [with respect to the probability measure over $(\pi_1, \ldots, \pi_n)$], the cost of $z, C[z(\pi)]$, will be covered in expectation or on average (see Groves and Loeb, 1975).

References

Green, J. and Laffont, J.-J. (1977) Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica* 45: 427–438.

Groves, T. (1973) Incentives in teams. *Econometrica* 41: 617–663.

Groves, T. and Loeb, M. (1975) Incentives and public inputs. *Journal of Public Economics* 4: 211–226.

Hurwicz, L. (1975) *On The Existence of Allocation Systems Whose Manipulative Nash Equilibria are Pareto-optimal*, Paper given at Third World Congress of the Econometric Society in Toronto.

Hurwicz, L. (1981) On the incentive problem in the design of non-wasteful resource Allocation Systems, in N. Assorodobraj-Kula *et al.* (Eds) *Studies in Economic Theory and Practice* (Amsterdam: North-Holland)

Walker, M. (1978) A note on the characterization of mechanisms for the revelation of preferences. *Econometrica* 46: 147–152.

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 6

BARGAINING AND FAIR ALLOCATION

Terje Lensberg

1 Introduction

In this chapter, we will look at the allocation problem from the point of view of bargaining or fair division, a tradition initiated by John Nash's (1950) pioneering essay on cooperative bargaining.

A typical allocation problem that would fit within this framework is one in which a number of firms plan to undertake a joint venture if they can agree on how to share the profits from the project. The firms may settle for any agreement they can think of, as long as it is unanimous. If there is no agreement, there will be no project and no profits to share. Thus, we consider a situation where no proper subset of agents can accomplish anything on their own, i.e., there is no room for coalitions. The question then is: what will be the outcome?

In Nash's framework, one attempts to gain some insight into the problem by means of the following model. First abstract from all physical characteristics of the particular problem at hand by representing it as a pair $(S,\mathbf{d})$, where S is a set of feasible von Neumann–Morgenstern utility vectors, one vector for each physical alternative or probability mixture of alternatives, and where $\mathbf{d}$ is the utility vector that will be the outcome in the case of no agreement between the agents (the status quo). Now one may consider whole families of such abstract bargaining problems and look for a set of general principles (axioms) that would describe the behavior of the agents in any given bargaining situation. By requiring that the solution outcome to any bargaining problem (denoted by $F(S,\mathbf{d})$) should obey these axioms, one can hope to narrow down the set of possible candidates for a solution (the set of possible functions F) to a class with sufficient structure to yield some predictive power.

Clearly, nothing prevents us from looking at these principles from a normative instead of a descriptive viewpoint, e.g., as principles of fair division (Harsanyi, 1955), or the values of an arbiter (Raiffa, 1953). Here we will allow for both interpretations of the model and refer to the problems $(S,\mathbf{d})$ under consideration as collective choice problems, or simply *choice problems*.

To take an example in the normative spirit, consider the problem of dividing the costs and benefits of a public utility among its clients. Is there a division scheme which is in some sense fair? One way to attack the problem would be to identify a set of axioms that might reflect popular ideas of what constitutes a fair division, and use them to limit the opportunities for discretionary action by the management of the public utility. In some cases, as the one studied by Nash, the axioms will eliminate discretion altogether by singling out a unique division scheme.

As a third example, we may look at a society where public decisions are made in the following decentralized fashion. There are many publicly owned or regulated firms, each servicing a subset of the members of society; some municipal electricity companies, a number of public universities, a few airlines and some broadcasting companies. Clearly, the production and pricing decisions of each unit will affect the welfare levels of those individuals who consume and pay for its services. Suppose each unit is instructed to achieve a fair allocation of costs and benefits among its clients. An interesting question is whether such decentralized public decision making will lead to allocations that are fair for society as a whole. Put differently, if such decentralized decision making is going to be consistent with some overall notion of fairness, then what are the implications as to the nature of the decision rules that would have to be followed by the decentralized units, and what restrictions, if any, would such a requirement impose on the notion of fairness itself?

This question will be the main topic of the present paper. Our analysis is based on two ingredients. The first one is an axiom first used by Harsanyi (1959) in connection with the Nash bargaining solution, and which expresses the kind of consistency requirement mentioned in the previous paragraph.

Harsanyi's axiom differs from the ones that are usually studied within the tradition of bargaining and fair division. It is a condition on the relationship between choice problems involving ***different sets of agents***, while in the traditional model, the set of agents is ***fixed***. Problems of collective choice with a variable number of agents was first studied in a systematic way by Thomson (1983a), and it is his model which is the second main ingredient in our analysis.

The results of this analysis are several new characterizations of familiar solutions that shed some new light on the nature of those solutions. We discuss the relationship of Harsanyi's axiom to the problem of decentralized public decision making mentioned above. The axiom can be seen as a necessary condition for decentralized decision making to be consistent with any overall notion of fair division, and we show that it has very precise implications as regards the nature of the decision rules that would have to be followed by the decentralized units.

In the next section, we present the model and the axioms that we use. Section 3 contains the results, and Section 4 contains some concluding remarks on the potential applicability of our results to practical allocation problems.

2 The Model

The classical axiomatic model of bargaining involves a fixed set P of agents, each equipped with a von Neumann–Morgenstern utility function. P may be taken to be a nonempty, finite subset of the natural integers. Let $|P|$ be the number of elements in P, and let R^P be the $|P|$-dimensional Euclidean space, indexed by the members of P. R_+^P and R_{++}^P denote the nonnegative and the strictly positive orthant of R^P, respectively. Given $\boldsymbol{x},\boldsymbol{y}$ in R^P, we write $\boldsymbol{x} \geqq \boldsymbol{y}$ if $\boldsymbol{x}-\boldsymbol{y} \in R_+^P$, $\boldsymbol{x} > \boldsymbol{y}$ if $\boldsymbol{x}-\boldsymbol{y} \in R_{++}^P$, and $\boldsymbol{x} \geq \boldsymbol{y}$ if $\boldsymbol{x} \geqq \boldsymbol{y}$ and $\boldsymbol{x} \neq \boldsymbol{y}$.

A $|P|$-person bargaining problem is a pair $(S,\boldsymbol{d})$, where S is a subset of R^P, $\boldsymbol{d}$ is an element of S, and where S satisfies the following properties:

$A1$: S is compact and convex.
$A2$: There exists $\boldsymbol{y} \in S$ such that $\boldsymbol{y} > \boldsymbol{d}$.

S is the set of utility allocations that can be achieved by the members of P through unanimous agreement, and $\boldsymbol{d}$ is the outcome that will result if they fail to agree. Thus, all subcoalitions of P can veto any outcome different from $\boldsymbol{d}$, while cooperation by all agents is required in order to achieve another outcome. The existence of a point in S which strictly dominates $\boldsymbol{d}$ guarantees that all agents are nontrivially involved in the bargaining problem. The compactness of S is a technical assumption, convexity follows if the agents may jointly randomize between outcomes.

For simplicity, we will assume that the utility functions are normalized such that the vector $\boldsymbol{d}$ is always the origin of R^P. We may then identify any bargaining problem $(S,\boldsymbol{d})$ by the set S only. We will also restrict the family of bargaining problems under consideration to sets S with the following additional properties:

$A3$: S is a subset of R_+^P.
$A4$: If $\boldsymbol{x} \in S$ and $\boldsymbol{y} \in R_+^P$ such that $\boldsymbol{y} \leqq \boldsymbol{x}$, then $\boldsymbol{y} \in S$.

$A4$ is usually referred to as *comprehensiveness* and amounts to assuming the free disposal of utility. Let Σ^P be the class of bargaining problems S for the set P of agents, such that S satisfies $A1$–$A4$. Such problems will be referred to as *choice problems*, and will be denoted S, S', T etc. A *solution* is defined to be a function $F:\Sigma^P \to R_+^P$, such that $F(S) \in S$ for all $S \in \Sigma^P$. Given $S \in \Sigma^P$, the vector $F(S)$ is called the *solution outcome* to S, interpreted as that compromise which is in some sense a best resolution of the conflict among the agents in P.

We have now given a description of the basic model of the collective choice problem. One can now proceed, as Nash (1950) did, to look for a set of axioms that would guide the agents in their search of a fair compromise. Nash suggested four axioms, which we state below.

(1) *Pareto-optimality* (PO): For all $S \in \Sigma^P$, if $\boldsymbol{x} = F(S)$, then $\boldsymbol{x} \geqq \boldsymbol{y}$ for all $\boldsymbol{y} \in S$.

Let Γ be the family of permutations on P. We will also sometimes treat $\gamma \in \Gamma$ as a function from R^P to itself, defined by $\boldsymbol{y} = \gamma(\boldsymbol{x})$ if $y_{\gamma(i)} = x_i$ for all $i \in P$.

(2) *Symmetry* (SY): For all $S \in \Sigma^P$, if $\gamma(\boldsymbol{x}) \in S$ for all $\boldsymbol{x} \in S$ and all $\gamma \in \Gamma$, then $F_i(S) = F_j(S)$ for all $i,j \in P$.

Symmetry says that if the geometry of a choice problem S does not distinguish between the agents, then the solution should not do so either. A slightly stronger variant of this axiom, that we shall make use of later on, is:

Anonymity (AN): For all $S \in \Sigma^P$, for all $\gamma \in \Gamma$, $F[\gamma(S)] = \gamma[F(S)]$.

Anonymity states that permuting the names of the agents should not change the solution outcome. Let Λ be the family of functions λ from R^P to itself, such that for some $\boldsymbol{a} \in R^P_{++}$, $\lambda_i(\boldsymbol{x}) = a_i x_i$ for all $i \in P$ and all $\boldsymbol{x} \in R^P$.

(3) *Scale Invariance* (S.INV): For all $S \in \Sigma^P$, for all $\lambda \in \Lambda$, $F[\lambda(S)] = \lambda[F(S)]$.

S.INV requires that a rescaling of the utility representation of one or more agents by a positive linear transformation should rescale the solution outcome in the same way. It can be seen as a reflection of the fact that von Neumann–Morgenstern utility functions are only unique up to positive affine transformations. The reason why the transformations λ contain no constant term, is that we have already used up this degree of freedom by fixing the disagreement point $\boldsymbol{d}$ at the origin.

(4) *Independence of Irrelevant Alternatives* (IIA): For all $S, S' \in \Sigma^P$, if $S' \subset S$ and $F(S) \in S'$, then $F(S') = F(S)$.

In other words, if S' is obtained from S by narrowing down the set of feasible alternatives while keeping the solution outcome to S a feasible alternative in S', then the solution outcomes to the two choice problems should be the same.

Nash showed that there exists one and only one solution that satisfies PO, SY, S.INV, and IIA. It is the solution that for all S picks the unique outcome that maximizes the product of the agents' utility levels on S. Strictly speaking, Nash stated his result only for the two-person bargaining problem, presumably because in a situation with more than two agents, there might be room for coalitions, a feature which is not captured by the model.

Nash's model has been elaborated by Harsanyi (1959, 1963, 1977), who showed how the problem of solving n-person bargaining problems could be reduced to the more familiar one of solving two-person problems. He argued that in any n-person bargaining problem, a particular payoff vector "... will represent the equilibrium outcome of bargaining among the n players only if no pair of players i and j has any incentive to redistribute their payoffs between them, as long as the other players' payoffs are kept constant" (Harsanyi, 1977, p 196). This condition, which we will refer to as ***bilateral stability***, was shown to imply that if, among a group of n agents, all two-person bargaining problems were solved by the two-person Nash solution, then all n-person bargaining problems had to be solved by the n-person Nash solution.

Before we give a formal statement of Harsanyi's condition, which differs from the ones already introduced by involving a varying number of agents, it will be convenient to modify the basic model by following Thomson (1983a), who deals with this case in a more explicit fashion.

Let there be a fixed, countably infinite set I of agents that may potentially become involved in some collective choice problem, and let $\boldsymbol{P}$ be the set of finite subsets of I. I may be taken to be the set of natural integers. Elements of $\boldsymbol{P}$ are denoted by P, P', Q, etc. For all $P \in \boldsymbol{P}$, let Σ^P be the set of all choice problems for the set P of agents. We redefine a *solution* to be a function F:

$$F\colon \bigcup_{P \in \boldsymbol{P}} R^P_+ \rightarrow \bigcup_{P \in \boldsymbol{P}} R^P_+ \quad ,$$

such that for all $P \in \boldsymbol{P}$ and all $S \in \Sigma^P$, $F(S)$ is an element of S. For all $P \in \boldsymbol{P}$, the restriction of F to Σ^P is called the ***P-component*** of F.

Given $Q \in \boldsymbol{P}$, a choice problem $T \in \Sigma^Q$, a point $\boldsymbol{x}$ in T, and a subset P of Q, we define the set $t_P^{\boldsymbol{x}}(T)$ to be the intersection of T with a copy of the hyperplane R^P through $\boldsymbol{x}$ (see Figure 6.1). Given $\boldsymbol{x}$ in R^Q, we denote by $\boldsymbol{x}_P$ the projection of $\boldsymbol{x}$ on R^P. We may now state the condition of

Bilateral Stability (B.STAB): For all $Q,P \in \boldsymbol{P}$ with $P \subset Q$ and $|P| = 2$, for all $T \in \Sigma^Q$ and all $S \in \Sigma^P$, if $\boldsymbol{x} = F(T)$ and $S = t_P^{\boldsymbol{x}}(T)$, then $F(S) = \boldsymbol{x}_P$.

It is illustrated in Figure 6.1 for $|Q| = 3$ and $|P| = 2$.

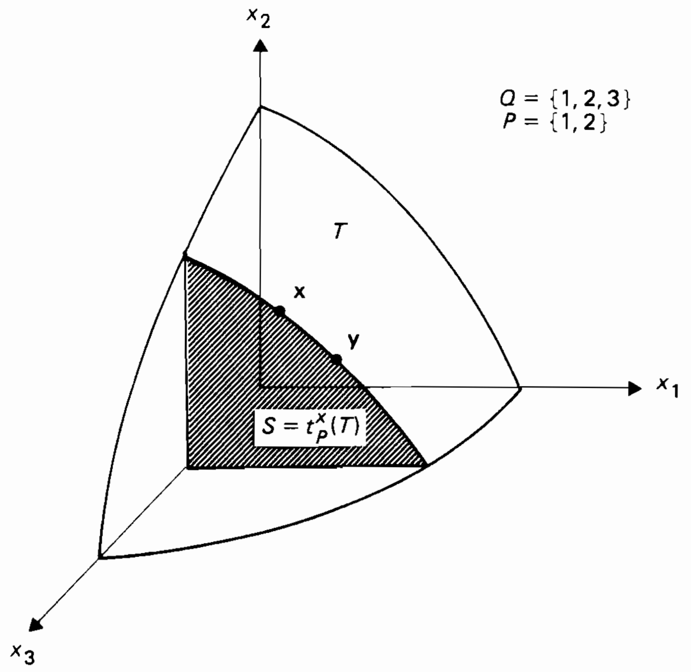

Figure 6.1 The axiom of bilateral stability.

Harsanyi motivates this condition by pointing out that a rational agent i will not accept a tentative agreement $\boldsymbol{x}$ for the bargaining problem T if he has reason to believe that he could successfully force some other agent j to make a concession in his favor. Suppose that the agents are all familiar with Nash's solution to the two-person bargaining problem, and that it is common knowledge among the agents that two-person problems are solved by that solution. Consider then agent 1, who is looking at the bargaining problem S involving only agent 2 and himself, by keeping the utility level of agent 3 fixed at x_3.

If the Nash outcome to this two-person bargaining problem is (y_1, y_2) where $y_1 > x_1$, then agent 1 will not accept x_1 but will demand y_1, arguing that agent 2 should lower his claim accordingly, by referring to their common knowledge of two-person bargaining theory. Whether agent 2 accepts or not does not really matter: if agent 1 rejects $\boldsymbol{x}$, then $\boldsymbol{x}$ cannot be the solution outcome to the three-person bargaining problem.

Seen positively, this means that a utility vector $\boldsymbol{x}$ can be the solution outcome to the $|Q|$-person choice problem T only if it agrees with the solution outcomes to all two-person subproblems S obtained from T by keeping the utility levels to all but two of the agents constant at the original outcome.

Because there seems to be no ***a priori*** reason why a dissatisfied agent should limit himself to challenging only one other agent at a time for concessions, it seems natural to consider the following generalization of B.STAB, which we call

> ***Multilateral Stability*** (M.STAB): For all $Q, P \in \boldsymbol{P}$ with $P \subset Q$, for all $T \in \Sigma^Q$ and all $S \in \Sigma^P$, if $\boldsymbol{x} = F(T)$ and $S = t_P^{\boldsymbol{x}}(T)$, then $F(S) = \boldsymbol{x}_P$.

In other words, the solution should be stable, not only with respect to two-person subproblems, but also with respect to subproblems involving any subset of the original group. As a principle of fair division, the axiom can be interpreted as a consistency requirement on the notion of fairness, saying that an allocation should not be declared a fair compromise for a given set of agents if it is unfair for some subset of those agents. As pointed out by Balinski and Young (1982), this seems to be a very natural condition to impose on the notion of fairness. Independently of Harsanyi's work, they have used an axiom in this spirit, called ***uniformity***, in their development of a theory of apportionment, e.g., for allocating seats in a parliament among political parties in agreement with the proportion of votes obtained by each party.

We close this section by introducing one more axiom and some additional notation that will be useful later on. The axiom is one of continuity, which states that similar choice problems should have similar solution outcomes:

> ***Continuity*** (CONT): For all $P \in \boldsymbol{P}$, for all $S \in \Sigma^P$, if $\{S^\nu\}$ is a sequence from Σ^P, converging in the Hausdorff topology to S; then $\{F(S^\nu)\}$ converges to $F(S)$.

The following additional notation will be used. Given P in $\boldsymbol{P}$, $\boldsymbol{e}_P$ is the vector in R^P with all of its coordinates equal to 1. If A is a subset of R_+^P, we denote by ***cch*** $\{A\}$ the ***convex and comprehensive hull*** of A, defined to be the smallest set containing A which satisfies $A1$, $A3$ and $A4$. Given S in Σ^P, we let $PO(S)$ denote the set of Pareto-optimal points in S, i.e.,

$$\mathrm{PO}(S) \equiv \{\boldsymbol{x} \in S \mid \nexists \boldsymbol{y} \in S,\ \boldsymbol{y} \geq \boldsymbol{x}\} \ .$$

Similarly,

$$\mathrm{WPO}(S) \equiv \{\boldsymbol{x} \in \mid \nexists \boldsymbol{y} \in S,\ \boldsymbol{y} > \boldsymbol{x}\}$$

is the set of weakly Pareto-optimal points in S. For example, if S is a rectangle in R_+^P with one corner at the origin, all interior points on the top edge of S will be weakly Pareto-optimal but not Pareto-optimal.

3 The Results

In this section, we give an outline of some recent results involving the stability axiom. In Section 3.1 we present a new characterization of the Nash solution, Section 3.2 is concerned with the Leximin solution, and in Section 3.3 we discuss a family of collectively rational and "decentralizable" solutions. Proofs will not be given here, but may be found in Lensberg (1982a,b, and 1983), respectively.

3.1 *Stability and the Nash solution*

We begin by stating the following two theorems, due to Nash (1950) and Harsanyi (1959):

Theorem 1 (Nash): *A solution F satisfies PO, SY, S.INV, and IIA if and only if for all $P \in \boldsymbol{P}$ and all $S \in \Sigma^P$,*

$$F(S) = N(S) \equiv \operatorname{argmax}\left\{ \prod_{i \in P} x_i \mid x \in S \right\} .$$

Theorem 2 (Harsanyi): *If a solution F satisfies CONT and B.STAB, and if F coincides with the Nash solution N for two-person problems, then $F = N$.*

Theorem 2 demonstrates how B.STAB can be applied to reduce the problem of solving n-person problems to one of solving two-person problems. In particular, if it is known *ex ante* that two-person problems are solved by the Nash solution, one is left with no degrees of freedom as regards the choice of a suitable n-person solution.

What then if nothing is known at the outset about the nature of the two-person solution? What kind of analytical power does the stability axiom have in this more general situation? It turns out that it can be used to give the following alternative characterization of the Nash solution:

Theorem 3: *A solution F satisfies PO, AN, S.INV and M.STAB if and only if it is the Nash solution.*

If we compare this result with Nash's own characterization, we see that except for a strengthening of SY to AN, the only difference is that IIA has been replaced by M.STAB. This is interesting, since the axiom of IIA has been somewhat controversial within the bargaining tradition (cf. Luce and Raiffa, 1957), and because of that, some authors (e.g. Kalai and Smorodinsky, 1975; Roth, 1977; Thomson, 1981) have replaced it with other axioms, and have arrived at different solutions. Here, however, we replace IIA by a version of Harsanyi's stability condition and still arrive at the Nash solution. Thus, it seems that the Nash solution does not rest so heavily on the axiom of IIA as is often thought.

As regards the connection to Harsanyi's theorem, we observe that, except for dropping the hypothesis that two-person problems are solved by the Nash solution, Theorem 3 uses the stronger version of the stability axiom, while CONT is not needed. Alternatively, we could weaken M.STAB to B.STAB, impose CONT and obtain the following variant to Theorem 3:

Theorem 4: *A solution F satisfies PO, AN, S.INV, CONT, and B.STAB if and only if it is the Nash solution.*

Going back to Theorem 3, it turns out that this is not the strongest result that one can prove. Specifically, the axiom of Pareto-optimality may be considerably weakened and still permit a characterization of the Nash solution. Suppose we weaken PO to require only that all one-person choice problems should be solved optimally:

Individual optimality (IO): For all $P \in \boldsymbol{P}$ with $|P| = 1$, for all $S \in \Sigma^P$, $F(S) = \max\{\boldsymbol{x} \mid \boldsymbol{x} \in S\}$.

Because IO and M.STAB together imply Pareto-optimality, we obtain:

Theorem 5: *A solution F satisfies IO, AN, S.INV and M.STAB if and only if it is the Nash solution.*

Let us also compare this result with Theorem 1. Nash's axioms seem to fall into two categories that are qualitatively quite different. The first category consists of SY and S.INV, which state that the solution outcome should not depend on information that is not contained in the model (Nash, 1953; Roth, 1979). In particular, S.INV is a reflection of the fact that von Neumann-Morgenstern utility functions are only unique up to positive affine transformations. In the second category are PO and IIA, both of which demand some form of collective rationality of the agents. Theorem 5 employs slightly modified versions of Nash's axioms in the first category, and replaces those in the second category by IO and M.STAB, both of which express a kind of individual (rather than collective) rationality when interpreted in a bargaining context.

The kind of collective rationality expressed by IIA can be seen more clearly by rephrasing it to say that if some alternative was declared to be "best" among a set of feasible alternatives, then it must also be "best" among any subset of those alternatives. This means in particular (Roth, 1979) that if the set of possible alternatives is expanded, then the solution either selects one of the new alternatives available, or it selects the solution outcome to the original problem. On the other hand, IIA does not have anything to say as to how the solution outcome should change *if* it changes as a result of an expansion in the set of feasible alternatives. For example, one might feel that if the set of alternatives is expanded in a direction which is particularly favorable to some agent, then that agent should gain, or at least should not be worse off, as a result of such a change in the problem.

Several authors have proposed and used axioms that express such a condition of *monotonicity* (Kalai and Smorodinsky, 1975; Kalai 1977a; Thomson and Myerson, 1980). In the next section we study the consequences of imposing such an axiom in conjunction with M.STAB.

3.2 *Stability and the Leximin solution*

We begin by stating Kalai's (1977a) version of the monotonicity axiom:

Individual Monotonicity (I.MON): For all $Q \in \boldsymbol{P}$, for all $S, S' \in \Sigma^Q$, if $S \subset S'$ and $S \cap R^P = S' \cap R^P$ where $P = Q - i$ for some $i \in Q$, then $F_i(S') \geqq F_i(S)$.

The motivation for this axiom is clear: The maximal set of utility vectors attainable in S by the members of $P = Q - i$ is $S \cap R^P$, and this set is not affected by the expansion from S to S'. This is illustrated in Figure 6.2, where $Q = \{1,2\}$. On the other hand, for each vector of utility for the set P, the maximal utility attainable by agent i increases, or at least does not decrease. It is in this sense that the expansion can be seen to be in agent i's favor, and it is natural then to require that this agent should not lose.

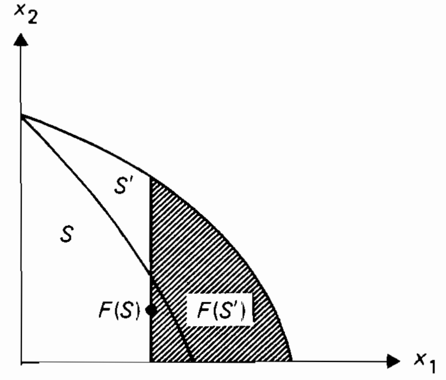

Figure 6.2 The axiom of individual monotonicity.

What do we get if we add individual monotonicity (I.MON) to the list of axioms in Theorem 3? The answer is "nothing", because the Nash solution does not satisfy I.MON, as shown in Figure 6.3. Thus, if we want the solution to satisfy I.MON, then some other axiom in Theorem 3 must go. The question is which one. If we look at Figure 6.3 again, we see that the reason why the Nash solution does not satisfy I.MON is that the level curves of the Nash product $\prod_{i \in P} x_i$ permit too much trade-off between the utility levels of agents 1 and 2. As it turns out (see Section 3.3) the axiom of Nash responsible for the particular shape of those level curves is scale invariance.

The question then is whether there are any solutions that satisfy PO, AN, I.MON and M.STAB. One possible candidate is the ***Egalitarian solution*** $\boldsymbol{E}$, which for each choice problem S picks the unique weakly Pareto-optimal point having equal coordinates (see Myerson, 1981; Thomson, 1983b). This solution does not admit any trade-off between the utilities of different agents, and so it would not violate I.MON in the example given in Figure 6.3. However, it satisfies neither PO nor M.STAB as is clear from Figure 6.4, where $Q \equiv \{1,2,3\}$, $P \equiv \{2,3\}$ and $T \equiv cch\,\{(1,2,3)\}$.

We have $E(T) = \boldsymbol{e}_Q = \boldsymbol{z}$ and $t_P^{\boldsymbol{z}}(T) = cch\{(2,3)\} = S$, so that $E(S) = 2\boldsymbol{e}_P$. Because $\boldsymbol{e}_Q$ is not a Pareto-optimal point of T, E does not satisfy PO, and because $2\boldsymbol{e}_P \neq \boldsymbol{z}_P = \boldsymbol{e}_P$, E does not satisfy M.STAB.

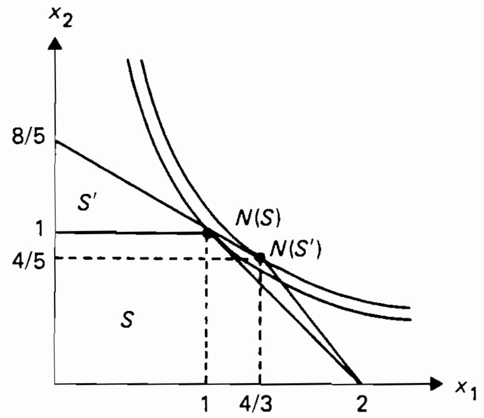

Figure 6.3 The Nash solution does not satisfy I.MON

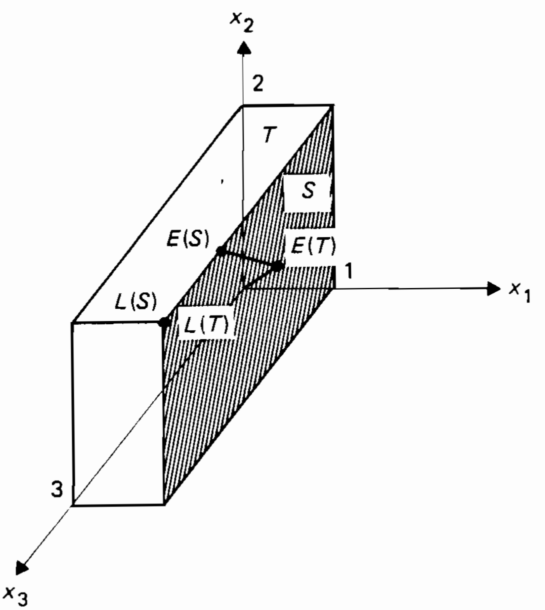

Figure 6.4 The Egalitarian solution satisfies neither PO nor M.STAB, but the Leximin solution does.

The Egalitarian solution is closely related to the Rawlsian maximin criterion (Rawls, 1971) which selects a feasible alternative maximizing the utility of the worst-off individual. In general, there may be more than one such alternative, as shown in Figure 6.4. Sen (1970) has suggested the following lexicographic extension of the Rawlsian maximin criterion which eliminates this indeterminacy. First, maximize the utility of the worst-off individual, then do the same for the next to worst-off individual, and so on, until all possibilities for increasing the utility of individuals has been exhausted. The solution obtained in this way is called the *Leximin solution* and is denoted L. It is illustrated in Figure 6.4, which also shows that L satisfies both PO and M.STAB in the given example. In fact, we have the following theorem:

Theorem 6: *A solution satisfies PO, AN, I.MON, and M.STAB if and only if* $F = L$, *the Leximin solution.*

Observe that the list of axioms used in Theorem 6 differs from the one used to characterize the Nash solution in Theorem 3 only in that S.INV has been replaced by I.MON. (Imai, 1983, has given a characterization of the Leximin solution that parallels Theorem 1 in a similar way.) Now, S.INV can be interpreted as a condition which rules out interpersonal comparisons between agents whose preferences are represented by (cardinal) von Neumann–Morgenstern utility functions. Theorems 3 and 6 show that S.INV and I.MON are in a sense polar opposites when used in conjunction with the three other axioms. The Leximin solution exploits to a maximum degree the possibilities for interpersonal comparability of relative utility levels that become available when S.INV is dropped, by admitting no trade-off between the utility levels of different agents.

One problem with the Leximin solution is that it is not continuous. This fact can be seen by considering any sequence $\{T^\nu\}$ of choice problems converging to the problem T depicted in Figure 6.4, such that each T^ν is strictly convex (in R^P_+). Then $L(T^\nu) = E(T^\nu)$ for all T^ν in the sequence, which means that $\{L(T^\nu)\}$ converges to $E(T)$. Because $E(T) = (1,1,1)$ while $L(T) = (1,2,3)$, this is a violation of CONT.

Thus, under the Leximin solution, it is not always the case that similar choice problems have similar solution outcomes. This might cause someone who was supposed to use it, say in a cost–benefit analysis, to worry about the quality of his data. Although there may be other problems to worry about in connection with implementing a collective decision rule, it will nevertheless be of interest to investigate the consequences of imposing continuity as an axiom in the model. This will be done in the next section.

3.3 *Stability and Collective Rationality*

Although the Nash solution and the Leximin solution are different in many respects, they have one thing in common: both are consistent with the maximization of some ***ordering*** on the space of alternatives; that is, a binary relation which is transitive, reflexive and complete. In the terminology of Richter (1971), such solutions are said to be *(collectively) rational*.

The Nash solution is collectively rational: for each $P \in \boldsymbol{P}$ and each $S \in \Sigma^P$, $N(S) = \{\boldsymbol{x} \in S \mid \boldsymbol{x} \succsim^N_P \boldsymbol{y}$ for all $\boldsymbol{y} \in S\}$, where the ordering $\succsim^N_P$ is defined on R^P_+ by

$$\boldsymbol{x} \succsim^N_P \boldsymbol{y} \quad \text{if and only if} \quad \prod_{i \in P} x_i \geq \prod_{i \in P} y_i \quad .$$

As regards the Leximin solution, Imai (1983) has shown that for all P and S, $L(S) = \{\boldsymbol{x} \in S \mid \boldsymbol{x} \gtrsim_P^L \boldsymbol{y}$ for all $\boldsymbol{y} \in S\}$, where for each P, $\gtrsim_P^L$ is the (symmetric) lexicographic extension of the ordering $\gtrsim_P^E$ of R_+^P defined by

$$\boldsymbol{x} \gtrsim_P^E \boldsymbol{y} \quad \text{if and only if} \quad \min_{i \in P} x_i \geqq \min_{i \in P} y_i \quad .$$

The property of being collectively rational is enjoyed by many other solutions as well. To fix ideas, one may think of the orderings $\gtrsim_P^N$ and $\gtrsim_P^L$ as Bergson–Samuelson social welfare functions (Bergson, 1938; Samuelson, 1947). It is then clear that from any Bergson–Samuelson social welfare function one obtains a social choice function (a solution) provided the maxima always exist and are unique on the relevant domain of choice problems.

A condition that is often imposed on social orderings is *separability* or *independence of unconcerned individuals* as it is also sometimes called. This condition (due to Fleming, 1952) says that if the utility levels for a subset of the agents of society is the same for some pair of alternatives, then the social ordering of those alternatives should not depend on the utility levels of those agents. This means that if $\gtrsim_Q$ is a social ordering of the utility space R_+^Q for a group Q of agents, then for every $P \subset Q$, the ordering $\gtrsim_P$ obtained from $\gtrsim_Q$ by restricting $\gtrsim_Q$ to any hyperplane parallel to R^P must be the same for all such hyperplanes. If the ordering $\gtrsim_Q$ is continuous, it can be shown (Debreu, 1960) that it has an *additively separable numerical representation*. In other words, there is a real-valued function f^Q on R_+^Q such that $f^Q(\boldsymbol{x}) \geqq f^Q(\boldsymbol{y})$ if and only if $\boldsymbol{x} \gtrsim_Q \boldsymbol{y}$, where f^Q is of the form

$$f^Q(\boldsymbol{x}) = \sum_{i \in Q} f_i(x_i) \quad .$$

The condition of separability is satisfied by any of the commonly used Bergson–Samuelson social welfare functions, such as *classical utilitarianism*, the Rawlsian maximin criterion and its lexicographic extension, as well as the *Nash social welfare function* (see d'Aspremont and Gevers, 1977; Deschamps and Gevers, 1978; Kaneko and Nakamura, 1978).

Intuitively, the stability axiom is a natural counterpart to separability in the sense that it imposes on a solution much the same requirement that separability imposes on a social ordering. What is more interesting, and perhaps less obvious, is that it imposes on the solution a fair amount of collective rationality as well, as the next theorem shows.

Let $\boldsymbol{F}$ be the family of all sequences $\{f_i\}_{i \in I}$ of strictly increasing, extended real-valued functions, where each f_i is defined on $R_+^{\{i\}}$, such that for all $P \in \boldsymbol{P}$, the function

$$f^P \equiv \sum_{i \in P} f_i$$

is strictly quasi-concave, i.e., the set of $\boldsymbol{x}$ such that $f^P(\boldsymbol{x}) \geqq f^P(\boldsymbol{y})$ is strictly convex in R_+^P for each fixed $\boldsymbol{y}$.

Theorem 7: *A solution F satisfies* PO, CONT *and* B.STAB *if and only if there exists a sequence of functions $\{f_i\}_{i \in I}$ from $\boldsymbol{F}$ such that for all $P \in \boldsymbol{P}$ and all $S \in \Sigma^P$, $\boldsymbol{F}(S) = \operatorname{argmax}\{\sum_{i \in P} f_i(x_i) \mid \boldsymbol{x} \in S\}$.*

It is interesting to see this result in relation to the problem mentioned earlier of attaining consistency in a system of decentralized public decision making, where each decentralized unit is trying to achieve a fair allocation among its own clients. When interpreted in this context, the stability axiom requires that if an allocation is to be considered globally fair for the whole society, then each decentralized unit should also regard the allocation as fair when considering only its own clients. This is clearly a necessary condition for decentralized public decision making to be consistent with some global notion of fairness: if it were not satisfied, then the globally fair allocation could never be obtained, because there would always be some local unit that would want to move away from it.

Theorem 7 shows that bilateral stability, when imposed in conjunction with PO and CONT, has some very precise implications concerning the nature of the decision rules that will have to be followed by the decentralized units. Firstly, the global (and local) notions of fairness must correspond to some Bergson–Samuelson social welfare function and the decision rules must be collectively rational. Secondly, the global social welfare function must be additively separable, which means that it can be split up and distributed among the decentralized units in such a way that each unit can make its decisions based on information about its own clients only. Clearly, there is in general no guarantee that this type of decentralized decision making will actually lead the society towards the globally fair allocation, but the point is that if the globally fair allocation exists at all (which it does, according to Theorem 7), then the decision rules will have to be of this form.

It can be shown that the axioms used in Theorem 7 are independent, in the sense that removing any one of them will permit solutions that are not collectively rational. Conversely, because the theorem characterizes a whole family of solutions, it is a useful framework for analyzing the implications of adding more axioms to the list in Theorem 7.

Adding symmetry (SY) to the list of axioms implies that all the functions f_i must be identical. Thus, all the information needed to solve any choice problem is contained in a single function of a real variable.

Next, we consider a weaker version of scale invariance, namely ***homogeneity*** (HOM), which says that if two choice problems are identical, except for a scale change, then their solution outcomes should also be identical, except for the same scale change.

Adding homogeneity to the list of axioms in Theorem 7 implies that the functions $\sum_{i \in P} f_i$ must be homothetic for all $P \in \boldsymbol{P}$. This means (Eichhorn, 1978, Theorem 2.2.1) that, except for arbitrary constant terms, there exists $\rho > -1$ and a sequence $\{\alpha_i\}_{i \in I}$ of positive real numbers such that $f_i(x_i) = -(\alpha_i / \rho)x_i^{-\rho}$ for all $i \in I$ if $\rho \neq 0$, and $f_i(x_i) = \alpha_i \log x_i$ for all $i \in I$ if $\rho = 0$. Thus $\sum_{i \in P} f_i$ is a CES type function for all $P \in \boldsymbol{P}$. If symmetry is also imposed, then f_i, and hence α_i, must be the same for all i. As $\rho \rightarrow 0$, we then obtain the Nash social welfare function, as $\rho \rightarrow -1$ we obtain classical utilitarianism, and as $\rho \rightarrow \infty$ we get the Rawlsian maximin criterion (see Roberts, 1980, for related results in the Arrow tradition of social choice theory).

Alternatively, dropping symmetry and strengthening homogeneity to scale invariance implies that $\rho = 0$, yielding a whole family of *nonsymmetric* Nash solutions. This family of solutions has been studied by Harsanyi and Selten (1972), Kalai (1977b), and Roth (1979).

It should be noted that the solutions U and E, derived from utilitarianism and the Rawlsian maximin criterion, respectively, do not themselves qualify as solutions, since the underlying social welfare functions do not always yield unique solution outcomes on the domain considered here. One may then consider single-valued selections, at the cost of relaxing either PO or CONT. For example, keeping PO and dropping CONT will admit the Leximin solution studied in the previous section.

4 Conclusion

We have attempted in this chapter to view the problem of allocating costs and benefits among a group of individuals as one of bargaining or fair division. A generalization, due to Thomson (1983a), of Nash's (1950) model of the bargaining problem has been used to explore the consequences of an axiom, due to Harsanyi (1959), which can be seen as a requirement that the solution to the allocation problem should be decentralizable in a certain sense.

Although this way of looking at the problem is fairly abstract, it does give some insight that may be useful when trying to solve real allocation problems. Our main result is that a certain amount of collective rationality in the decision-making process is a necessary prerequisite for an allocation procedure to be decentralizable. Thus, when faced with a practical problem, the theory tells us to look for a social welfare function in order to rank the given physical alternatives. In order to take care of the decentralization aspect, the social welfare function should be additively separable in individual utility levels. For practical purposes, this means that the composite function $f_i(x_i(\cdot))$, where x_i is agent i's unobservable utility function and f_i is the i th component of the social welfare function, can be looked upon as a standard of living index for agent i, depending on the physical benefits or costs allocated to i.

This indicates that the problem of solving real allocation problems in a decentralized setting is one of establishing a procedure for project evaluation in the public sector. In both cases, the basic problem consists in specifying an appropriate set of standard of living indices for (groups of) individuals to be used as a criterion for selecting among the physical alternatives available in any given choice situation. Moreover, in a given choice situation, a description of the problem consists in specifying the effect of each physical alternative on the standard of living index for each individual or group of individuals. Thus, our results suggest that if one is interested in normative aspects of the allocation problem in a decentralized setting, then the problem can be attacked by means of the familiar tools of cost–benefit analysis.

Acknowledgments

I am indebted to H.P. Young and the participants at the IIASA conference for extensive comments and helpful discussion. Special thanks are due to William Thomson with whom I have had the pleasure to collaborate while working on this topic. Financial support from the University of Rochester is gratefully acknowledged.

References

Bergson, A. (1938) A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics* 52: 310–34.

Balinski, M. and Young, H.P. (1982) *Fair Representation: Meeting the Ideal of One Man, One Vote* (New Haven, CT: Yale University Press).

d'Aspremont, C. and Gevers, L. (1977) Equity and the informational basis of collective choice. *Review of Economic Studies* 44: 199–209.

Deschamps, R. and Gevers, L. (1978) Leximin and utilitarian rules: a joint characterization. *Journal of Economic Theory* 17: 143–63.

Debreu,G. (1960) Topological methods in cardinal utility theory, in K.J. Arrow, S. Karlin, and P. Suppes (Eds) *Mathematical Methods in the Social Sciences* (Stanford, CA: Stanford University Press).

Eichhorn, W. (1978) *Functional Equations in Economics* (Reading, MA: Addison–Wesley).

Fleming, M. (1952) A cardinal concept of welfare. *Quarterly Journal of Economics* 66: 366–84.

Harsanyi, J.C. (1955) Cardinal welfare, individual ethics and interpersonal comparisons of utility. *Journal of Political Economy* 63: 309–21.

Harsanyi, J.C. (1959) A bargaining model for the cooperative n-person game, in A.W. Tucker and R.D. Luce (Eds) *Contributions to the Theory of Games*, Vol. IV, (Princeton, NJ: Princeton University Press).

Harsanyi, J.C. (1963) A simplified bargaining model for the n-person cooperative game. *International Economic Review* 4: 194–220.

Harsanyi, J.C. (1977) *Rational Behavior and Bargaining Equilibrium in Games and Social Situations* (London: Cambridge University Press).

Harsanyi, J.C. and Selten, R. (1972) A generalized Nash solution for two-person bargaining games with incomplete information. *Management Science* 18(5)II: 80–106.

Imai, H. (1983) Individual monotonicity and lexicographic maximin solution. *Econometrica* 51: 389–401.

Kalai, E. (1977a) Proportional solutions to bargaining situations: interpersonal utility comparisons. *Econometrica* 45: 1623–30.

Kalai, E. (1977b) Nonsymmetric Nash solutions and replications of two-person bargaining. *International Journal of Game Theory* 6: 129–33.

Kalai, E. and Smorodinsky, M. (1975) Other solutions to Nash's bargaining problem. *Econometrica* 43: 513–18.

Kaneko, M. and Nakamura, K. (1979) The Nash social welfare function. *Econometrica* 47(2): 423–35.

Lensberg, T. (1982a) *Stability and the Nash Solution* (mimeo) Norwegian School of Economics and Business Administration, Bergen.

Lensberg, T.(1982b) *Stability and the Leximin solution* (mimeo) Norwegian School of Economics and Business Administration, Bergen.

Lensberg, T.(1983) *Stability and Collective Rationality*. Discussion Paper 03/83, Norwegian School of Economics and Business Administration, Bergen.

Luce, D. and Raiffa, H. (1957) *Games and Decisions* (New York: Wiley).

Myerson, R.B. (1981) Utilitarianism, egalitarianism and the timing effect in social choice problems. *Econometrica* 49: 883–97.

Nash, J.F. (1950) The bargaining problem. *Econometrica* 18: 155–62.

Nash, J.F. (1953) Two-person cooperative games. *Econometrica* 21: 128–40.

Raiffa, H. (1953) Arbitration schemes for generalized two-person games, in H.W. Kuhn and A.W. Tucker (Eds) *Contributions to the Theory of Games II*, Annals of Mathematical Studies, No. 28 (Princeton, NJ: Princeton University Press).

Rawls, J. (1971) *A Theory of Justice* (Cambridge, MA.: Harvard University Press).
Richter, M.K. (1971) Rational choice, in J.S. Chipman *et al.* (Eds) *Preferences, Utility, and Demand* (New York: Harcourt, Brace, Jovanovich).
Roberts, K.W.S. (1980) Interpersonal comparability and social choice theory. *Review of Economic Studies* 48: 421–39.
Roth, A.E. (1977) Independence of irrelevant alternatives and solutions to Nash's bargaining problem. *Journal of Economic Theory* 16: 247–51.
Roth, A.E. (1979) *Axiomatic Models of Bargaining* (New York: Springer-Verlag).
Samuelson, P.A. (1947) *Foundations of Economic Analysis* (Cambridge, MA: Harvard University Press).
Sen, A. (1970) *Collective Choice and Social Welfare* (San Fransisco: Holden-Day).
Thomson, W. (1981) A class of solutions to bargaining problems. *Journal of Economic Theory* 25(3): 431–41.
Thomson, W. (1983a) The fair division of a fixed supply among a growing population. *Mathematics of Operations Research* 8(3): 319–26.
Thomson, W. (1983b) Problems of fair division and the egalitarian solution. *Journal of Economic Theory* 31(2): 211–26.
Thomson, W. and Myerson, R.B. (1980) Monotonicity and independence axioms. *International Journal of Game Theory* 9: 37–49.

PART II APPLICATIONS

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 7

PERCEIVED ECONOMIC JUSTICE: THE EXAMPLE OF PUBLIC UTILITY REGULATION

Edward E. Zajac

1 Introduction

Why do public utility regulators, and the public generally, find it so hard to accept and apply the principles of economic efficiency – principles that are so obvious to trained economists? Is it simply that the public is woefully economically illiterate? Or is it the economists who are out of step, insisting that everyone march to their drummer, when in fact they are deaf to a more fundamental beat that drives society?

I first became involved in public utility regulation in 1965 when I was assigned from Bell Laboratories to an ATT task force. The task force's job was to prepare a strategy for Federal Communications Commission Docket 16258, launched as a comprehensive investigation of Bell System rate levels and rate structures and how they should be set so as to serve the public interest. In response to this and subsequent dockets (e.g., Dockets 18128 and 20003) the Bell System undertook studies to apply the most skilled neoclassical economic analysis to determine pricing policies in the best public interest, and these studies in turn spawned a large literature by non-Bell economists. All of this research profoundly deepened economists' understanding of public utility pricing's relation to economic efficiency (for an overview of this research see Sharkey, 1982, and Baumol *et al.*, 1982). However, it did not cause the FCC to scrap its fairness-based "distributed costs" pricing in favor of an economic efficiency based pricing policy.

My continual immersion in public utility regulation since then has gradually led me away from "the public is economically illiterate" and more toward the "economists are deaf" view. In my experience, economists are often deaf to pleas for economic justice because of their reliance on the "economic efficiency-income distribution" paradigm: for a given income distribution the economist's job is to recommend policy actions that will move the economy toward Pareto optimality through a Pareto improvement, while it is someone else's job to determine what the income distribution ought to be. In theorizing, this paradigm is very useful for sorting out positive and normative issues. But in the practical world, there are generally *no* Pareto improving actions; every policy change generates *some* losers. My observation is that economists generally ignore this fact and suggest "approximate Pareto improving" moves, perhaps ones where the winners greatly outnumber losers, or where winners could potentially compensate losers many times over, or where a large dead-weight loss could be eliminated.[1] Such "approximate Pareto improving" moves are of course a form of utilitarianism, with its ethical danger of harming the innocent in order to benefit the many. Even with a *single* loser, an approximate Pareto improvement may be, ethically, miles removed from a strict Pareto improvement.[2]

As long as policy makers ignored economists, their espousal of "approximate Pareto improving moves" in whatever guise was not an issue. However, recently professional economists have become members of and even headed regulatory bodies. They routinely train lawyers and graduates of schools of public administration to base policy on the facilitation of economic efficiency. Although I have observed that the FCC has shied away from embracing economic efficiency for pricing, we have seen the massive movement to deregulate airlines, trucking, financial institutions, and telephone utilities – a tribute in large measure to the leadership and persuasiveness of professional economists. The ethical grounding of economic analysis has thus become a serious business.

How is this serious business to be approached? To demand only strict Pareto improvements is to demand policy paralysis, since such improvements generally do not exist. Further, working too hard to avoid losers risks incurring the wrath of potential winners, who may feel that the receipt of justly deserved gains is being unjustly thwarted. But to implement approximate Pareto improving moves on the naive assumption that this will help achieve Pareto optimality is to risk imposing ethical fascism. At issue are ancient ethical questions such as individual versus societal welfare (should an individual's rights and freedom be violated to advance the social good?); the ends versus means dilemma, reflected currently in the tension between Rawls' end state theory of justice and Nozick's procedural justice theory (if we agree that a particular Pareto optimal end state is just, how do we get there justly?)

In my view, the starting point for approaching these issues is clear. At bottom, economic policy is driven by considerations of economic justice. That being the case, the further development of both positive and normative economic theory requires a deeper understanding of economic justice, especially of how it is perceived.

For example, the currently popular "economic" or "political" theories of regulation are models of self-interest groups interacting with legislators or regulators. But what causes the groups to form? One observes that the groups typically legitimate their actions by claims of correcting or preventing economic injustices. In my experience, economists usually dismiss such justice arguments as smokescreens for the furthering of self-interest and insist on approaching group formation or almost any other issue in terms of rational actors balancing benefits and costs at the margin. But if the justice arguments are such transparent frauds, why are they advanced in the first place and why are they given serious attention by the regulatory process? If a theory of "sincere justice" is to be rejected, then one must explain the success of "insincere justice" arguments, and one should be able to construct a theory of "insincere justice argumentation."

On the normative side, economic analysis is an indispensable calculus for dealing with economic justice. Its great virtue is its ability to calculate benefits and costs that are diffused throughout the population, as well as those that are concentrated in narrow self-interest groups. But the practical application of this calculus can be stymied by economists ignoring, or worse, being unintentionally high-handed about, justice issues. As a case in point, consider Katherine Sasseville, former chairwoman of the Minnesota Public Service Commission, discussing economists' proposals to deregulate telecommunications

> ... they [economists] are also idealists, hopelessly out of touch with reality. Like the flower children of the 1960s they live in a perfect world of harmony and joy where all conditions are at Pareto optimality, and perfect competition perfectly allocates all the resources of society, providing a free-market panacea for all the woes of man. Blinded by their belief in the theoretical world of competition, they have made the leap of faith to assume that what works in theory will work in the real world. Convinced that efficient allocation is fair and just allocation, they ignore the absence of any empirical evidence that, in the real world, unregulated competition can assure the fair distribution of an essential social good or service...Theoretical economists who dream perfervid dreams of perfect competition in communications services should leave Washington for a few months and travel about the small towns of America, listening to the people describe their goals, the public's idea of social goals for communications service. I have heard sworn testimony from people who watched their house or barn burn down because of unreliable telephone service. I also heard a woman testify that a baby is alive today because the telephone system at St. Mary's Village, Alaska, had been installed the week before the child became ill, and it was possible to summon the bush pilot and fly the child to a hospital (Sasseville, 1981, pp 111–114).

2 The Argument in Brief

This paper does not attempt to tell economists how to inject ethics into their analyses. Its aim is much more modest – to make a start on constructing a positive theory of how economic justice is perceived in public utility regulation. The paper argues that there are usually several economic justice principles that regulatory policymakers must take into account. These principles in effect form an unwritten constitution of economic justice,[3] or at least are elements of such a constitution. The unwritten constitution is tantamount to a descriptive theory of how the public perceives economic justice in public utility regulation. Part of this unwritten constitution is a direct outgrowth of the Fifth, Tenth, and Fourteenth Amendments to the United States Constitution. Other parts have their roots in the law on limited liability and bankruptcy. I summarize these principles into six "propositions" that form at least a part of the unwritten constitution.

Briefly, my thesis is that the noneconomist public, uneducated in the importance of Pareto optimality and gains through exchange, tends to see life as a struggle for a just share of an economic pie of fixed size. One member of society's getting a share of the pie is viewed as denial of that share to other members. The public recognizes life to be unfair, but still, in the struggle for a share of the economic pie, no one should receive less than a minimum amount (Proposition 1). Proposition 2 says that if circumstances make it feasible to distribute what are deemed to be equal shares, then such a distribution should occur. Proposition 3 states that if one has fairly acquired a share of the pie, it should not be confiscated; moreover, the share should not be lost because of circumstances beyond one's control (Proposition 4).

Proposition 5 deals with the perceived injustice of economic inefficiency. Individuals will balk at market impediments when the impediments cause them to lose potential gains through exchange. However, at the collective level, substantial evidence is required to engender the perception that some rearrangement of resources can result in a larger pie. Because the public has little conception of what constitutes Pareto improvement, the injustice of economic inefficiency is perceived not as a denial of gains through exchange, but as the work of a small group thwarting potential large-scale benefits for society. This is especially the case if those who thwart are seen as merely resisting the loss of some unfair advantage they already enjoy.

The last Proposition relates to the public's desire for due process mechanisms to prevent the unjust exercise of a utility's monopoly power.

The outline of the paper is as follows. First, we briefly review the role of economic justice in English and American economic history, then review and critique the scant literature on descriptive or positive theories of economic justice. This is followed by a discussion of the economic justice effects which form a constitutional framework for public utility policymakers.[4]

3 Historical Perspective

One starting point might be to try to apply to public utility regulation the old and well developed literature on the positive and normative ethical foundations of the law (see Murphy and Coleman, 1984, for an introduction to this literature). However, although much of the philosophical legal literature has direct relevance to our subject, it cannot be used once and for all to resolve all economic justice issues, for public utility regulation has special characteristics. By and large, regulatory bodies were created to short-cut existing legal processes in order to expedite the furthering of the common good. For example, the rules of evidence do not usually apply in regulatory hearings; rather, regulatory bodies' procedures are governed by administrative law. The bodies are generally created by enabling legislation that is revised seldom and only in response to large changes. (The FCC is still operating under the Federal Communications Act of 1934, in spite of the upheavals in the telecommunications industry.)

Laws are generally not passed to tell regulators how to rule on specific substantive issues (for an exception, see the New York State "Utility Bill of Rights" in the sequel Section 5.2.1). As long as the bodies operate within the administrative law and the enabling legislation, they have considerable freedom to set policy on such matters as rate structures and levels, costing methodology used to arrive at rates, and the research and development expenditures that are justifiably supported by rate payers. This in effect allows them to act as legislators. In adjudicating disputes that come before them, regulators also perform a judicial function (the former title of "hearing examiner" is gradually being replaced by "administrative law judge"), and their actions are subject to judicial review. Finally, regulatory bodies are usually a part of the executive branch of government. Thus, they perform a mixture of legislative, judicial, and executive functions.

Moreover, regulatory bodies generally do not resolve two-party disputes in the relative privacy of a law office or courtroom. Rather, they operate in the public fishbowl of coverage by the media, which often perceives regulatory bodies to be resolving disputes between "the public" and public utility corporations. Unlike the market, which offers no relief to those who object to a price rise, regulation provides a place for the citizen to complain if he or she feels a proposed water, gas, telephone, or electric power rate hike is too high. If the citizen forms a group with similarly situated citizens, say "senior citizens", or other residents in the citizen's community, or other small businessmen, and the group hires a lawyer and petitions for status as an intervenor, the citizen's complaint will probably be given serious attention. If the complaint appeals to the press as an example of David fighting the public utility Goliath in the name of justice, the citizen might succeed beyond expectation. On the other hand, utility executives must be very vigilant. A seemingly innocuous proposal for a policy change in rate level or structure for example, may galvanize citizens into actions that will capture the fancy of the press and make the utility appear to the public like an oppressive monster trying to take unfair advantage of helpless persons.

Hence, public utility regulators may have to be particularly responsive to the public's notions of economic justice. Like legislators and unlike judges, regulators need not be and often are not lawyers, and thus may not be conditioned to base their actions on a legal conceptual framework. Thus, the notions of economic justice that emerge in the regulatory process are liable to be intuitive rather than the finely honed result of argument and counter-argument by scholarly minds over several centuries.

A thesis that suggests itself is that the closer an institution is to immediate public accountability, the less it will be informed by a coherent economic logic and the more by primitive notions of morality and justice. This is in fact Posner's (1972) assertion about the common and statute laws:

> Our survey of the major common law fields suggests that the common law exhibits a deep unity that is economic in character...The common law method is to allocate responsibilities between people engaged in interacting activities in such a way as to maximize the joint value, or, what amounts to the same thing, minimize the joint cost of the activities. (p 98)
> ... If much of the common law seems informed by an implicit economic logic, the same cannot be said for statute law. Unsystematic as our survey of statute law has been in this book, it does suggest that statutes exhibit a less pervasive concern with efficiency and a much greater concern with wealth distribution. (p 327)

If anything, one might expect public utility regulation to be even more concerned with wealth distribution than is statute law.

The highly public nature of public utility regulation may give the framework or context of economic decisions an unusually significant role. As will be stressed in the remainder of this chapter, little empirical research has been done on "framing" or "contextual" effects on the public's perception of what is economic justice. However, history, experiment (what there is of it), and my own observation of public utility regulation suggest that contextual effects may be very important. We begin by considering the historical evidence.

3.1 *The Importance of Context on Perceptions of Economic Justice: Historical Evidence*

Modern economics training in the United States focuses so much on the market organization of an economy that it is easy to forget that other alternatives are possible. But the free market economy as a legitimate institution is a relatively modern concept, as stressed by Polanyi. According to Pearson's Comment on Polanyi (1977, p xxix):

> Polanyi was ... at pains to point out that the market-ordered institutional complex does not identify the economy in all societies. Whether we look to the evidence from anthropology or history, it is clear that the competitive market-money-price complex, operating in its legal context of private property and free contract and its 'economizing' cultural context, has either been absent or has played a subordinate role through most of man's history. ... The give and take relations between persons in regard to material things in these societies are typically embedded in a broad network of social and political commitments that do not allow the individual to maximize his 'economic' advantage in these relationships.

Polanyi's arguments are elaborated in his book *The Livelihood of Man* (1977). For example, he remarks (p 11):

> As regards man, we were made to accept the view that his motives can be described as either 'material' or 'ideal' and that the incentives on which everyday life is organized necessarily spring from the material motives...In fact, human beings will labor for a large variety of reasons so long as they form part of a definite social group. Monks traded for religious reasons, and monasteries became the largest trading establishments in Europe. The kula trade of the Trobriand Islanders, one of the most intricate barter arrangements known to man, is mainly an esthetic pursuit. Feudal economy depended largely on custom and tradition. With the Kwakiutl, the chief aim of industry seems to be to satisfy a point of honor. Under mercantile despotism, industry was often planned so as to serve power and glory.

Historically, in the English tradition, one of the "broad social and political commitments" mentioned by Polanyi has been a moral or ethical commitment, a heritage of the medieval Christian view of how an economy should operate "justly". As is well known, medieval Christian Europe held to the doctrine of the "just price". Phillips quotes Glaeser on the importance of this doctrine and its historical role in regulation (Phillips, 1969, p 52).

> Regulation of private industry has been attempted by government from earliest times. All attempts at such regulation owed much to a very ancient ideal of social justice, which, as applied to economic life by the early Church Fathers, became their very famous doctrine of ***justum pretium***, i.e., they opposed this idea to the contemporaneous doctrine of ***verum pretium***, i.e., 'natural price,' which the Roman law had derived from Stoic philosophy. As contrasted with the doctrine of natural price, which justified any price reached by agreement in effecting exchanges between willing buyers and willing sellers, the 'just price' doctrine drew attention to the coercion which may reside in economic circumstances, such as a food famine where a buyer is made willing by his economic necessities. Hence, in order to draw the sting

> of coercion, the early Church Fathers, following St. Augustine, considered only that trading to be legitimate in which the trade paid a 'just price' to the producer, and in selling, added only so much to the price as was customarily sufficient for his economic support. There was to be no unjust enrichment.[5]

"Just prices" in the grain markets prevailed as the official doctrine in England all through the 18th century, even though the doctrine progressively came to be disregarded. However, it always hovered in the background, to be appealed to especially in times of scarcity. But most interestingly, it captured a popular ethical view. E.P. Thompson calls this ethical outlook, the "paternalistic model" (Thompson, 1971, p 83) and describes its working in the grain markets as follows:

> The paternalist model existed in an eroded body of Statute law, as well as common law and custom. It was the model which, very often, informed the actions of Government in times of emergency until the 1770s; and to which many local magistrates continued to appeal. In this model, marketing should be, so far as possible, direct, from the farmer to the consumer. The farmers should bring their corn in bulk to the local pitching market; they should not sell it while standing in the field, nor should they withhold it in the hope of rising prices. The markets should be controlled; no sales should be made before stated times, when a bell would ring; the poor should have the opportunity to buy grain, flour, or meal first, in small parcels, with duly-supervised weights and measures. At a certain hour, when their needs were satisfied, a second bell would ring, and larger dealers (duly licensed) might make their purchases. Dealers were hedged around with many restrictions, inscribed upon the musty parchments of the laws against forestalling, regrating and engrossing, codified in the reign of Edward VI. They must not buy (and farmers must not sell) by sample. They must not buy standing crops, nor might they purchase to sell again (within three months) in the same market at a profit, or in neighboring markets, and so on. Indeed, for most of the eighteenth century the middleman remained legally suspect, and his operations were, in theory, severely restricted.

The 18th century was a time of great improvements in agriculture in England and in most years agricultural markets operated without incident and without great regard for the "eroded body of statute law" cited by Thompson. However, to quote Thompson again (p 79):

> Those years which brought English agriculture to a new pitch of excellence were punctuated by the riots – or as contemporaries often described them, the 'insurrections' or 'risings of the poor' – of 1709, 1740, 1756–7, 1766–7, 1773, 1782, and, above all, 1795 and 1800–1. This buoyant capitalist industry (agriculture) floated upon an irascible market which might at any time dissolve into marauding bands, who scoured the countryside with bludgeons, or rose in the market-place to 'set the price' of provisions at the popular level.

Thompson describes dramatically the fury of the mobs, in times of great scarcity, directed at the middlemen who "forestalled" the path of grain to the market but bought cheaply in bulk to sell dearly in small lots ("engrossing") or otherwise "unnecessarily" elevated the grain's final price ("regrating"). He also emphasizes that the mobs' actions were driven by a strong ethical purpose – to have the grain markets function according to the paternalistic model (p 111):

> the movement of the crowd from the market-place outwards to the mills and thence ... to farms, where stocks were inspected and the farmers ordered to send grain to market at the price dictated by the crowd – all this is commonly found.

Thus, the work of the mob was to "set the price" fairly even while confiscating the grain, when the mob could just as easily take the grain without payment.

Thompson argues that the paternalistic model gave way at the end of the 18th century to the "new political economy" (p 89):

> Few intellectual victories have been more overwhelming than that which the proponents of the new political economy won in the matter of the regulation of the internal corn trade. The model of the new political economy may, with convenience, be taken as that of Adam Smith.

Interestingly, Thompson points out that neither the paternalistic model nor the model of the new political economy were clearly verified empirically (p 91):

> Whereas the first appeals to a moral norm – what ought to be men's reciprocal duties – the second appears to say: 'this is the way things work, or would work if the State did not interfere'. And yet if one considers these sections of ***The Wealth of Nations*** they impress less as an essay in empirical enquiry than as a superb, self-validating essay in logic ... When we consider the actual organization of the eighteenth-century corn trade, empirical verification of neither model is to hand.

What about the effects of Adam Smith's great moral triumph on the economic development of the United States? What is American history as regards the morality of market versus nonmarket controls? A fascinating account of the interplay of the use of market and nonmarket controls is given by J.R.T. Hughes in *The Governmental Habit* (1977). He does not imbed his account in a framework of contextual effects on popular perceptions of economic justice. Nevertheless his account implies such effects. Briefly, the colonies took over into their laws the extensive controls over economic life that existed in contemporary English law. To quote Hughes (p 36):

> Two medieval concepts, the prohibition of usury and the doctrine of a just price, were imbedded in the colonial laws. ... Georgia, in mid-eighteenth century, passed a law that contained the actual medieval words, enjoining merchants against "forestalling, engrossing and unjust exactions".

In summary, the colonial period set the stage (p 49):

> Thus we see that during the colonial era virtually every aspect of economic life was subject to nonmarket controls. Some of this tradition would not survive, some would become even more powerful, while some would ascend to the level of federal control. The colonial background was like an institutional gene pool. Most of the colonial institutions and practices live on today in some form, and there is very little in the way of nonmarket control of the economy that does not have a colonial or English forerunner. American history did not begin in 1776.

Since the establishment of the republic, Hughes sees a pattern much like that described by Thompson for England before Adam Smith, but with the acceptance of the market and without a strict moral sanction against it. As in 17th and 18th century England, in times of abundance, the market is allowed free rein, but (p 8):

> ... When the market creates shortages of goods and services in response to the expressed desires of consumers, the consequences are resented and the call goes out for the government to intervene. And it has, over time, case by case, with the results we now face.

As a specific example, consider Hughes' description of the treatment of land from colonial times to the continental expansion (pp 59–60):

> ... we discern the general pattern. The colonial regime, with its township distribution, headright grants, quit rents and socage tenures, and long leases with feudal incidents (in New York), can be viewed as pursuing a policy of keeping land relatively scarce, while still encouraging its settlement ... Freed from these constraints, the new republic at first mixed tradition with the market and then relied increasingly upon the market. A plentiful land supply rendered a complex social apparatus of distribution superfluous, for straight sales were satisfactory to government and to buyers alike. But when the line of settlement reached beyond the 100th meridian into territories of low rainfall, good land became scarce again, and settlers demanded, and got in the Homestead Act, government rather than free-market land distribution.

Following the continental expansion, roughly in the period following the panic of 1873 and before the First World War, there was a continual and increased turning to ***federal*** rather than local nonmarket controls (p 95):

> ... Control gained the upper hand at the federal level over the new laissez faire because powerful groups rejected the decision of the free market regarding the ownership and distribution of economic power.

The powerful groups were many: farmers demanding protection against the economic devastation of crop failures, unions seeking ways of ameliorating worker risk in the "free wage bargain", businessmen trying to shield themselves from the effects of financial panics, and the general public wanting protection from the perceived abuses of monopoly power.

Subsequently, according to Hughes, America has gone beyond specific groups turning to the federal government for relief. By the end of the Second World War, it had come to expect the federal government to apply nonmarket controls to provide economic stability (p 148):

> Federal responsibility for the maintenance of a satisfactory level of economic life was a radical break with tradition. Some hoped for it in the thirties, but it was not until 1946 that a majority of Congress went along.

Hughes' work appeared in 1977, immediately before the great deregulatory movement in airlines, agriculture, financial markets, transportation, and telephones. It is not clear whether or not he would now make a categorical statement like (p 97):

> ... One only rarely sees in American history a demand to return to the free market once an apparatus of nonmarket control has been established.

But, as I argue below, if one imbeds Hughes theory in a framework of perceived economic morality, the deregulation movement is not necessarily a contradiction. All of what Hughes describes from a sweeping overall view of American economic history accords with what can be observed in public utility regulation. One of the principles that has emerged there, and it seems also in the economy generally, is the perceived injustice of special interest groups thwarting the conferring of widespread, diffused benefits. As we discuss in some detail in the sequel when we consider Proposition 5, it is in this form that economic efficiency appeals to public notions of fairness and justice.

At any rate, whether from 17th and 18th century England or from 19th and 20th century America, the historical evidence seems to indicate that in times of economic stress, economic justice looms large in governmental policy. Policies that are accepted unquestioningly in times of plenty are challenged as unfair, and either forgotten governmental mechanisms are resurrected or new ones are created to "correct" what are perceived to be economic injustices. Once in place, these mechanisms provide a framework for routinely examining the justice of economic actions that go unchallenged in a market setting.

4 Recent Literature on Descriptive Theories of Economic Justice

4.1 *Prescriptive Theory as Descriptive Theory*

In a sense, the distinction between descriptive and prescriptive theories of economic justice is artificial. To have predictive value, a descriptive theory must assume stable human behavior over time, which implies that human beings act according to some set of ethical principles – i.e., they act according to a prescriptive theory. Likewise, a prescriptive theory must be empirically based on some descriptive model of innate human behavior. So prescriptive or normative theories of economic justice, which go back to antiquity, obviously have descriptive empirical content. Unfortunately, prescriptive or normative notions of economic justice are often so broad and so intent on grinding a particular axe that they are not precise in stating their descriptive or positive bases. For example, Okun (1975, p 119) espouses the slogan "the market needs a place, and the market needs to be kept in its place." He champions a truncated income distribution as the primary economic justice goal – the lower boundary of family income should be at one-half the mean. Thus, he accepts equality, but also a measure of economic efficiency. On the other hand, Hayek (1944) finds repugnant all proposals for government interference in the name of economic justice and predicts that, unchecked, they will inexorably lead to totalitarianism. So, Hayek wants the economy to accord to the "new political economy", considered completely legitimate after Adam Smith, and to notions of individual liberty contained in the Bill of Rights. Needless to say, the Okun–Hayek tension is mirrored in the contrast between two recent books, ***The Zero-Sum Society*** (1980) by Lester Thurow and ***Free to Choose*** by Milton and Rose Friedman (1979), as well as by countless political debates between the "left" and "right".

In the philosophical literature, the recent works of Rawls (1971) and Nozick (1974) have received considerable attention from economists. Rawls suggests maximizing the lot of the worst-off members of society as an "end state" ethical goal, while Nozick rejects "end state" goals in favor of justice as a process. He argues that "a distribution is just if it arises from another (just) distribution by legitimate means". A literature is now emerging at the intersection of philosophy and economics (see Murphy and Coleman, 1984; Buchanan, 1985), which however is generally normative and has not turned its attention to the specific area of public utility regulation.

4.2 *Equity Theory in Social Psychology*

To my knowledge, the most intense empirical work on how economic justice is actually perceived has been done by social psychologists. Unfortunately, their work has been almost exclusively confined to two-person interactions and they have done little on the interaction of one person with a group or with the state. Nevertheless, according to Walster *et al.* (1979) over 400 empirical studies have been done on what social psychologists call "equity theory," and this literature offers some interesting insights. Walster *et al.* summarize what in their words is the "heart of equity theory" by four propositions. These are:

> *Proposition I:* Individuals will try to maximize their outcomes (where outcomes equal rewards minus costs).
>
> *Proposition IIA:* Groups can maximize collective reward by evolving accepted systems for equitably apportioning resources among members. Thus, groups will evolve such systems of equity, and will attempt to induce members to accept and adhere to these systems.
>
> *Proposition IIB:* Groups will generally reward members who treat others equitably, and generally punish (increase the costs for) members who treat others inequitably.
>
> *Proposition III:* When individuals find themselves participating in inequitable relationships, they will become distressed. The more inequitable the relationship, the more distress individuals will feel.
>
> *Proposition IV:* Individuals who discover they are in an inequitable relationship will attempt to eliminate their distress by restoring equity. The greater the inequity that exists, the more distress they will feel, and the harder they will try to restore equity.

Walster *et al.* further formulate the basis of what is considered an equitable relationship by a "definitional formula". The formula essentially says that the ratio of rewards must be equal to the ratio of contributions if two parties are to view an interaction as just. The definitional formula is traced to Aristotle in his *Nichomachean Ethics*.

The Aristotelian notion, ratio of rewards equal to ratio of contributions, is fraught with difficulties for an economist, even in simple two-party interactions. These difficulties are also apparent to social psychologists, as succinctly described by Homans (1976, p 232):

> Thus distributive justice always entails a comparison by the parties of the contributions each makes and the reward each receives, but what they compare is not the subjective value and cost of the rewards and contributions – for there can be no comparison of subjective values – but rather the outward

and visible amounts of the rewards and contributions. Thus workers in a factory compare their earnings but not how much these earnings 'mean' to each of them.

Note that if all parties are to accept a distribution as just, they must agree on three different points. First, they must agree on the rule of distributive justice itself: that rewards ought to be proportional to contributions. Second, they must agree on what kinds of rewards and contributions are to be legitimately taken into account in applying the rule. Third, they must agree in their assessment of the amounts of these contributions each makes and the amounts of these rewards each receives. Experience seems to indicate that people are much more likely to reach agreement on the rule itself than on the other two issues. They agree on the rule but not on its concrete applications.

4.3 *Experiments by Game Theorists and Economists*

Game theorists and economists have done a small number (relative to "equity theory") of experiments involving subjects agreeing on how to divide resources or money. Of particular interest are the experiments of Hoffman and Spitzer (1985); these explicitly address the effect of context on how subjects perceive what is fair. Their initial experiments gave startling (to them, and to economists generally) results involving bargains struck between

> ... two subjects who had opposing payoff functions and who had full information of one another's payoffs. ***By a flip of a coin*** one subject could choose a non-cooperative outcome, unilaterally: the winner of the coin toss could simply choose an outcome which gave him $12 and left the other subject nothing, regardless of whether or not the other subject agreed. However, if the two subjects cooperated they could obtain $14 from the experimenter, which could be split between the subjects in any ***mutually agreed on manner***. Cooperative game theory predicts that the subjects will cooperate and divide the rewards $13–$1 (the Nash bargaining solution: an even division of the $2 gain from trade). Under no circumstances should the winner of the flip settle for less than $12, according to game theory. In our experiments all of the subject pairs chose the joint profit-maximizing outcome, which the theory predicts, but they then divided the rewards $7,$7. In effect, each winner of a coin flip agreed to take $5 *less* than the $12 that he could have obtained ***without*** the other subject's cooperation.[6]

Although these results seem at first glance to be irrational to someone steeped in economics or game theory, Hoffman and Spitzer offer an explanation in terms of rational behavior with a moral basis:

> In the results reported above, we hypothesize that subjects behave in accord with a theory of distributive justice which says that flipping a coin is not a just way of allocating unequal property entitlements. Subjects perceived no morally justified difference between themselves, even though one "legally" owned a substantial property entitlement and the other did not. Because they were "morally equal", an equal split seemed to be the only fair allocation.

To test this theory, Hoffman and Spitzer varied the moral context in two ways: Winners did not win the flip of a coin but rather won a game of nim, and winners were instructed that they had "earned the right" to control the amount of money they and the loser were to receive (rather than having been "designated" to control). This treatment allowed Hoffman and Spitzer to test three theories of perceived distributional justice:

> ... observing how subjects' payoff distributions change, we can infer which theories of distributive justice subjects seemingly hold. ... Specifically, we consider the self-regarding, economic man paradigm from utilitarian theory, an extreme egalitarian theory of equal sharing, and the Lockean theory of earned desert.

Here, by "Lockean theory of earned desert", is meant a normative theory similar to the Aristotelian notion of "rewards proportional to contributions" of equity theory. Hoffman and Spitzer define it as follows:

> The Lockean theory posits that an individual deserves, as a matter of natural law, a property entitlement in resources that have been accumulated or developed through the individual's expenditure of effort. The individual deserves the entitlement because he "mixed his labor" with the resource.

With the addition of the "moral authority" and "earning" variables, the results came out strongly in favor of the Lockean theory:

> ... The results suggest that subjects behaved in accord with neither the self-regarding utilitarian theory nor the egalitarian theory of distributive justice, but rather in accord with the Lockean theory of earned desert. These experimental results show strongly that different methods of assigning property entitlements lead to significant differences in the frequency of self-regarding *vs.* equal payoff divisions. Subjects appeared to treat their entitlements as rights to unequal payoff divisions when the experimental institutions led those subjects to believe that they had "earned" those entitlements.

Similar results are reported in Selten (1978) who reviews a number of experiments testing "individual rationality". He concludes that many of the results may be summarized by a ***general equity principle***, namely, if the subjects perceive themselves to have contributed to the achievement of outcomes that yield resources or rewards, the Aristotelian "rewards proportional to contributions" formula will be followed; where contributions cannot be distinguished, rewards will be divided equally.

The results from "equity theory" of social psychology and from the experiments of game theorists and economists are suggestive of a strong "framing" or contextual effect on what the public perceives as economically fair. In normal times, driving a "gas guzzler" automobile evokes no public reaction. But in times of a gasoline crisis, such an activity may suddenly be considered very unfair, to be stopped immediately by the authorities. Thompson's observation of a different ethical norm existing during an era of officially "just prices" from an era of a freely acting market economy supports a contextual effect, as does Hughes' thesis that severe economic setbacks usually result in pleas for governmental intervention. This would predict that what is perceived to be economically just behavior in a regulatory context might differ markedly from economically just behavior in a market context, especially since regulation has a structure in place for the hearing of economic grievances. Unfortunately, to my knowledge, no specific experiments to test this hypothesis have been performed.

4.4 *Economic Theories of Regulation*

Several "economic" or "political" theories of regulation have recently been proposed, which Levine (1981) calls "revisionist" theories. These theories try to revise the older "public interest" theory – that regulation does and should exist to remedy market failures – in favor of a theory that regulation exists for the benefit of rational agents acting to further self-interest. All of the revisionist theories are necessarily based on some behavioral model of economic agents, and thus directly or indirectly on some theory of how individuals deal with economic justice. By and large, the underlying behavioral assumptions in these models are crude and are presented as being self-evident and self-validating, not requiring detailed explication or empirical support.

Some of the principal economic or political regulatory models are the implicit contract model of Goldberg (1976), the facilitator model of Fiorina and Noll (1978), the political model of Wilson (1980), the economic theory of regulation of Stigler (1971) and of Peltzman (1976), and the strategic game models of McDonald (1975) and of Owen and Braeutigam (1978). (For a review and critique of the economic theories of regulation, see Posner, 1974). I will not attempt to discuss all of these theories, but will comment on Stigler–Peltzman (SP) and Owen–Braeutigam (OB) to illustrate my general dissatisfaction with them as a class.

In brief, Stigler starts with the observation that the state has a unique power, namely, the power to coerce or police. Self-interest groups in society will, according to Stigler, try to commandeer this unique power to further their own ends. For example, members of a profession will demand licensing laws that erect high barriers for those who would enter the profession – thereby giving existing members monopoly power. The grab for the state's police or coercive power creates the demand side of a market for regulation. On the other hand, society gives legislators or regulators the right to exercise the government's police powers. But legislators or regulators have their own self-interest goals: power, glory, re-election, high-paying jobs when ,their term ends, etc. They disburse governmental power in return for things, such as votes, which will further these goals, thereby creating the supply side of a market for regulation. In the Peltzman elaboration of Stigler's basic model, the essential commodity being transacted in the political market is a transfer of wealth. The regulator seeks to maximize the net votes or a majority in his favor, by benefiting a group through taxing those not in the benefited group. The probability of the regulator's obtaining support from the benefited is assumed to increase with the group's received net benefit (benefit minus costs of campaigning, lobbying, organizing into a group, etc.). On the other hand, the probability of opposition by the nonbenefited is assumed to increase with the tax they suffer and to decrease with the per capita expenditures used on education to quell their opposition. Peltzman puts these ingredients together into a mathematical model that can be solved for equilibria and examined for insight into the relevant importance of the various factors.

Owen and Braeutigam take as their starting point the bold and seemingly weakly supported assertion that the administrative process has two main effects: it confers beneficial property rights in the status quo and has the government act as insurer against economic uncertainty. For example, Davis' (1958) standard text on administrative law states (p 7):

> The pervasiveness of the effects of the administrative process on the average person can quickly be appreciated by running over a few samples of what the administrative process protects against: excessive prices of electricity, gas, telephone, and other utility services; ... uncompensated injuries related to employment; cessation of income during temporary unemployment; subminimum wages; poverty in old age; ... loss of bank deposits.

A true believer in the OB theory might charge, for example, that, in the name of protecting the consumer, the administrative process will bias utility and transportation rates against increases, that is, the process will confer property rights in the status quo on those whom the status quo benefits. Likewise, in the name of "protection against chaotic conditions in broadcasting" the process will give those who like classical music a property right in the status quo of its broadcast; a classical music radio station will be required to go through elaborate procedures to change to country rock music. Likewise, a believer in OB might conclude that "protection" against "uncompensated injuries related to employment", "cessation of income during temporary unemployment", "poverty and old age", and "loss of bank deposits" are examples of the government acting as an insurer when private insurance could possibly be superior.

In the OB theory legislators and regulators are assumed to be seeking the favor of the "median voter", much as in the SP theory. Regulatees, however, are assumed to be capitalizing on the attempt of the administrative process to treat them fairly and equitably. In particular, regulatees have a powerful tool – the delay of regulatory action by demanding the due process of regulation to run its course. The net result, according to OB, is the opportunity for regulatees to "game" the system, acting strategically in the light of regulatory and administrative rules.

The SP and OB theories are both appealing, as are revisionist theories of regulation already mentioned. But they all lack a firm behavioral foundation, which must be grounded in notions of economic justice that are considered legitimate. As already mentioned, none of the theories attempt to explain how "self-interest groups" form in the first place. Likewise, even if we understand the economic forces driving individuals to coalesce into self-interest groups, how do we explain when legislators can afford to trade the government's policy powers for votes without incurring the wrath of the nonbenefited voters? What is the boundary between logrolling, porkbarreling, and other perjoratively described legislative acts and legislators' allowing members of a profession to get away with the usurpation of monopoly power?

In the OB theory, there is a different deficiency: In explaining the driving forces of regulation by appeals to elemental notions of economic justice OB have not cast their net widely enough. The feeling that one has a property right in the status quo is indeed a powerful force. Likewise, "the government should be my insurer" is also a strongly held ethical belief. But, as I argue below, these are not the only two effects at work.

5 Six Propositions of Perceived Economic Injustice

5.1 *Introduction*

In this section we describe the six principal economic justice effects that seem to circumscribe policymakers in public utility regulation. At the outset, three remarks are in order.

First, it may be more useful for policy analysis to focus on economic injustice rather than on justice. This point is cogently made by E.N. Cahn in *The Sense of Injustice* (1949). Cahn first discusses attempts to ground justice in "natural laws" motivated by arguments like the following (p 4): "There are natural laws to govern the relation of stars and planets on their journeys through the heavenly spheres, and, by like token, there must be natural laws to govern the relations of human beings in their social and political orders." Cahn then describes Hobbes' argument that natural laws are arbitrary, and goes on to observe that attempts to axiomatize natural laws have proved sterile, and, further, that basing justice on experience or particular instances leads to infinite regress ["how are we to know that the instances selected are themselves wholly 'just'?" (p 13)]. Cahn goes on to say (p 13):

> Why do we speak of the 'sense of injustice' rather than the 'sense of justice'? Because 'justice' has been so beclouded by natural-law writings that it almost inevitably brings to mind some ideal relation or static condition or set of preceptual standards, while we are concerned, on the contrary, with what is active, vital, and experiential in the reactions of human beings. Where justice is thought of in the customary manner as an ideal mode or condition, the human response will be merely contemplative, and contemplation bakes no loaves. But the response to a real or imagined instance of injustice is something quite different; it is alive with movement and warmth in the human organism. For this reason, the name 'sense of injustice' seems much to be preferred.

Cahn summarizes his own candidates for perceived economic injustice as follows (p 22):

> The sense of injustice may be described as a general phenomenon operative in the law. Among its facets are the demands for equality, desert, human dignity, conscientious adjudication, confinement of government to its proper functions, and fulfillment of common expectations. These are facets, not categories. They tend to overlap one another and do not together exhaust the sense of injustice.

The point raised by Cahn is not trivial. If we think in terms of a mathematical formulation, the quest for an economically just state suggests axioms that evoke a social preference function to be optimized under feasibility constraints that reflect resource limitations, such as Yaari's (1981) mathematization of Rawls, or axiomatic approaches to "just" cost allocation that lead to the Shapley value (see Chapters 1 and 3).

On the other hand, the quest for a lack of economic injustice suggests inequality constraints on possible resource allocations and perhaps a region in the space of resource allocations that is considered free of economic injustice, a view reminiscent of the core and related solution concepts in cooperative game theory (see, for example, Zajac, 1978; also Chapters 1, and 8).

Second, as in the case of the US Constitution's articles and amendments, we can expect conflicts between the six economic injustice Propositions set forth below. The freedom of the press guaranteed by the US Constitution's First Amendment suggests that a reporter's sources should be protected; while the right of the accused to confront his accusers suggests that the sources should be divulged. Every Supreme Court term is devoted to resolving conflicts over which of two or more possible Constitutional guarantees should prevail. Likewise, as will become apparent from the six Propositions, in certain instances, more than one concept of economic justice may be applicable and different groups may be winners or losers depending on which concept society thinks is paramount. The resolution of the resulting conflict may be difficult and controversial, requiring lengthy legislative, judicial, or regulatory proceedings.[7]

In other terms, policymakers in the public utility regulation collectively act, in effect, as "ethical pluralists" as described by Gordon (1980, p 46):

> The basic problem of ethics arises from the fact that criteria of goodness may conflict with one another. The ethical pluralist accepts this and does not attempt to eliminate such conflicts by methods such as utopian reconstruction, reduction, or lexicographical ordering. The main tasks of pluralist ethics are to clarify ethical criteria, identify conflicts of values, and improve our ways of mediating them.

Third, my conjecture is that many, perhaps all, of the Propositions can be formulated as models of rational economic agents acting to achieve self-interest rather than as moral maxims. For example, why do queues form at a bus stop? The moral answer might be that it is only fair that those who arrive at the stop first get the first chance to board a crowded bus. A rational-actor-model explanation might be that order is a public good; those who ride buses have learned the benefits of this public good and hence form queues in their individual self-interest. For parsimony and to make the discussion accessible to noneconomists, I have not attempted to construct such rational-actor models for any of the Propositions. A deeper issue is whether or not a theory of positive economic justice can be constructed with no reference to morality. In the philosophical legal literature, such an extreme view of "legal positivism" has a long history but is now generally rejected (see Murphy and Coleman, 1984, Chapter 1). An open question is whether reasonable explanatory and predictive power would be provided by a similar extreme "economic justice positivism" in terms of rational actors seeking Pareto superior moves.

5.2 *The Six Propositions*

5.2.1 Economic rights

In the Reagan administration, it is currently fashionable to talk of "safety nets". This is a recent manifestation of the notion that every individual somehow has a right, or an entitlement to a minimum level of economic "necessities" such as food, shelter, clothing, a job, health care, education – and in the United States, utilities. The very fact that a conservative administration feels it must espouse an economic rights concept such as a safety net is significant, for such espousal has not always been felt necessary and, for that matter, still has its critics.

Indeed, the movement toward economic rights has been so strong in the United States and Western Europe as to generate support for economic rights to be given constitutional status. Perhaps the most dramatic example of this has been the attempt to get adopted a United Nations Convenant on economic rights.[8]

Recently in New York State, the notion of economic rights has been extended to cover utility services. In 1981, the New York State Legislature and the governor approved a "Utility Bill of Rights".[9] The *New York Times* (July 4, 1981) reported:

> With the bill's sponsors asserting that gas and electric service 'are not a luxury, but a necessity', the Senate tonight approved a measure defining the rights of consumers in their dealings with utility companies ... Under the bill, consumers who are approved for utility service must receive that service within five days – instead of the current 10 – and cannot be required to pay security deposits. Companies must provide service to welfare recipients, even if they cannot pay, and the State Department of Social Services must cover up to four months of arrears and guarantee the payment for future services to such recipients. The payments would be guaranteed for no more than two years. The so-called 'Utility Bill of Rights' requires utility companies to follow specific procedures when they want to cut off gas and electricity. The companies are also prohibited from turning off heat during cold weather without first contacting an adult resident of the household at least 72 hours before shutoff. A utility cannot terminate service if a doctor has certified that a person in the household is experiencing a medical emergency. The bill permits customers who are behind in bill payments to pay the debt on a deferred schedule keyed to their ability to pay. The legislation also limits a utility's use of estimated bills. The Companies are also required to inform the customers annually, in writing, of their rights.

Economists might be taken aback by New York State's "Utility Bill of Rights". For example, Leland Johnson (1981), in commenting on the author's paper, "Is Telephone Service an Economic Right" (1981) states, "Taken literally this (an economic right) suggests the jolting notion that allocation of resources in meeting basic needs (however they are defined) should proceed independently of market signals." Johnson further warns:

> We can agree that, in some sense, people should have an economic right to telephone service. But the concept of economic right is extraordinarily fuzzy and must be examined most carefully before making policy recommendations in specific situations. If we grant that a particular service should be provided as an economic right, do we mean that the service should be provided free of charge to all comers, or that the user should pay enough to cover at least some portion of that cost? Or, once we have agreed upon some price that is 'proper' for service to which a user has an economic right, does this mean that any increase in that price is to be prohibited in the future, despite inflation, changing market structure, or changing social needs? ... I, for one, am disturbed by any notion of economic rights that casts in concrete a structure within which the user pays nothing, or a specified amount for a specific service, regardless of changing circumstances.

Johnson's warnings are right on target. But his view, and the views of the majority of economists, are not governing. As I said in my paper:

> Whether one believes telephone service should be an economic right is beside the point. Individuals do not make the decision; rather, society acting through the democratic process decides it. In my view, a disinterested, objective observer of the American scene would conclude that a basic level of telephone service is regarded as an economic right.

The evidence leads me to the first and perhaps most important perceived economic injustice proposition:

> *Proposition 1:* It is now accepted that every individual has basic economic rights to adequate food, shelter, heat, clothing, health care, education, and, in the United States, to basic utility services. Deprivation of basic economic rights is considered unjust.

5.2.2 Equality of gain and pain

We live in a society of widely disparate incomes and individual wealth. These disparities and others in possessions, status, health, etc., are accepted as individuals go about their daily business. But envy, i.e., coveting what the other fellow has, is a deeply felt emotion. When circumstances precipitate feelings of envy, the reactions are liable to be swift and violent, with strong feelings expressed over the outrage of unequal treatment.

Musgrave's *The Theory of Public Finance* (1959, Chapter 5) traces economists' attempts, starting in the middle of the sixteenth century with an essay by Guicciardini, to formulate workable schemes for minimizing envy and treating individuals equally. The efforts of subsequent social philosophers culminated in a classic essay by John Stuart Mill arguing in favor of "equal sacrifice" as the proper basis of taxation. After pointing out the difficulty of making the concept of "equal sacrifice" precise (should it be equal, absolute, proportional, or marginal sacrifice?), Musgrave states his position with regard to interpersonal utility comparisons as follows (pp 108–109):

> This entire discussion rests on the assumption that interpersonal utility comparisons can be made in a meaningful fashion. This assumption is basic to a subjective view of the ability-to-pay doctrine. Yet it is an assumption generally rejected by the "new" welfare economics. If such rejection is valid, the entire concept of equal sacrifice becomes so much nonsense and must be discarded – lock, stock, and barrel ... I hesitate to go this far. While we cannot assume that the utility schedules of individuals are known, the new welfare economics may have gone too far in its categorical rejection of interpersonal utility comparisons. Such comparisons are made continuously, and in this sense have operational meaning ... For purposes of policy formation in a democracy – or, for that matter, in a nondiscriminating dictatorship – the best solution may well be to follow Robbins' formula and to proceed as if individuals were alike. Thereby, the concept of subjective utility is translated into one of social income utility ... If we proceed along these lines, the principle of ability to pay ceases to be the subjective matter that J.S. Mill had thought it to be. It becomes a question of social value, and the problem is how the values can be determined. In a democracy they must be traced to the preferences of individuals, and a political mechanism must be designed by which this can be accomplished.

Thus, Musgrave rests his faith on the possibility of a utilitarian approach, based on a "political mechanism" that allows the policy maker to infer "social value" from individual preferences. Musgrave's language immediately brings Arrow and the problem of collective choice to mind. Since Musgrave was writing in 1959, only eight years after Arrow's *Social Choice and Individual Values* (1951), the difficulties of collective choice were not as well appreciated as they are today, and did not seem to be overwhelming obstacles to Musgrave:

> Arrow's particular set of conditions for social ordering are based on norms that are met where majority voting leads to an unambiguous result. This is an interesting case because majority voting is widely used; but it is only one possible case, without claim to general validity ... It remains to be seen whether the results obtained under a system of qualified majority, plurality, or point voting might not be superior.

I think the modern view among economists is that the interpersonal-comparison-of-utilities problem cannot so easily be dismissed. Rather, the economist must approach questions like equal tax treatment with great humility. Thus, in motivating their theoretical work on effects of taxes, based on a Bergsonian welfare function, Atkinson and Stiglitz in their recent *Lectures on Public Economics* (1980, pp 422–423) state:

> In deciding on the direction of such research, one must bear firmly in mind the purpose of this kind of literature. The aim is not to provide a definite numerical answer to the question, 'how progressive should the income tax be?' ... The purpose is rather to explore the implications of different beliefs about how the world works or about how governments should behave...

Still another possible theoretical approach to the issue of equality is to observe that inequality is a fact of life that is tolerated and to argue that the search for equality is at heart a search for lack of envy. This then suggests the literature on "envy-free", technically called "equitable", resource allocations, where envy has Foley's definition: "I prefer your bundle of goods to mine." This literature is still in the making (see Varian, 1975, for an introduction) and, as yet, has little to offer in the way of practical policy prescriptions.

As far as the public is concerned, my guess is that it finds economists' worries about avoiding interpersonal utility comparisons, the difficulties of establishing universal measurements of gain and pain, and the subtle but surprisingly disastrous implication of the voter's paradox in collective choice all to be the most arcane academic hairsplitting. This is especially the case where an apparently valid measuring rod appears at hand – for example, gains and losses measured in dollars or in time. The social psychology equity theory literature seems to teach this, and regulation is in fact replete with applications of the principle of equal treatment. The regulation of natural monopoly is based on the equal treatment notion that investors in a natural monopoly should get a return equal to that of other investors who bear comparable risk. Those who would price discriminate in public utility pricing bear the burden of proof that price differentials are based on cost differentials; appeals to economic efficiency as a basis of price discrimination are difficult to sustain.

This brings us to:

Proposition 2: Equal treatment of individuals is seen as a just basis for policy, especially when common measurements, such as dollars or time, of individual gain or sacrifice are at hand. Unequal treatment of individuals is considered unjust.

5.2.3 *Right in the status quo*

> In 1978, my wife and I together with another couple made a one week tourist trip to Cuba. On arrival in Cuba, we were escorted to two buses, each of which had a guide who was to accompany us for the entire week of an intensely planned trip. The buses were filled in order as we came out of customs; those coming out first filled the first bus and the overflow the second bus. The four of us happened to be in the first bus, which also happened to have a very vivacious and interesting young lady as the guide. As a girl, she had lived for several years in Miami and spoke fluent English. Although she worked hard at propagandizing us, she readily answered all questions and offered even more information and interesting sidelights than was requested. The young lady guide in the second bus can only be described as a dolt. She spoke English very haltingly, was not receptive to questions, and offered as little information as she could get away with. The disparity between the guides was apparent within the first day, but no one attempted to change buses until the third day. On that morning, my wife and I and our friends were the last to get on the first bus. We discovered that all of the seats were taken. Four persons from the second bus had filled them. What to do? Tempers quickly came to a boil, not only ours but those of friends we had already made on the first bus. The guide finally announced, "I know who you people are. This bus is not leaving until you go back to ***your*** bus." Two people finally sheepishly departed to ***their*** (the second) bus. Silence! The guide then implored two single people to go to the second bus as a special favor to her. They did, and the trip resumed. The two interlopers were of course recognized and completely ostracized for the rest of the trip, not only by the occupants of ***our*** bus but by the occupants of ***theirs*** as well.

Students seat themselves at random the first day of class, and continue to occupy the same seats; after a few class meetings, each student feels he owns his or her seat. Queues are normally formed voluntarily in front of ticket windows, bank tellers, to enter a bus, or generally when many are served by a few. A place in the queue is felt to be owned by its occupier, a fact that sometimes led to bloodshed in the 1974 and 1979 gasoline crises in the United States.

The sense of ownership in the status quo is a commonplace phenomenon and its justness may be felt so strongly as to outweigh feelings that equality should prevail, as the above anecdote illustrates. I have already remarked that this phenomenon is a behavioral cornerstone of Owen and Braeutigam's theory of regulation as a game. It is also an element in Goldberg's contract theory of regulation, where the regulator is viewed as the customers' agent with the job of negotiating an implicit contract with the regulated firm that supplies them; part of the implicit contract is that a certain quality of service will continue to be provided at rates that are determined by the regulatory process.

Aside from Owen and Braeutigam and from Goldberg, the principal researcher who has explicitly focused on this phenomenon seems to be W.J. Samuels (Samuels, 1974; Samuels and Mercuro, 1976). Several recent events strikingly illustrate the phenomenon:

(1) Long-standing rent control establishes a notion that status-quo rents should be raised only to cover "fairly" incurred costs, e.g., due to inflation, and not merely to respond to market forces. Attempts at the official raising of rents are resisted by tenants, sometimes violently. For example, in 1981 the New York City Rent Guidelines Board voted increases for lease renewals for the 900 000 apartments it controls. The increases were 10%, 13%, and 16% for one, two, and three year leases respectively. The *New York Times* article (of June 16, 1981) remarked that the board's meeting was "raucous", resulting in twelve demonstrators being arrested. It went on to observe:

> The persistent din of yesterday's meeting seemed to surprise even longtime observers of the normally noisy board meetings. The tenants, who set up a picket line outside the building before the meeting convened, marched into the auditorium chanting and clapping and refused to stop. At one point, the meeting was adjourned for 40 minutes because the demonstrators drowned out the deliberations ... More than 25 police officers were on hand at the peak of the demonstrations.

Subsequently, on June 29, 1981, the *Times* reported that New York legislature leaders had decided to extend for another two years the seven-year-old *Emergency Tenant Protection Act* which had established the board.

(2) As western US cities develop, successive rings of new developments generate successively increasing water costs. Residents and businesses often feel it is only fair that they should continue to pay the rates that prevailed before more recent development on the grounds that the new developments caused the higher water costs. The classic economic efficiency counterargument is put forth by J.T. Wenders (1981):

> Both old and new customers are equally responsible for the capacity and operating costs of the LVVWD (Las Vegas Valley Water District). An old customer, by occupying facilities which could otherwise be transferred to new customers and save the LVVWD the cost of building new facilities at current cost, is just as responsible for the growth in the size of the LVVWD delivery system as is the new customer ... It is clear that the LVVWD has chronically mis-priced water. By failing to recognize that summer, peak-period usage is much more costly to provide than is winter, off-peak usage, the District has severely underpriced summer water usage. This has caused the uneconomic stimulation of summer water usage in Las Vegas, resulting in an uneconomic overexpansion of the LVVWD delivery system. Thus, the LVVWD rate structure has encouraged more investment in water delivery facilities than would be needed had the District adopted a rate structure which more accurately tracked costs.

The issue is even more forcefully joined in a 1976 *Public Utilities Fortnightly* article by A.E. Kahn and C.A. Zielinski, both former chairmen of the New York Public Service Commission, as quoted in Wenders (1981):

> We may expect therefore to be confronted constantly with proposals for vintaged rates – lower rates for existing subscribers and higher rates for new.

> Since most commissions have probably been presented with similar reasoning in the case of electricity – namely, that it is the new electric customers, or those whose demands are growing, who are responsible for our having to incur the higher current costs, and that rates approximating incremental costs should therefore be applied discriminatorily to them – it seems to us very important to get the basic principles straight. So far as generating and transmission costs, at least, are concerned, new customers coming on the line are no more responsible causally for the incurrence of higher current costs than old customers staying on the line. In purely economic terms, it is just as important for the latter as for the former to confront a price that reflects the additional cost to society of their taking more, or the amount that society would save if they took less. And while we recognize the intuitive attraction of protecting old customers from the burden of inflation, we cannot understand the case even on moral grounds for protecting wealthy consumers whose level of consumption, already extremely high, is not growing, at the expense of poor consumers whose consumption is growing, possibly because they are just getting married, or just having children. Vintaging of this kind is flatly, and in our judgment inexcusably, discriminatory.

Thus, the feeling of a property right in the status quo, if the status quo is preferred to a proposed change, can be intense, as illustrated by the rent control and old-versus-new water-users examples. If a Rockefeller can acquire a property right in economic wealth by the accident of birth, why shouldn't I have a property right to a seat on the first Cuban bus by the accident of having gotten through customs before others did, or why shouldn't the early settlers of Tucson, Arizona have a property right to cheap water by the accident of having moved to Tucson sooner than others? This idea may be summarized as

> *Proposition 3:* The beneficial retention of a status quo is considered a right whose removal is considered unjust.

This Proposition on the right to a beneficial status quo is perhaps the most interesting of the six and the most fruitful for further study. It is intimately related to the notion, extensively discussed in the literature, of property or ownership as a bundle of rights (see Gordon, 1980, pp 84–89 for an introduction). One of the forces obviously behind this Proposition is the public's appreciation that *stability* or order is a public good of enormous importance to society. This is discussed at length in Rawls (1971). Usher (1981) formulates a game theory approach to stability, arguing that the game-theoretic instability of majority rule means that legislatures in a democratic society will avoid explicitly dealing with income or wealth redistribution *per se*, the implication being that society recognizes that stability is too important to be put in jeopardy. Since wealth is a bundle of claims to property rights, the argument extends to an argument about the stability rights to the status quo. Much more can be said in this vein but would lengthen an already long chapter.

5.2.4 Society as an insurer

It is impossible to protect oneself against all contingencies. No matter what level of disaster or misfortune we provide for, something worse is possible. So, it is common in almost all societies to provide shelter and food when natural disasters such as earthquakes, floods, or volcanic eruptions destroy homes, or medical care when an unexpected grave illness strikes. Likewise, institutions that in economic disasters shift the responsibility for economic burdens from the individual to someone else are common. The idea that such a shift is only fair is, of course, behind the notion of economic rights. But it is also behind the notion of society protecting the individual against the fickleness of the marketplace. Again, it is easy to assume that modern institutional arrangements have always existed and to forget that they have evolved over time, with a concomitant evolution in ethical norms. Significantly, perhaps the two most important modern forms of transference of economic responsibility — the limited liability corporation and bankruptcy — are of relatively recent origin.

In a society with a strong ethical belief that each individual should be responsible for his or her debts, the idea that several persons should be allowed to limit the responsibility for their collective debts may seem like a moral outrage, a scheme to encourage persons to act irresponsibly. This is in fact Adam Smith's view in *The Wealth of Nations* (1776, p 264, vol 2):

> This total exemption from trouble and from risk, beyond a limited sum, encourages many people to become adventurers in joint stock companies, who would upon no account, hazard their fortunes in any private copartnery... The directors of such companies, however, being the managers rather of other people's money than of their own, it cannot well be expected, that they should watch over it with the same anxious vigilance with which the partners in a private copartnery frequently watch over their own ... Negligence and profusion, therefore, must always prevail, more or less, in the management of the affairs of such a company. It is upon this account that joint stock companies for foreign trade have seldom been able to maintain the competition against private adventurers. They have, accordingly, very seldom succeeded without an exclusive privilege; and frequently have not succeeded with one. Without an exclusive privilege they have commonly mismanaged the trade. With an exclusive privilege they have both mismanaged and confined it.

Smith follows the above passages with a recitation of some of the spectacular mismanagements and failures involving the Royal African, Hudson's Bay, South Sea, and East India Companies, and with Abbé Morellet's list of fifty-five failed joint stock companies. He sums up (p 280, vol 2):

> To exempt a particular set of dealers from some of the general laws which take place with regard to all their neighbors, merely because they might be capable of thriving if they had such an exemption, would certainly not be reasonable. To render such an establishment perfectly reasonable, with the circumstance of being reducible to strict rule and method, two other circumstances ought to concur. First, it ought to appear with the clearest evidence, that the undertaking is of greater and more general utility than the greater part of common trades; and secondly, that it requires a greater capital than can easily be collected into a private copartnery.

Smith concludes that only in banking, insurance, canals, and water works are his conditions fulfilled and only in these industries should joint stock companies be allowed.

Smith's ethical outlook was evidently shared by others. Although limited liability, joint stock companies were first chartered in England in 1662 (see Hill, 1967, p 186), E.S. Mason (1968) points out that they had a checkered history, with reactions to curb them setting in after spectacular failures. For example, the collapse of the South Sea bubble was followed by the Bubble Act of 1720, which was "An Act to Restrain the Extravagant and Unwarrantable Practice of Raising Money by Voluntary Subscriptions for Carrying on Projects Dangerous to the Trade and Subjects of the Kingdom." The act prohibited (quoted in Mason):

> The acting or presuming to act as a corporate Body or Bodies, the raising or pretending to raise transferable Stock or Stocks, the transferring or pretending to transfer or assign any Share or Shares in Such Stock or Stocks without Legal Authority either by Act of Parliament or by any charter from the Crown to warrant such acting as a Body Corporate...

Mason then goes on to remark that, in England, only in 1855 was limited liability granted as a matter of course to companies wishing to incorporate.

Debtor's prisons were still common in Dickens' day, and bankruptcy has a similar, parallel history to corporate limited liability, with the escaping of debts being legitimated only recently. According to Clark (1939):

> The original purpose of bankruptcy laws was to secure full realization of assets and just distribution among creditors ... Not until 1705 was provision made for the release of the debtor from the undischarged balance of his debts, and not until 1826 in England (1841 in the United States)[10] could the debtor voluntarily go into bankruptcy and so avail himself of this privilege of discharge.

Thus, we seem to have evolved, with many ethical oscillations, to policies of ever greater protection of individuals and groups of individuals from the exigencies of the market. In recent times, we have seen the "bail-outs" of Lockheed and Chrysler, and the fierce debates about the extent of protection that should be extended to the steel and auto industries against the onslaught of foreign competition. It is perhaps not surprising that individuals in society should, as a matter of course, expect to be granted the same kind of governmental protection as is given to the large corporation or even to the profligate neighbor who has declared personal bankruptcy. Thus, the economist may view such things as fair trade laws as a cynical grab for monopoly power on the part of small businessmen, but small businessmen may view them as justifiable insurance against "destructive competition" by large businesses. Owen and Braeutigam state this viewpoint well (23–24):

> Market forces, particularly those associated with innovative activity, necessarily pose a threat to human beings with less than instantaneous adaptive capabilities. It is not merely that investments in physical capital with few alternative uses may be threatened, but also investments in human capital: specialized skills, knowledge of an industry or firm, and the like. Political activities designed to protect these investments from sudden unexpected reductions in value may be indistinguishable from actions designed to achieve an increased return on investment through monopoly. But to the extent legislators see the proposals for regulation as being principally an attempt to obtain the benefits of due process, with its slow deliberation, in order to protect human investment, they may reasonably be sympathetic. This point of view, when coupled with the symbolic political usefulness of the notion that regulation is to protect the consumer from monopoly prices or unsafe practitioners and products, may be quite persuasive. In a sense, then, regulation is not much different from unemployment insurance and agricultural price

> supports, both of which are intended to protect human as well as financial interests from the shocks and blows of market forces.

Thus, we see federal policies to "stabilize the market" in agriculture, and uniform prices across airlines (until their recent deregulation). This discussion may be summarized by:

> *Proposition 4:* Society is expected to insure individuals against economic loss because of economic changes. Failure to insure is considered unfair.

5.2.5 *Economic efficiency*

The desire to trade or transact to mutual advantage seems to be a fundamental human trait. Suppressing this desire by government authority is apt to meet with as much success as attempts to suppress love, sex, or original sin. This is brought out by Barbara Tuchman in *A Distant Mirror*. After describing the medieval doctrine of the "just price" she goes on to explain that it was to a large extent honored in the breach (Tuchman, 1978, p 37–38):

> ... banker, merchant, and businessman lived in daily commission of sin and daily contradiction of the moral code centering upon the 'just price'. ... To ensure that no one gained an advantage over anyone else, commercial law prohibited innovation in tools or techniques, underselling below a fixed price, working late by artificial light, employing extra apprentices or wife and under-age children, and advertising of wares or praising them to the detriment of others. As restraint of initiative, this was the direct opposite of capitalist enterprise. It was the denial of economic man, and consequently even more routinely violated than the denial of sensual man ... Merchants regularly paid fines for breaking every law that concerned their business, and went on as before...

The point is also made by opponents of command or planned economics, who argue that attempts to control prices, to ration, and otherwise to monitor the flow of resources by a central authority inexorably lead to black markets or to other forms of noncompliance with rules and regulations. We had a black market in the United States during World War II and we now have the claim that a large underground economy has arisen in direct response to excessive regulation. In 18th century England, even though the statute books were full of laws governing the setting of just prices, the laws were ignored except in times of famine.

In the US, we have seen the movement in recent years toward deregulation. The airlines present a striking example of deregulation in the name of facilitating economic efficiency. Levine (1981) recounts its history in arguing that it resulted primarily from Congress's desire to serve the public interest and to correct perceived economic injustice. According to Levine (p 191):

> In 1938, it appeared to the general public and the Congress that uncontrolled markets did not work very well for the public over the long run. Although markets produced low prices during the Depression years, many producers went out of business. The airline business was relatively new, Congress had little experience with it, and there was no reason not to apply general skepti-

> cism about markets to airlines ... it seemed to Congress that mistakenly optimistic entrepreneurs were seeking profitable operations where none were possible, draining away resources needed for further extension of the airline system in fruitless and profitless competitive struggles for existing business ... It was therefore not difficult for the airline industry to persuade Congress and the public that the fledgling airline industry would go the way of many other Depression-era firms and that the full potential of aviation could not be developed in a free-market environment. This would be both inefficient and unfair.

Panzar (1980) expresses a similar opinion that an economist viewing the demand and supply characteristics of the fledgling airline industry in the mid-thirties would have concluded that regulation was the policy that best served the public interest. However, later experiments like the California and Texas unregulated air service cast doubt on how well the system worked. As Levine goes on to remark (p 193):

> Over time, scholarly assessments of the performance of the regulated industry suggested that airlines operated better without regulation. Congressional leaders thought it politically beneficial, and perhaps even consistent with their legislative duty, to make changes which might benefit the public.

Finally, in spite of having a system that worked well in many respects, and in spite of strong opposition, deregulation was accomplished (pp 193–194):

> Certainly the industry opposed deregulation, as did many members of Congress and the public ...
> But the telling fact for our purposes is that this initially doubtful congressional faction (including Senator Cannon, the powerful chairman of both the Aviation Subcommittee and the full Commerce Committee) ultimately helped control the legislative process in favor of deregulation, which by then was perceived to be in the public interest. And it did so in the face of diehard opposition by factions (including the industry) whose positions were undermined by the ultimate transparency of the degree to which their positions were motivated by purely private, rather than 'public-interest' considerations of gain and loss ...
> This scenario may seem painfully quaint to most readers. It is certainly not self-evidently correct. But it is consistent with the rhetoric of the legislative and administrative history, with the facts of airline deregulation, and with the existence of a broader movement to deregulate other industries as well.

Offsetting the current deregulatory movement are the many years of motion toward more regulation rather than less that are described by Hughes (1977). Evidently the examples of waste and inefficiency because of the impeded action of markets must be numerous and flagrant in order to evoke popular sentiment for unfettering market action. Hence, we postulate:

Proposition 5: The existence of numerous and significant economic inefficiencies is considered unjust, especially if their existence is seen as conferring benefits on special interest groups who oppose their removal.

5.2.6 Resentment of abuse of monopoly power

Since 1877, the United States Supreme Court has struggled with the problem of defining the boundaries of regulation. To a large extent, the struggle has been a conflict between individual and corporate rights to control property and the police powers of individual states. The battles have centered on the Fifth Amendment to the Constitution ("No person ... shall be deprived of ... property without due process of law"), the Tenth ("The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people"), and the Fourteenth ("No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United Sates; nor shall any State deprive any person of ... property without due process of law ...").

A typical history of a dispute that has found its way to the Court has been the following: The citizens of a state become sufficiently outraged over some perceived abuse of economic power to cause the state legislature to pass a law curbing the power; those curbed sue on the grounds that their constitutional rights have been abridged.

To summarize this history briefly, I follow Phillips (1969, Chapter 3) (see also Hughes, 1977). The first, landmark decision was *Munn v. Illinois* in 1877. At the time, virtually all of midwestern grain flowed through Chicago, the "gateway of commerce". Nine firms owned all of Chicago's grain elevators and met periodically to fix storage rates. In 1871, the Illinois legislature passed a law fixing the rates the firms could charge. Munn and Scott, owners of one of the firms, ignored the law and were sued for failing to comply. Upholding Illinois, Chief Justice Waite argued for the majority in a famous passage:

> ... we find that when private property is "affected with a public interest, it ceases to be ***juris private*** only." This was said by Lord Chief Justice Hale more than two hundred years ago ... and has been accepted without objection as an essential element in the law of property ever since. Property does become clothed with a public interest when used in a manner to make it of public consequence, and affect the community at large. When, therefore, one devotes his property to a use in which the public has an interest, he, in effect, grants to the public an interest in that use, and must submit to be controlled by the public for the common good, to the extent of the interest he has thus created ...

Justice Bradley, in a later case, summarized the Court's findings more succinctly:

> The inquiry ... was as to the extent of the police power in cases where the public interest is affected; and we held that when an employment or business becomes ... a practical monopoly, to which the citizen is compelled to resort, and by means of which a tribute can be exacted from the community, it is subject to regulation by the legislative power.

Munn vs. Illinois was followed by two cases that expanded the states' power to regulate grain and how it was loaded and stored from ships (*Budd vs. New York*, 1892) and to the fixing of storage rates within an entire state (North Dakota) not necessarily at a "gateway of commerce" (*Brass vs. Stoeser*, 1894). The grain cases were followed by a case that permitted Kansas to regulate the fire insurance business (*German Alliance Insurance Company vs. Lewis*, 1914). Then the trend to granting the states broad powers to regulate was halted in a series of decisions in the twenties. In *Wolff Packing Company vs. Court of Industrial Relations* (1923) Kansas was not allowed to fix prices and wages, as it wanted to in a variety of industries; in *New State Ice vs. Liebmann* (1925), Oklahoma was prevented from regulating the ice business; in *Tyson and Brother vs. Banton* (1927) New York was not allowed to fix prices of theater tickets; Tennessee was not allowed to fix the price of gasoline in *Williams vs. Standard Oil Company* (1929); and New Jersey was not allowed to fix employment agency fees in *Ribnik vs. McBride* (1928).

These cases were not generally unanimous. However, the decisions eventually evolved to the conferring of broad powers to the states, with the last landmark case being *Nebbia vs. New York* (1933). Here, the Court allowed New York to regulate the milk industry and to fix milk prices. Justice Roberts wrote:

> It is clear that there is no closed class or category of business affected with a public interest, and the function of courts in the application of the Fifth and Fourteenth Amendments is to determine in each case whether circumstances vindicate the challenged regulation as a reasonable exertion of governmental authority or condemn it as arbitrary or discriminatory ... So far as the requirement of due process is concerned, and in the absence of other constitutional restriction, a state is free to adopt whatever economic policy may reasonably be deemed to promote public welfare, and to enforce that policy by legislation adapted to its purpose. The courts are without authority either to declare such policy, or, when it is declared by the legislature, to override it.

Thus we see the confirmation of Justice Waite's view in *Munn vs. Illinois* that when "one devotes his property to a use in which the public has an interest, he, in effect, grants to the public an interest in that use, and must submit to be controlled by the public for the common good." Public utility executives are aware of the strength of this feeling on the part of the public. The constant affirmation by them of their dedication to serve the public is not idle rhetoric, but grounded in years of experience of incurring the public's wrath if service quality falls below the public's expectations or if prices appear exorbitant or exploitative.

Casual empirical observations suggest that, generally speaking, the fewer the close substitutes for a regulated firm's products and services that are available, the more intense is the feeling on the part of the public that it has an interest in the use of the firm's property. Dissatisfaction with service or price at a shoe store is easily remedied; one simply trades at another shoe store down the street. But there is generally no alternative telephone, electricity, or water company down the street. Further, as already remarked, basic utilities are now generally considered necessities that should be provided by society as an economic right. If a complaint against a utility providing a "necessity" is not satisfactorily resolved, one can be left with a profound feeling of frustration, with no place to turn, like a prisoner despairing of freedom.

It is thus perhaps not surprising that the history of regulation of public utilities is replete with disputes between the state's police powers and personal (or corporate) freedom to deploy property, not only on such matters as service quality and rates, but on matters affecting the basic management of the utility, with regulators asserting an almost ownership say in the utility's activities. In an editorial, L.E. Smart (1982), editor of the *Public Utilities Fortnightly*, discusses movements in state legislatures or by initiative and referendum to exert more control over utilities. He cites a proposal in Maine as an extreme:

> The proposal in Maine goes furthest, it would seem, in trying to weaken investor-owned electric utilities and perhaps eventually supplant them with public power. It would decree that the new Maine Energy Commission (which would replace the public utility commission, board of railroad commissioners, state water storage commission, and office of energy resources) would have responsibility and authority to create a state energy budget, a semiannual energy demand study, a plan for meeting the demand ascertained by the study, an energy conservation plan, and an energy development fund to raise the money to finance projects in the state energy budget. Certainly, if all these features of the proposed law revision became enacted, it is hard to see what initiatives, if any, would be left for electric utilities. They would be in place simply to carry out the will of the all-powerful state energy commission. The motivating force behind these proposals would seem to be not so much the desire to improve regulation as to enlarge the role of the public sector as against that of the private sector.

This brings us to our last proposition of perceived economic injustice:

> *Proposition 6:* The fewer the substitutes for a regulated firm's output, and the more the output is considered an economic right, the more the public expects to exert control over the firm. Denial of control is considered unjust.

5.2.7 Summary

In an earlier era of "political economy," economists did not confine themselves to narrow questions of economic efficiency. Unfortunately, political economy led to doctrine rather than science because of its intermingling of value-laden, normative judgments of "what ought to be" with objective, value-neutral analyses of "what is" (Schumpeter, 1954, p 1141). Great theoretical advances came with the abandonment of doctrinaire "political economy" in favor of the "positive-normative" and "economic efficiency-income distribution" dichotomies. But with the greater ascendancy of economists and their students into responsible policy-making positions, and the greater influence generally of neoclassical economics on public policy, the ethical issues that these dichotomies avoided must once again be confronted. In particular, economists must realize that a policy that attempts to approximate Pareto improving moves is not ethically harmless; in fact, it can have profound ethical consequences. On the other hand, the great ability of economic analysis to analyze diffused as well as concentrated benefits and costs makes it an indispensable tool for the calculus of economic justice. It is suggested that the place to begin to mobilize this tool is to understand how the public actually perceives economic justice.

This chapter attempts to develop such an understanding in the case of public utility regulation. It points out that public utility regulation is a very public activity, that often has extensive media coverage. This in turn gives importance to "framing" or "contextual" effects in what the public perceives to be fair and just. There is a great need for investigations of such effects. The evidence from history, casual observation, and the few experiments that have been performed all point to different perceptions of economic justice in a market setting from the perceptions in a regulated setting.

Acknowledgments

A preliminary version of this paper was presented at a symposium at Airlie House, Warrenton, Virginia, "A Regulatory Medley – From Economic Analysis to Economic Justice", sponsored by Washington and Lee University in cooperation with the Chesapeake and Potomac Telephone Companies, August 18–21, 1981.

This paper represents an ambitious excursion into a number of fields, especially philosophy and law, which are outside of my specialty of public utility economics. To prevent the most obvious blunders of ignorance, omission and commission, to say nothing of logic, I have gone to great lengths to solicit comments from a large number of persons. So many in fact that it is difficult to remember all of them. In the very initial drafts, my colleagues at Bell Laboratories, especially Roy Radner, Peter Linhart, and Frank Sinden were especially helpful and encouraging. In addition, from the Bell System, J. Falk, G.R. Faulhaber, T. Hodge, S.P. Perles, and M. Wish offered comments on early drafts, while A.E. Buchanan, John Carver, Alfred E. Kahn, K.P. Hamister, W.J. Samuels, John Wenders and Jack Wiseman did the same from outside the system. I would like to thank all of these. In addition, I would like to thank the attendees of various seminars that I have given for the comments they have made on the material in this paper. Needless to say, none of these are to be blamed for errors in the paper or for its point of view. For these, only I am responsible.

Notes

1 For years I lectured Bell executives on dead-weight losses and how their elimination could potentially improve the lot of many without hurting anyone. Somehow I avoided dealing with the ethical problems involved in ***actually*** eliminating dead-weight losses. It was only when I tried in my monograph ***Fairness or Efficiency*** to explain the practical relevance of the theoretically economically efficient (Ramsey) prices for a regulated firm that the extent of the ethical problems of approximate Pareto improving moves sank in. To my economists colleagues at Bell Labs and elsewhere I posed the simple question "Why should a policy maker move toward Ramsey prices when the move will clearly generate unhappy losers, sometimes politically very powerful losers?" I could get no fully satisfactory answer. Finally, in writing the monograph, I retreated to pointing out that while the move to Ramsey prices would generate losers, ***not*** making the move meant that potential winners would be denied gains and in a sense would also be losers. I concluded that it was essentially up to the regulators to pick the winners and losers (see Section 5).

2 Space does not permit a discussion of the defense of utilitarianism and variants of it that try to overcome the "harm of innocents" objection (see Murphy and Coleman, 1984, Chapter 2, for an introduction). For example, Harsanyi's (1976) longstanding defense of utilitarianism as an ethical norm is well known in the economics literature. For an extensive ethical critique of utilitarianism, see Rawls (1971); Rawls' critique is discussed by Arrow (1973).

3 Throughout the chapter, I use "constitution" in the sense of Buchanan and Tullock (1962), that is, a set of rules or procedures by which a group or a society can make decisions as a collective.

4 The present chapter is similar in spirit to Scott Gordon's ***Welfare, Justice, and Freedom*** (1980). His taxonomy of property, fair exchange, desert, equality, equality of opportunity, and freedom is very similar to my six "Propositions". The chapter gives concrete examples from the arena of public utility regulation to illustrate these and other factors at work, and can thus be viewed as a case study of some of Gordon's theses.

5 For a more extensive discussion of the doctrine of the just price and its relation to modern regulation, see Manus (1975).

6 In the second set of experiments to test their theory Hoffman and Spitzer (1985) replicated the initial "coin flip/no moral authority" experiments described at the beginning of this section, except for some changes in wording of the instructions to the participants. The wording changes resulted in 61% rather 100% of the pairs opting for an even $7, $7 division, but the overall statistical results were largely unaffected by the instructional effect (see Footnote 35 of Hoffman and Spitzer).

7 The problem of how laws or legal processes are or ought to be divided into disjoint, nonoverlapping sets has also occupied legal philosophers in both positive and normative theorizing. Space limitations do not allow us to review the extensive literature that has resulted (see Murphy and Coleman, 1984). Very briefly, most theories conclude that overlap exists and is inevitable and must be resolved by judges applying their own standards.

8 Such a Covenant was proposed to implement the UN Universal Declaration of Human Rights, which was adopted by unanimous vote (but with abstentions) in 1948. The United States' adoption of the Universal Declaration and its nonadoption of the subsequent Covenant is an interesting story that illustrates the difficulties of delineating the boundaries of perceived economic injustice. (The subsequent discussion of the UN Declaration and the proposed Covenant is based on Green, 1956.)
The introduction of economic rights into the United Nations was evidently unanticipated. After the horrors of World War II, the United Kingdom and the United States felt it important for the United Nations to take a stand on human rights, having in mind civil or political rights, like freedom of speech, assembly, or press, that can only be removed by due process of law. However, the communist bloc responded to the English-speaking initiative by arguing that the concept of human rights should encompass basic economic rights as well as civil rights. The result was the UN Declaration of Human Rights of 30 articles, 23 of which pertain to civil rights, five to economic rights, and two to administrative matters.

A critical distinction between the two forms of rights is that, by and large, civil rights require few economic resources to guarantee, whereas the guaranteeing of economic rights requires considerable resources. This distinction has evoked different emotions among economists. For example, Okun (1975) is not particularly bothered by the fact that economic rights require resources (p 16):

> It is much less expensive, in every sense, to fulfill the right to free speech than a 'right' to free food. But society does provide some costly or resource-using rights, like public education. And one way proponents of equality seek to narrow the differences in standards of living among Americans is to lengthen the list of resource-using rights.

Hayek (1976), on the other hand, is outraged by the whole idea. He argues (p 102), "Nobody has a right to a particular state of affairs unless it is the duty of someone to secure it", and continues (p 103):

> There have recently been added new positive 'social and economic' human rights ... These are claims to particular benefits to which every human being as such is presumed to be entitled without any indication as to who is to be under the obligation to provide those benefits ... It is, of course, meaningless to describe them as claims on 'society' because 'society' cannot think, act, value, or 'treat' anybody in a particular

> way. If such claims are to be met, ... the cosmos of the market would have to be replaced by a taxis whose members would have to do what they are instructed to do ... From this it follows that the old civil rights and the new social and economic rights cannot be achieved at the same time but are in fact incompatible; the new rights could not be enforced by law without at the same time destroying that liberal order at which the old civil rights aim.

Hayek reserves his particular outrage for the Universal Declaration of Human Rights (p 105): "Even the slightest amount of ordinary common sense ought to have told the authors of the document that what they decreed as universal rights were for the present and for any foreseeable future utterly impossible of achievement, and that solemnly to proclaim them as rights was to play an irresponsible game with the concept of 'right' which could result only in destroying the respect for it."

Hayek's worst fears were never completely realized for, to my knowledge, the Draft Covenant to implement the UN Declaration on Human Rights has never been adopted by the US, much less by the entire United Nations. Article 2, Paragraph 1 of the Draft Covenant stated (Green, 1956, p 190): "Each State Party hereto undertakes to take steps, individually and through international cooperation, to the maximum of its available resources, with a view to achieving progressively the full realization of the rights recognized in this Covenant by legislative as well as by other means." In the Senate debates of 1952 and 1953, the economic rights to be implemented by clauses such as this were referred to as "socialist rights" (Green 1956, p 61) and US support for the Covenant was withdrawn.

9 The bill (Senate Bill 5578B) was signed into law by Governor Carey on August 6, 1981 and is now Chapter 713 of New York State Laws of 1981.

10 More specifically, Countryman (1970) recounts that in the United States the first congressional bankruptcy statute was enacted in 1800, to be repealed in 1803 as being more favorable to the merchant class than to farmers. This was followed by a statute passed in 1841 (referred to by Clark), but repealed in 1843 "as a manifestation of the then common view that debtor relief was immoral" (Countryman, p 121). Then came a statute in 1867, repealed in 1878 as being too beneficial to debtors. Finally,

> The last statute was enacted in 1898 in belated response to demands for debtor relief originating in the panic of 1893, and has apparently become a permanent institution. It has been frequently amended and was extensively revised in 1938; there has been no agitation for its repeal.

References

Arrow, K.J. (1951) *Social Choice and Individual Values* (New York: Wiley).

Arrow, K.J. (1973) Some ordinalist–utilitarian notes on Rawls' theory of justice. *Journal of Philosophy* 70: 245–63.

Atkinson, A.B. and Stiglitz, J.E. (1980) *Lectures on Public Economics* (New York: McGraw-Hill).

Baumol, W.J., Panzar, J.C. and Willig, R.D. (1982) *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt, Brace, Jovanovich).

Buchanan, A.E. (1985) *Ethics, Efficiency, and the Market* (New York: Rowman and Allanheld).

Buchanan, J.M. and Tullock, G. (1962) *The Calculus of Consent* (Ann Arbor, MI: University of Michigan Press).

Cahn, E.N. (1949) *The Sense of Injustice* (New York: New York University Press).

Clark, J.M. (1939) *The Social Control of Business* (New York: McGraw-Hill).

Countryman, V. (1970) Bankruptcy. *Encyclopedia Britannica*, Chicago, pp 121–123.
Davis, K.C. (1958) *Administrative Law* (St. Paul, Minn.: West).
Fiorina, M.P. and Noll, R.G. (1978) Voters, bureaucrats and legislators. *Journal of Public Economics* 9: 239–54.
Friedman, M. and Friedman, R. (1979) *Free to Choose* (New York: Harcourt, Brace, Jovanovich).
Goldberg, V.P. (1976) Regulation and administered contracts. *Bell Journal of Economics* 7: 426–48.
Gordon, S. (1980) *Welfare, Justice, and Freedom* (New York: Columbia University Press).
Green, J.F. (1956) *The United Nations and Human Rights* (Washington, DC: Brookings Institution).
Harsanyi, J.C. (1976) *Essays on Ethics, Social Behavior, and Scientific Explanation* (Dordrecht, Holland: Reidel).
Hayek, F.R. (1944) *The Road to Serfdom* (Chicago, IL: University of Chicago Press).
Hayek, F.R. (1976) *Law, Legislation and Liberty* vol. 2 (Chicago, IL: University of Chicago Press).
Hill, C. (1967) *Reformation to Industrial Revolution* (Harmondsworth, UK: Penguin).
Hoffman, E. and Spitzer, M. (1985) Entitlements, rights, and fairness: an experimental examination of subject's concepts of distributive justice. *Journal of Legal Studies* 14.
Homans, G.C. (1976) Commentary, in L. Berkowitz and E. Walster (Eds) *Equity Theory: Toward a General Theory of Social Interaction* (New York: Academic Press) pp 231–44.
Hughes, J.R.T. (1977) *The Governmental Habit* (New York: Basic Books).
Johnson, L.L. (1981) Comments, in *Energy and Communications in Transition*, Proceedings of the Institute of Public Utilities Eleventh Annual Conference (East Lansing, MI: Michigan State University) pp 117–120.
Levine, M.E. (1981) Revisionism revised? Airline deregulation and the public interest. *Law and Contemporary Problems 44* (School of Law, Duke University) pp 179–95.
Loehman, E. and Whinston, A. (1971) A new theory of pricing and decision making for public investment. *Bell Journal of Economics* 2: 606–25.
McDonald, J. (1975) *The Game of Business* (New York: Doubleday).
Manus, P.C. (1975) Competition in regulated industries—economics. *Essays on Economic Issues* (Macomb, IL: Center for Business and Economic Research, Western Illinois University).
Mason, S.E. (1968) Corporation, in *International Encyclopedia of the Social Sciences* (New York: Macmillan and the Free Press) pp 396–403.
Murphy, J.G. and Coleman, J.L. (1984) *The Philosophy of Law* (New York: Rowman and Allanheld).
Musgrave, R.A. (1959) *The Theory of Public Finance* (New York: McGraw-Hill)
Nozick, R. (1974) *Anarchy, State and Utopia* (New York: Basic Books).
Okun, A. (1975) *Equality and Efficiency: The Big Tradeoff* (Washington, DC: Brookings Institution).
Owen, B.M. and Braeutigam, R. (1978) *The Regulatory Game: Strategic Use of the Administrative Process* (Cambridge, MA: Ballinger)
Panzar, J.C. (1980) Regulation, deregulation, and economic efficiency: the case of the CAB. *American Economic Review* 70: 311–15.
Peltzman, S. (1976) Toward a more general theory of regulation. *Journal of Law and Economics* XIX: 211–40
Phillips, C.F. (1969) *The Economics of Regulation* (Homewood, IL: Irwin).
Polanyi, K. (1977), in H.W. Pearson (Ed) *The Livelihood of Man* (New York: Academic Press).
Posner, R.A. (1972) *Economic Analysis of Law* (Boston: Little, Brown).
Posner, R.A. (1974) Theories of economic regulation. *Bell Journal of Economics* 5: 335–58.

Rawls, J. (1971) *A Theory of Justice* (Cambridge, MA: Belknap Press of Harvard University).

Samuels, W.J. (1974) Commentary: an economic perspective on the compensation problem. *Wayner Law Review* 21: 113–34.

Samuels, W.J. and Mercuro, N. (1976) Property rights, equity, and public utility pricing, in *New Dimensions in Public Utility Pricing* (East Lansing, MI: Division of Research of Michigan State University).

Sasseville, K.E. (1981) Comments, in *Energy and Communications in Transition*, Proceedings of the Institute of Public Utilities Eleventh Annual Conference (East Lansing, MI: Michigan State University) pp 110–16.

Schumpeter (1954)

Selten, R. (1978) The equity principle in economic behavior, in H.W. Gottinger and W. Leinfelner (Eds) *Decision Theory and Social Ethics: Issues in Social Choice* (Dordrecht, Holland: Reidel) pp 289–301.

Sharkey, W.W. (1982) *The Theory of Natural Monopoly* (New York: Cambridge University Press).

Smart, L.E. (1982) *Public Utilities Fortnightly* 108: 4–8.

Smith, A. (1776) 1976 edition: *The Wealth of Nations* E. Cannan (Ed) (Chicago: University of Chicago Press).

Stigler, G. (1971) The theory of economic regulation. *Bell Journal of Economics* 2: 3–21.

Thompson, E.P. (1971) The moral economy of the English crowd in the eighteenth century. *Past and Present* 50.

Thurow, L.C. (1980) *The Zero-Sum Society* (New York: Penguin).

Tuchman, B.W. (1978) *A Distant Mirror* (New York: Knopf).

Usher, D. (1981) *The Economic Prerequisite to Economic Democracy* (New York: Columbia University Press).

Varian, H.R. (1975) Distributive justice, welfare economics, and the theory of justice. *Journal of Philosophy and Public Affairs* 4: 223–47.

Walster, E., Walster, G.W., and Berscheid, E. (1979) *Equity: Theory and Research* (Boston: Allyn and Bacon).

Wenders, J.T. (1981) Testimony in *Southern Nevada Homebuilders Association, Inc. et al. vs. Las Vegas Valley Water District*, Case No. A-154270, Eighth Judicial District Court of the State of Nevada in and for the County of Clark.

Wilson, J.Q. (1980) *The Politics of Regulation* (New York: Basic Books).

Yaari, M.E. (1981) Rawls, Edgeworth, Shapley, Nash: theories of distributed justice reexamined. *Journal of Economic Theory* 24: 1–39.

Zajac, E.E. (1978) *Fairness or Efficiency: An Introduction to Public Utility Pricing* (Cambridge, MA: Ballinger).

Zajac, E.E. (1981) Is telephone service an economic right?, in *Energy and Communications in Transition*, Proceedings of the Institute of Public Utilities Eleventh Annual Conference (East Lansing, MI: Michigan State University) pp 94–109.

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 8

ECONOMIC AND GAME-THEORETIC ISSUES ASSOCIATED WITH COST ALLOCATION IN A TELECOMMUNICATIONS NETWORK

William W. Sharkey

1 Introduction

In recent years, scientific interest in the theory and practice of allocating joint and common costs has been extensive and has led to a substantial literature on the subject. A significant part of this literature has been motivated by the problems faced by public enterprises, regulatory agencies, and private firms in setting prices or making other economic decisions in which a specific methodology of cost allocation is required. The telecommunications industry has been and continues to be both a prominent source of problems for theorists and a testing ground for practical methodologies.

In this chapter I describe some of the economic and technological characteristics of the telecommunications industry and some of the attempts to apply game-theoretic concepts to issues involving pricing and allocation of costs in the industry. The most important conclusion reached in the following analysis is that cost allocation in the telecommunications industry cannot be considered without regard for the fundamental interaction of telecommunications with the remainder of the economy. Telecommunications services are consumed as final outputs by households, but in addition are used as inputs in most other productive sectors of the economy. Some sectors of the industry have the characteristic of a natural monopoly with few close substitutes. Other sectors of the industry are currently or prospectively highly competitive because of the erosion of natural boundaries between telecommunications and information processing.

These technological and economic characteristics of the telecommunications industry must be taken into account in any attempt to allocate telecommunications costs. In Section 2, I briefly consider the more important technological and economic characteristics of the industry. Then, in Section 3, I describe some of the progress that has been made in applying game-theoretic concepts to issues of cost allocation.

2 Economic and Technological Characteristics of the Industry

2.1 *Telecommunications Technology and the Need to Allocate Costs*

The single most important characteristic of the telecommunications industry for the discussion herein is that it is a multiple-output industry.[1] Indeed, there are many different ways in which one may describe the outputs of the industry and, generally speaking, with each particular disaggregation of total industry output, there are important issues involving cost allocation. Basic telephone service involves the production of three distinct outputs: access to the network, local usage, and toll usage. Of course, access to the network is a prerequisite for obtaining either local usage or toll usage, but the costs of providing access are totally independent of actual usage.

The distinction between toll and local usage is essentially arbitrary and is based more on social and geographical boundaries than on technology. All end users are connected to a central office, containing a local switch, by means of copper wire, optical fiber, or other technology. The cost of providing and maintaining this local loop comprises the nontraffic-sensitive cost of access. Local usage consists of electronic messages that are transmitted to the end office switch and then back to another subscriber in the same central office or to a different central office in the same local service. Toll usage consists of messages that follow a similar sequence of paths, except that a typical toll call travels through a more complex hierarchy of switching offices before reaching its final destination.

Perhaps the most fundamental issue of cost allocation is to assign the fixed costs of providing access to customers or services so as to satisfy standards of equity and of economic efficiency. Currently in the US, charges for access to the network are bundled together with a flat fee for a minimum amount of local usage. In European countries, access is priced separately and local usage is often priced on the basis of number of calls, duration, and time of day. In both North American and European systems, it is generally true that the revenues from local service and access together are less than the total costs of providing these services (Mitchell, 1979). The difference is made up from a surplus of revenues over costs for toll traffic.[2]

Going beyond the basic subsidy from toll usage to access and local usage, one can find innumerable instances of cost allocation based upon alternative disaggregations of output. For example, even if access is unsubsidized by other services, it would be necessary to allocate the fixed costs of access among business and residential subscribers, urban and rural subscribers, and so on. In the case of toll service, a problem arises because the same transmission and switching equipment is used to provide dedicated private line service, ordinary toll service, and various bundled services, such as Wide-Area Telephone Service (WATS) in the United States. While these services are technologically similar, they are provided in substantially different markets to largely distinct sets of subscribers. Therefore, in order to determine prices, it is necessary for the supplier of these services to have an explicit cost allocation methodology.

As a final observation on the disaggregation of output, it should be noted that telecommunications services, both toll and local, consist ultimately of the point to point transmission of messages between fixed geographic locations. In order to design an efficient network, given the economies of scale that are possible to achieve in transmission, it is desirable to concentrate traffic in a relatively small number of high-capacity routes. However, since telephone traffic is also probabilistic, and varies periodically over hours of the day and days of the week, it is most efficient to follow a practice known as alternative routing.[3] This means that as a typical toll call enters the lowest level in the hierarchy of switching machines, it is assigned a direct route to its final destination, which is generally the shortest geographic path through the network. If the direct path is being used to full capacity, an alternative route is selected which is geographically longer, but which utilizes higher-capacity transmission media and which therefore has a lower probability of being blocked. If the first alternative route is blocked, a second alternative is chosen, and so on, until a final routing is reached at the highest level of the switching hierarchy.

As a result of alternative routing and other techniques in network design, it is difficult to directly assign the costs of individual pieces of equipment to specific outputs in the network. Consequently, higher levels of aggregation are required in order to allocate the total costs of toll usage. In both the United States and Europe, prices are set on the basis of average cost per unit of distance. While this method of allocation is simple to administer, it has given rise to a substantial subsidy in toll revenue from the high-density, low-cost routes to the low-density, high-cost routes.

2.2 *Competition in Telecommunications Markets*

I have argued above that the technology of telecommunications gives rise to a large number of situations in which joint and common costs must be allocated among different outputs or classes of service in order to set prices. In this section, I argue that in setting prices, it is also necessary to account for the different degrees of potential and actual competition in telecommunications markets. Since the opening of telecommunications markets to competitive entry is considerably more extensive in the US than elsewhere, this section refers more specifically to telecommunications in the US.

The growth of the telecommunications industry in the US has followed a rather uneven course.[4] After the invention of the telephone by Alexander Graham Bell in 1876, the Bell Telephone Company held a monopoly in the industry due to its original patents. After the basic patents expired in 1893, there was competitive entry in the markets for local telephone services as independent companies established themselves in communities not yet served by Bell and occasionally in direct competition with already established Bell companies. During this period the Bell companies, under a newly formed parent company known as the American Telephone and Telegraph Co. (AT&T), began to interconnect the local service companies with long-distance lines. In the early years of the twentieth century, both AT&T and the independent companies grew rapidly. The period was also a time of rate wars, litigation over patents, and occasional mergers between Bell and independent companies. In 1913 the dominant firm in the industry, AT&T, agreed to refrain from further merger activity and to interconnect its long-distance network with the remaining independent companies. In 1934, the Communications Act was passed establishing the Federal Communications Commission

(FCC) with regulatory authority over the industry. AT&T owned approximately 80% of the telephones in the country and most of the long-distance network, while approximately 1500 independent companies provided local service in the remainder of the country.

For the next thirty years, the entire telecommunications industry could be characterized as a regulated monopoly. However, in the early 1960s, several decisions were made by the FCC and by the courts, which set in motion a chain of events that will almost certainly end the era of regulated monopoly in all sectors of the industry.[5]

In 1959, the FCC decided to allow entry into the market for private lines services by private firms who wished to construct their own communications networks. In 1969, the FCC agreed to allow a "specialized common carrier" known as Microwave Communications Inc. (MCI) to construct a microwave link between Chicago and St. Louis. In 1977, a federal court ruled that MCI could not be prevented from offering a service similar to ordinary toll service, thereby overturning an earlier FCC ruling denying such authorization. Five years later, in 1982, AT&T agreed to a consent decree with the US Department of Justice, thus ending a long-running antitrust case against the company. In the decree, which was effective from 1984, AT&T agreed to divest itself of its 22 wholly owned local operating companies. In return, AT&T was given permission to enter new, unregulated markets such as data processing, and it was understood by both parties to the decree that AT&T's toll service markets would ultimately become deregulated as well.

It is almost certainly true that the ultimate consequences of the actions taken by the FCC, the courts, and the executive branch were not correctly foreseen at the time they were made. Furthermore, the consequences were unforeseen largely because of a misunderstanding of, or a lack of appreciation of, the difficulty of allocating costs for a regulated firm serving markets with differing degrees of competition. As noted in the previous section, Bell System rates for both private line and toll services were traditionally set on the basis of systemwide average cost per unit of distance. Thus, a private line or toll call between two rural locations was charged the same amount as a line or call between two urban locations separated by the same distance. Due to the pronounced economies of scale in transmission, it was therefore possible for a rival company to enter selected high-density markets and provide service that cost less than average Bell System rates. While many firms perceived the opportunity, however, only a few had the persistence and political astuteness to win the prolonged regulatory battles.

Once a few competing firms gained a foothold in the industry, however, the situation changed dramatically. Under totally regulated monopoly, cost allocation decisions were made on the basis of administrative convenience, simple rules of thumb, and loose interpretations of regulatory mandates. Under regulated competition – that is, when AT&T served both monopoly markets and markets open to competitive entry – the survival of new entrants was often at stake in the choice of a cost allocation procedure. For example, when entry was first allowed in private line markets, AT&T's response was to lower its own private line rates to reduce incentive for the construction of private networks. Suppliers of equipment for the private networks complained that AT&T was using its monopoly services – that is, toll service – to subsidize its newly competitive services. Evidence of the alleged subsidization was based upon a particular method of fully distributed cost allocation, which showed the rate of return from toll service to be higher than the rate of return from private line service. AT&T responded that its prices for private line service were greater than incremental costs and therefore that the private line service was not subsidized in an economically meaningful sense. While most economists supported the latter position, it was never fully accepted by the FCC, in part because data to measure the true incremental cost could come only from AT&T itself.

Ultimately, the test for cross-subsidization was conducted in the political arena. When the market share of new entrants was insignificant, the FCC was reluctant to accept cost allocation rules that would favor AT&T at the expense of its rivals. However, cost allocation rules that were able to protect incumbent rivals also had the effect of encouraging new entry. Soon it became clear that the FCC would either have to expand its authority to, in effect, dictate market shares and levels of entry in all markets or, alternatively, to initiate the process of total deregulation of all markets.

The potential for competition in local service markets is only now beginning to be perceived.[7] As in the case of toll and private line markets, competition can only occur in markets in which price is greater than cost. As noted in the previous section, in both the US and European systems, prices for access and local service are generally less than incremental cost, and the difference has traditionally be subsidized from the revenues of toll services. Furthermore, since prices for toll service are usage-sensitive, it is the very large toll users who have contributed disproportionately to the subsidy.[8] Consequently, it is the large toll users that constitute the most attractive market for entry by firms who provide alternative access to end users, so that they can avoid subsidizing the access of other subscribers. As in the case of entry into private line services, the first instances of entry occurred when large business users constructed private networks in order to bypass completely the public network. New technologies, such as cellular mobile radio, optical fibers, and satellites, as well as traditional microwave technology, may also be used to replace copper wires in the local loop (Bell Systems Operating Companies and AT&T, 1982). The attractiveness of these alternative technologies depends on the pricing policies of existing local service telephone companies. Their prices in turn depend on the method they use to allocate costs, including the fixed costs of access.

3 Game-Theoretic Analysis of Cost Allocation in Telecommunications

3.1 *Cost-Based Tests for Cross-Subsidization*

It was argued in the previous sections that while allocations of joint and common costs are essential in setting prices for telecommunications services, it is also necessary to account for the economic characteristics of telecommunications markets. In this section, I illustrate some of the ways in which game-theoretic analysis could be potentially useful in telecommunications pricing and cost allocation.

I believe that the most important overall contribution that game-theoretic analysis has made and will continue to make is the definition of a rigorous framework in which issues of both fairness and efficiency can be addressed. Both issues are very much apparent in the literature on cross-subsidization.[9] The question of cross-subsidization arises primarily when a regulated or public sector firm must allocate joint or common costs among two or more product lines that are sold in distinct markets. In the simplest case, one can imagine two discrete outputs that can be produced at a total cost of

$$C(1,2) = C_0 + C_1 + C_2$$

where C_0 is a common cost while C_1 and C_2 are directly attributable costs of outputs one and two, respectively. Clearly, if r_1 and r_2 represent the revenues collected from customers in markets one and two, then one may say the revenues or prices are subsidy-free if $r_1 \geq C_1$ and $r_2 \geq C_2$, so that each output covers at least its directly attributable cost. If, in addition, total revenues must cover the total costs so that $r_1 + r_2 = C(1,2)$, then there are no uniquely determined subsidy-free revenues. The economic approach suggests allowing demand conditions in markets one and two to determine a unique allocation.[10] However, if only the cost function is used as a basis for allocation, little more can be said.

One contribution of game theory to the theory of cross-subsidization has been to extend the simple analysis above to more complex situations involving three or more outputs (Faulhaber, 1975). For example, if there are n discrete outputs and $N = \{1, \ldots, n\}$ denotes the set of all outputs, then a revenue vector $\boldsymbol{r} = (r_1, \ldots, r_n)$ is ***subsidy-free*** if

$$\sum_1^n r_i = C(N) \tag{1}$$

$$\sum_{i \in S} r_i \geqq C(N) - C(N \setminus S) \quad \text{for all } S \subseteq N \quad . \tag{2}$$

Condition (1) is merely the break-even constraint. Condition (2) represents an "incremental cost test" for subsidization, since it requires that revenues from each collection S must be at least as great as the incremental cost of producing that collection. Clearly, given (1), condition (2) can be rewritten as

$$\sum_{i \in S} r_i \leqq C(S) \quad , \tag{3}$$

which suggests a "stand alone test" for subsidization. Namely, the revenue vector $\boldsymbol{r}$ is subsidy-free if each collection S pays no more than would be required if it were to stand alone.

Either one of the above equivalent tests for cross-subsidization seems to embody a reasonable concept of equity. However, it is important to note that when there is competition in one or more of the markets, the tests must also be passed in order to guarantee the most basic requirement for economic efficiency. As an example, let us assume that the firm or public enterprise is a natural monopoly, in the sense that the cost function C is strictly subadditive. That is,

$$C(S) + C(T) > C(S \cup T) \quad \text{whenever } S \cap T = \phi \quad .$$

Suppose that a particular coalition $N \setminus S$ is subsidizing the complementary coalition S. Then it follows that

$$\sum_{i \in S} r_i < C(N) - C(N \setminus S) < C(S) \quad , \tag{4}$$

where the first inequality follows from the incremental cost definition of subsidization and the second follows from the definition of subadditivity.[11] Equivalently, one may write

$$C(N) - C(S) < C(N \setminus S) < \sum_{N \setminus S} r_i \quad . \tag{5}$$

If there is a competitive supplier of outputs $N \setminus S$ who can provide them at a cost $G(N \setminus S)$ such that

$$C(N) - C(S) < G(N \setminus S) < \sum_{N \setminus S} r_i \quad , \tag{6}$$

then the customers of outputs $N \setminus S$ will surely defect. But then equation (4) shows that the customers remaining in markets S must pay not only their incremental costs but also their total stand alone costs. Furthermore, the total costs of production are larger since

$$G(N \setminus S) + C(S) > C(N) - C(S) + C(S) = C(N) \quad . \tag{7}$$

Thus, subsidization when there is also competition can harm both the subsidized customers and aggregate economic efficiency.

Conditions (1) and (2) define the incremental cost test for cross-subsidization and are equivalent to the conditions defining the core of a game, in which the characteristic function is the incremental cost of production. However, subadditivity of the total or stand alone cost does not imply superadditivity of the incremental cost. An alternative game that is superadditive is defined by the characteristic function $v(S) = -C(S)$. In either case, it is well known that cores need not exist unless more restrictive conditions are placed on the cost function. Thus, another contribution of game theory to the literature on cross-subsidization is the proposition that subsidy-free prices may not exist.

The definition of cross-subsidization that is based only on the cost function has a number of important limitations. As already noted there are, in general, many subsidy-free allocations and there is no apparent method of choosing one subsidy-free allocation rather than another. More importantly, the cost-based definition of subsidization does not make use of any available additional information concerning competitive conditions in the various markets or the willingness of customers to pay. From the discussion of competition in telecommunications markets, it should be clear that this is a serious omission. When there are alternative sources of supply, or other constraints based on consumer demand, it clearly follows that not all subsidy-free cost allocations are viable.

Suppose that, in the example discussed above, the revenue vector $\boldsymbol{r}$ is adjusted so that the group $N \setminus S$ no longer subsidizes S and that no other subsidization constraints are violated. Nevertheless, it may still happen that there is an alternative source with costs $G(N \setminus S) < C(N \setminus S)$ such that

$$C(N) - C(S) < G(N \setminus S) < \sum_{N \setminus S} r_i \leqq C(N \setminus S) \quad . \tag{8}$$

As before, the customers of $N \setminus S$ have the incentive to defect. If they do, then the remaining customers of S bear the cost and aggregate efficiency is lowered, as again illustrated by inequalities (7).

3.2 *Analysis of Benefit Games*

To extend the game-theoretic analysis to include information on consumers' willingness to pay, one may proceed as follows. Suppose that $\boldsymbol{b} = (b_1, \ldots, b_n)$ is the vector representing the maximum willingness to pay of consumers of each product.[12] Then if $C(S)$ represents the cost of producing the collection S of outputs, it is possible to define the following characteristic function for a benefit game,[13]

$$v(S) = \max_{R \subseteq S} \left\{ \sum_{i \in R} b_i - C(R) \right\} . \tag{9}$$

The function v represents the maximum surplus obtainable by the coalition S of markets. It is easy to see that v is monotonic, i.e., that $v(S) \geqq v(T)$ whenever $S \supseteq T$; that v is superadditive if C is subadditive; and that $v(S) \geq 0$ for all S, since $C(\phi) = 0$. Furthermore, it follows that

$$v(N) = \sum_N b_i - C(N)$$

if and only if

$$\sum_{i \in S} b_i \geqq C(N) - C(N \setminus S) \quad \text{for all } S \subseteq N \quad . \tag{10}$$

In other words, it is optimal to produce the entire product set N if and only if the combined willingness to pay of every coalition S is at least as great as the incremental cost of producing S.

It was shown in the previous section that one definition of subsidy-free prices or revenues could be made with reference to the core of an appropriate cost game. Similarly, when there are demand constraints, subsidy-free prices or revenues can be determined by examining the core of the benefit game. Let $\boldsymbol{y} = (y_1, \ldots, y_n)$ be a vector representing the consumer surplus in each market. Then, assuming that conditions (10) are true, it is a straightforward exercise to show that there exists a vector $\boldsymbol{y}$ in the core of the benefit game if and only if there exists a subsidy-free revenue vector $\boldsymbol{r}$ for which no market is charged more than its willingness to pay. That is, the core of the benefit game is nonempty if and only if there exists a vector $\boldsymbol{r}$ such that

$$\sum_N r_i = C(N) \tag{11}$$

$$\sum_N r_i \geqq C(N) - C(N \setminus S) \text{ for } S \subseteq N \tag{12}$$

$$r_i \leqq b_i \quad \text{for } i \in N \tag{13}$$

Clearly, the subsidy-free prices in a benefit game are a subset of the subsidy-free prices in the cost game, because of the additional constraints (13). Furthermore, it is possible that the core of the benefit game is empty even though the core of the cost game is nonempty. However, for some technologies, it is possible to guarantee that the core of the benefit game is nonempty for every set of benefit vectors $\boldsymbol{b}$ that satisfy the optimality conditions (10).

Let $C^*(S) = C(N) - C(N \setminus S)$ be the incremental cost function. It has been demonstrated (Sharkey, 1981, 1982b) that if the incremental cost function C^* is convex, that is, if

$$C^*(S) + C^*(T) \leqq C^*(S \cup T) + C^*(S \cap T) \tag{14}$$

for all coalitions S and T, then for any willingness to pay vector satisfying (10) there is a revenue vector that satisfies (11)–(13). Equivalent to (14) is the condition that the total cost function be "concave" so that

$$C(S) + C(T) \geqq C(S \cup T) + C(S \cap T) \quad . \tag{15}$$

Condition (15) describes a property known as "cost complementarity", which has proven to be useful in the study of multiple-product cost functions. It should be emphasized that cost complementarity is sufficient but not necessary for the existence of a subsidy-free revenue vector satisfying (11)–(14). The following section will describe a different class of nonconcave cost functions for which the core of the benefit game is always nonempty.

3.3 *Cost Allocation in a Spanning Tree Game*

There is a type of cost allocation game known as a minimum cost, spanning tree game that has particular relevance for technologies involving construction of a network.[14] Minimum cost, spanning tree games may be relevant in allocating the cost of access to local communications networks. In this game, one takes as given the set of nodes $N \cup \{0\}$ and the matrix $\{C_{ij}\}$ representing the cost of constructing an arc linking nodes $i, j \in N \cup \{0\}$. Node $\{0\}$ is a distinguished node representing the common supplier of network services and N represents the set of potential subscribers to the network service. For any coalition $S \subseteq N$ one then defines the cost function $C(S)$ representing the sum of costs of the arcs in a minimum cost spanning tree which connects the nodes $S \cup \{0\}$. It is well known that minimum cost, spanning tree games always have cores (Granot and Huberman, 1981). In fact, the revenue vector that assigns to each subscriber the cost of the associated link in a minimum cost spanning tree is itself in the core. Since there exist efficient algorithms for calculating minimum cost spanning trees, it would appear that finding a subsidy-free revenue vector in a spanning tree game is a straightforward task. However, when the subscribers' willingness to pay is also known no simple algorithm exists for finding revenues in the core of the benefit game. While a spanning tree game with willingness to pay constraints does have a nonempty core, it is necessary to solve explicitly the $2^n + n$ inequality constraints given in (11)–(14) in order to determine a subsidy-free revenue vector (Sharkey, 1985).

When the spanning tree game is generalized in various ways, subsidy-free revenues may no longer exist. For example, if there is more than one potential location for the source node or if new nodes may be created in order to construct a minimum cost tree, then the core of the cost game need not exist (Bird, 1976; Megiddo, 1978a). Moreover, if there are positive incremental costs when two or more users share a link, as there are in telecommunication networks, the core may also be empty (Sharkey, 1985). Thus, while the minimum cost, spanning tree game has an interesting mathematical structure that may ultimately prove useful in resolving the cost allocation question, such games appear now to have virtually all of the complexity of general cost allocation games.

4 Concluding Comments

In this chapter, I have attempted to survey the technological and economic characteristics of the telecommunications industry and to indicate how some of the methods of cooperative game theory might be used in formulating rules for cost allocation. While the technology of telecommunications is such that cost allocation questions frequently arise, one must consider the economic characteristics of telecommunications markets before making a final choice of a method of allocation. While game-theoretic concepts have frequently been applied to questions of pure cost allocation, where economic constraints do not intrude, I believe that game theory has also been shown to be a useful aid in understanding some of the economic issues.

There is, of course, much additional work that could be done. Highly structured games, such as the minimum cost, spanning tree game have not been extensively studied. Important economic issues, such as the basic externality inherent in communications services, have not been incorporated in most game-theoretic discussions. Finally, there is much to be done in the area of finding useful computational results. What has been achieved, however, is the specification of a simple, yet relevant model of the cost allocation process that can be used to guide current policy and to encourage further research.

Notes

1. See Sharkey (1982a, Chapter 9) for further discussion.
2. The subsidy from toll service to local service and access has been justified traditionally by the perceived need to promote universal service – that is, to maximize the number of households having a telephone. Additional arguments for subsidized access may be based upon the externalities inherent in a communications service. See Sharkey (1982a, Chapter 9) and the references cited therein for further discussion.
3. The practice of alternative routing is described more fully in Bell Laboratories (1977).
4. See Brock (1981) for a more detailed account of this period.
5. See Johnson (1982) for additional discussion of recent competitive inroads into the industry.
6. The method, known as FDC-1, is described more fully in Johnson (1982).
7. See Bell System Operating Companies and AT&T (1982) for a discussion of bypass technology and the potential for competition in local service markets.
8. In Bell System Operating Companies and AT&T (1982), it is reported (p 92) that "10% of the residence customers account for approximately one-half of all residence interstate MTS messages; 10% of the business locations account for approximately three-fourths of all business interstate MTS revenues; and 10% of the interstate WATS locations account for approximately 60% of all interstate WATS revenues".
9. See Faulhaber (1975) and references cited therein for further discussion of cross-subsidization.
10. This approach, known as Ramsey pricing, is explained in Baumol and Bradford (1970), Boiteux (1971), and Ramsey (1927).
11. See Sharkey (1982a) for further discussion and analysis of subadditivity.
12. The demand vector $\boldsymbol{b}$ might also be interpreted as a vector of competitive prices from outside suppliers. Note also that under the willingness to pay interpretation, it is assumed that demands are independent across markets.
13. Benefit games are discussed further in Littlechild (1975) and Sharkey (1982c).
14. Cost allocation on a spanning tree was first discussed by Claus and Kleitman (1973). Game-theoretic analysis is contained in Bird (1976), Granot and Huberman (1981), and Megiddo (1978a,b).

References

Baumol, W.J. and Bradford, D.F. (1970) Optimal departures from marginal cost pricing. *American Economic Review* 60: 265–83.

Bell Laboratories (1977) *Engineering and Operations in the Bell System* (Indianapolis, Indiana: Western Electric Co.).

Bell System Operating Companies and American Telephone and Telegraph Company (1982) Comments in the matter of MTS and WATS market structure. *Response to the Fourth Supplemental Notice of Inquiry and Proposed Rulemaking* (FCC Docket 78–72).

Bird, C.G. (1976) On cost allocation for a spanning tree: a game theoretic approach. *Networks* 6: 335–50.

Boiteux, M. (1971) On the management of public monopolies subject to budgetary constraints. *Journal of Economic Theory* 3: 219–40.

Brock, G.W. (1981) *The Telecommunications Industry: The Dynamics of Market Structure* (Cambridge, MA: Harvard University Press).

Claus, A. and Kleitman, D.J. (1973) Cost-allocation for a spanning tree. *Networks* 3: 289–304.

Faulhaber, G.R. (1975) Cross-subsidization: pricing in public enterprises. *American Economic Review* 65: 966–77.

Granot, D. and Huberman, G. (1981) Minimum cost spanning tree games. *Mathematical Programming* 21: 1–18.

Johnson, L. (1982) *Competition and Cross-Subsidization in the Telephone Industry*. Rand Report R-2976-RC/NSF (Santa Monica, CA: Rand Corporation).

Littlechild, S.C. (1975) Common costs, fixed charges, clubs, and games. *Review of Economic Studies* 42: 117–24.

Megiddo, N. (1978a) Cost allocation for Steiner trees. *Networks* 8: 1–6.

Megiddo, N. (1978b) Computational complexity and the game theory approach to cost allocation for a tree. *Mathematics of Operations Research* 3: 189–96.

Mitchell, B. (1979) Pricing policies in selected European telephone systems, in H.S. Dordick (Ed) *Proceedings of the Sixth Annual Telecommunications Policy Research Conference* (Lexington, MA: Lexington Books).

Ramsey, F. (1927) A contribution to the theory of taxation. *Economic Journal* 37: 47–61.

Sharkey, W.W. (1981) Existence of sustainable prices for natural monopoly outputs. *Bell Journal of Economics* 12: 144–54.

Sharkey, W.W. (1982a) *The Theory of Natural Monopoly* (London: Cambridge University Press).

Sharkey, W.W. (1982b) Cooperative games with large cores. *International Journal of Game Theory* 11: 175–82.

Sharkey, W.W. (1982c) Suggestions for a game-theoretic approach to public utility pricing and cost allocation. *Bell Journal of Economics* 13: 57–68.

Sharkey, W.W. (1985) *Cores and Implications for Pricing in a Model of Production with Fixed Costs and Shared Facilities*. Bell Communications Research Economic Discussion Paper No. 1 (Morristown, NJ: Bell Communications Research).

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 9

COST ALLOCATION AND PRICING POLICY: THE CASE OF FRENCH TELECOMMUNICATIONS

Nicolas Curien

1 Introduction

This chapter surveys some of the potential applications of cost allocation methods and principles to the field of telecommunications. We present a simple methodology for estimating marginal costs of different outputs and, using this as a reference system of prices, we study the revenue trade-offs that exist between users and between different outputs under current pricing policy in the French telecommunications network. This amounts to a pragmatic estimation of the extent of cross-subsidization between different classes of users and outputs. We discuss the implication of these results for French pricing policy for telephone services in terms of equity and economic efficiency.

Cost allocation principles and techniques have many potential applications in the field of telecommunications (for surveys more oriented toward the North American context, see Chapters 7 and 8). One of these is how to allocate telephone costs between basic access to the network, lines operations, local usage, and long-distance usage. Such allocations are needed to establish reference costs for carrying out economic studies and designing pricing policies.

Similarly, it may be necessary or desirable to share costs between different categories of usage in the network. Peak users cost more to the telephone industry than off-peak ones, access is more costly to provide in rural than in urban areas, toll calls cost more than local ones but less than tariffs tend to show, etc. Comparing actual rates to a theoretically fair and/or efficient pricing system allows us to draw a map of cross-subsidizations, i.e., revenue trade-offs that occur between types of customers and between classes of usage.

Third, there is the issue of how to share the common costs of a trunk telephone network between different regional operating units, in order to build a decentralized accounting system in which long lines and local operations are treated as separate cost centers.

Another problem is how to allocate the common costs of transmission and switching capacity between several different telecommunications services using the same network, such as voice and data transmissions. This issue is important in France, where narrow-bandwidth nonvoice traffic (facsimile and videotext) is accommodated in part by the basic telephone network. Economic rationality in pricing might suggest charging per information-bit transmitted regardless of the nature of messages, whereas commercial purposes might encourage a differentiation of prices according to services. Moreover, if these new services were to be run in the future by operators outside the monopoly, it would become a major concern for PTT to set cost-based rates for the network facilities supplied to operators and to price in a consistent way the intermediate demand (operators) and the final demand (users).

Finally, cost-sharing rears its head in the use of transcontinental telecommunications equipment. Each national company collects all of the revenue from users' outgoing traffic and pays to the intermediate or final destination countries (or to international organizations for satellite and underseas circuits) a fee based on the cost of equipment utilized. Conversely, each company receives a fee per minute of incoming or transit traffic run through its network. These fees, which constitute an internal pricing system between telephone companies, are determined through a negotiation process that should allow for allocation principles.

In this chapter, we concentrate on the first two applications: building reference costs and assessing the revenue trade-offs inherent in the tariffs. We then discuss the French telephone pricing system and explain the main orientations of the policy planned by the Direction Générale des Télécommunications toward "efficient" pricing and "fair" control of trade-offs.

2 Normative Pricing Schemes

In theory, a telecommunications monopoly such as the French PTT should price at marginal costs in order to maximize the overall social surplus, i.e., the sum of the industry's profits and consumer surplus (see Abraham and Thomas, 1966). But if, for instance, owing to increasing returns to scale, marginal pricing does not allow total costs to be recovered, the budget balance can be restored through taxes. In this trade-off, consumers yield to the industry part of their surplus and the overall social surplus remains constant and maximal. However, such a tax system is certainly more theoretical than realistic and some care is needed so that they are shared among consumers in a Pareto-optimal way. Moreover, it might be undesirable insofar as it creates poor incentives for a cost-minimizing policy for the industry. Economic theory then recommends setting "second-best" prices that maximize consumer surplus, subject to profit constraints or to budget balances, the so-called Ramsey prices (see Ramsey, 1927; Baumol and Bradford, 1970; Boiteux, 1971). Despite their second-best status, Ramsey prices possess a number of properties that are liable to disqualify them as a reference system. First, reference prices should be a representation of industry costs, and therefore should rely only upon production parameters. Ramsey prices do not satisfy this condition since they also depend on demand elasticities. Assuming zero cross-elasticities of demand, they depart from marginal costs in inverse proportion to elasticities. For instance, the color of telephone sets could influence Ramsey prices if demand is color-sensitive, even though production costs are indifferent to this factor. Second, Ramsey prices reveal a disputable notion of equity, based on a willingness-to-pay principle: inelastic segments of demand are marked up more than elastic ones, which would lead (in the example of telephone usage) to overcharging intermediate demand (businesses) because it is less elastic than final demand (households). This may seem unfair, since the former needs the telephone service as an input to production and thus represents a more captive market than the latter.

Third, it has been shown that for a natural monopoly with economies of scale, which is reputed to be the case for at least a portion of the telecommunications industry (for an estimation of economies of scale and economies of scope in telecommunications, see Christensen *et al.*, 1983; Kiss *et al.*, 1983), Ramsey prices are not necessarily subsidy-free. In other words, some outputs may generate less revenue than their incremental cost (the extra cost of supplying these outputs under provision of the others) and some may generate more revenue than their stand-alone cost (the cost of supplying these outputs without providing the others). In this case the latter subsidize the former (a discussion of this point is given in Rheaume, 1983, and a model yielding welfare-optimal, subsidy-free prices is developed.). This opens Ramsey prices to criticism as a norm. From a policy standpoint there is another serious problem with Ramsey prices – or indeed any prices that allow cross-subsidization. If barriers to entry are not too strong – as in the North American context – such prices may encourage inefficient competition in the market for subsidizing services and discourage it in the market for subsidized ones where it could be sociably desirable (see Chapter 8).

There are price systems other than Ramsey that are cost sharing and do not exhibit the above-mentioned defects. For instance, prices that maximize surplus subject to being subsidy-free and satisfying a profit constraint; or Aumann–Shapley prices which, under suitable conditions on the cost function, are subsidy-free, coincide with marginal costs in the case of constant returns to scale, and generalize in a natural way the notion of average cost for a one-output industry (see Chapters 1 and 3). The diversity of possible pricing systems is in fact merely an indicator of the range of suboptimality, each separate system pointing to a different concept of equity, or monopoly sustainability, or second-best efficiency, etc.

Faced with these issues and with the numerous possible answers, we considered both the availability of data, which does not permit very elaborate estimation techniques, and the purpose of this study. Rather than assessing the discrepancies between actual pricing and some desirable and/or attainable goal, the aim was to build an "ideal" ("first best") reference system with which we could compare all feasible pricing systems consistent with budget-balancing constraints. Among these are actual pricing, Ramsey pricing, and other "suboptimal" cost-sharing systems. From this standpoint, the evaluation of marginal costs appeared to be the first and the most useful step in the process of gaining better knowledge of production conditions in French telecommunications.

3 Estimating Marginal Costs as a Reference Base

In computing marginal costs one has to account for the multi-output character of telecommunications, including access to the network, lines operations, local traffic, long-distance (toll) traffic, and international traffic. These are in fact joint outputs; traffic cannot be provided without access, nor toll and international traffic without local traffic. Moreover, some equipment is shared by at least two functions. For instance, in local exchanges some transit switching equipment run both local and toll traffic; buildings, energy and logistics serve all activities.

Although the cost function is not separable according to outputs, the pricing function is, and the consumer pays separately for access through an initial connection fee, for lines operations through a bimonthly rental charge, and for traffic through a distance and time-of-day-sensitive rate based on periodic pulses.

Assessing marginal costs and comparing them with tariffs thus requires either evaluating the total cost function from which marginal costs can be derived, or estimating marginal costs directly by allocating elementary expenses. The first method means evaluating a multi-output, nonseparable cost function by econometric techniques. (The functional form of the cost function as a translog second-order approximation is mentioned in Fuss, 1983; Chistensen *et al.*, 1983; Kiss, *et al.*, 1983.) This involves a number of difficulties including specification of the form of the cost function and a multicollinearity among the variables (such as changes in traffic and mainstation stock), not to mention the lack of consistent cost data over space and time due to imperfections and changes in the accounting system.

The second method consists of referring to detailed accounts and measuring the expenditure (investment and operations) occasioned over a period of one year by the increase in the various outputs. (A more detailed presentation of marginal cost calculations from the French telecommunications accounting system is given in Curien and Pautrat, 1983.) In this approach, joint and common costs have to be allocated, first to telephone (as opposed to other services such as telex or telegraph using some of the same network facilities) and then to connections, lines operations, and traffic.

Although less attractive than the first method from a theoretical standpoint, the second was preferred because of its stability under uncertainty and incompleteness of data, its structural consistency with the accounting system and, not least, its better understandability by management. In order to obtain results quickly, the cost allocation process was carried out using classical empirical rules. (Such rules are referred to as "easy joint cost allocation approaches" in Thomas, 1980). For instance, technical expenditures are shared between lines and traffic in the same proportion as the equipment they contain. Such a choice was dictated by the relative weakness of data, and a belief that not much would be gained from taking account of sophisticated tools related to game-theoretic concepts (see, for instance, Littlechild, 1970; Faulhaber, 1975). Such tools generally require more inputs than are available from basic accounts, and their useful application assumes that data quality fits methodological refinement, which was certainly not the case when marginal cost calculations were first made for French telecommunications in 1976. Since then, accounts have been checked and improved through usage, and cost data are more reliable, so that further reference to principles of allocation theory should be possible and worthwhile in the future.

Table 9.1 shows the marginal costs per unit of output, computed as indicated above, for the years 1976, 1978, 1979 (in 1977 important changes in the accounting system do not allow reliable computation of costs); investment and operations costs are separated. The mainstation (MS) is the unit used to measure lines connections or operations outputs, and the erlang (E) unit is used to measure the traffic capacity of the network. The volume of traffic that can be handled by the telecommunications equipment is 1 erlang-hour if the equipment is busy during a one-hour period. The intensity of traffic, i.e., the volume of traffic per time unit is then measured in erlangs, as is the traffic capacity of a network or piece of equipment: 1 E is the capacity that can carry 1 E of peak traffic intensity.

Table 9.1 Marginal cost estimation (thousands of French francs, FF, current value).

Operations

Output	*Subscribers lines* (MS)	*Coin telephone lines* (MS)	*Local traffic* (E)	*Trunk traffic* (E)	*International traffic* (E)
1976	0.64		9.31	22.50	
1978	0.59	3.64	8.83	16.35	176.29
1979	0.59	4.04	9.91	15.04	194.20
Annual % increase (constant FF)					
1978/1976	−12.2		−11.0	−22.1	
1979/1978	−9.8	6.3	1.8	−16.6	−0.001

Investment

Output	*Subscriber connections* (MS)	*Coin telephone connections* (MS)	*Local traffic* (E)	*Trunk traffic* (E)	*International traffic* (E)
1976	3.01		112.0	159.8	
1978	3.43	16.03	144.45	186.90	397.58
1979	3.28	15.39	168.57	219.54	388.41
Annual % increase (constant FF)					
1978/1976	−2.4		3.9	−1.1	
1979/1978	−13.3	−13.0	5.8	6.5	−11.4

Several trends are noteworthy. Investment costs decreased in constant FF except for local and trunk erlangs, due to a policy of decreasing equipment usage rates; more equipment was installed to satisfy a given level of traffic demand, thus lowering the probability of call rejection and improving the quality of service supplied. Operating costs declined as well between 1975 and 1980 (except for coin telephones), mainly in the trunk network. However, it should be observed that the decrease in marginal costs does not necessarily imply increasing returns to scale, for this decrease was caused not only by the growth of output but also by the reductions in equipment and labor costs. More observations than available would be necessary to separate the two effects.

4 Analysis of Trade-offs

The classical way to carry out cross-subsidy tests is to compare, for each set of joint outputs, the revenues on the one hand and the incremental and stand-alone costs on the other. (Such tests have been carried out for the Canadian interregional telecommunications network and are described by Autin and Leblanc, 1983). However, the evaluation of these costs is not yet available for French telecommunications. Therefore we used a simpler approach to measuring cross-subsidization by analyzing the trade-offs associated with variations in consumers surplus. When telecommunications prices deviate from marginal reference costs, this may be measured by the change in surplus generated between the reference state and the actual state. If actual prices exceed marginal costs, the overall change in surplus reflects a global subsidy from consumers to the industry. The variations in surplus for each elementary segment of the market, when compared with "fairly" allocated variations, indicate revenue trade-offs between groups of users and between types of outputs.

More precisely, let q denote the telecommunications output vector and p the associated price vector. Market equilibrium requires that

$$q = q(p) \quad , \tag{1}$$

where $q(p)$ is the (joint) demand function. The surplus is not generally totally differentiable, so that its definition depends on the selection of a path in price space between the marginal cost price reference point p_0 and the actual price p_1. The surplus variation in a transition from marginal to actual price is then

$$S = -\int_{p_0}^{p_1} q(p)\ \mathrm{d}p \quad , \tag{2}$$

where the integral is computed along the selected path from p_0 to p_1. If (as is the case in France) marginal prices p_0 are far from actual prices p_1, the theoretical demand $q_0 = q(p_0)$ is unknown, so that it looks sensible and convenient to assume perfect inelasticity in this price range. Let

$$q_0 = q_1 = q(p_1) \quad , \tag{3}$$

which means approximating the demand function $q(p)$ between p_0 and p_1 by a step in computing the surplus integral (2). This integral no longer depends on the price trajectory and may be written explicitly as

$$S = -q_1 \cdot (p_1 - p_0) \quad , \tag{4}$$

Denote classes of users (business, residential, ...) by an index i and classes of outputs (connections, lines operations, traffic, ...) by an index j. The surplus variation between reference and actual states may then be written

$$S = -\sum_{i,j} q_1^{ij} (p_1^{ij} - p_0^{ij}) \quad , \tag{5}$$

where each term in the sum is the surplus variation of market segment (i,j):

$$S^{ij} = -q_1^{ij}(p_1^{ij} - p_0^{ij}) \quad . \tag{6}$$

The transition from p_0 to p_1 may conceptually be split into two steps. In the first step all users together yield the surplus S to telecommunications, this subsidy being "fairly" allocated in the sense that profitability is assumed to be the same in each market segment. Thus reference share of surplus $\hat{S}^{ij}$ is proportional to the revenue generated in the reference state:

$$\hat{S}^{ij} = S\frac{q_1^{ij}p_0^{ij}}{q_1 \cdot p_0} \quad . \tag{7}$$

Using expression (4) for S this becomes

$$\hat{S}^{ij} = -q_1^{ij}\left[\frac{q_1 \cdot p_1}{q_1 \cdot p_0}p_0^{ij} - p_0^{ij}\right] \tag{8}$$

In the second step the surplus S is reallocated in order to reach the actual allocation S^{ij}, and the trade-off T^{ij} received or given by segment (i,j) is defined as the change in surplus obtained from reallocation, i.e.,

$$T^{ij} = S^{ij} - \hat{S}ij \quad . \tag{9}$$

Substituting expressions (6) and (8) for S^{ij} and $\hat{S}ij$:

$$T^{ij} = -q_1^{ij}\left[p_1^{ij} - \frac{q_1 \cdot p_1}{q_1 \cdot p_0}p_0^{ij}\right] \quad . \tag{10}$$

Let $\hat{p}_0$ denote the price vector proportional to the marginal price vector p_0, adjusted in order to yield the same income as the actual price vector p_1, i.e.,

$$\hat{p}_0 = \frac{q_1 \cdot p_1}{q_1 \cdot p_0}p_0 \tag{11}$$

The share borne by segment (i,j) in the first step of the "fair" allocation is then

$$\hat{S}^{ij} = -q_1^{ij}\left[\hat{p}_0^{ij} - p_0^{ij}\right] \quad . \tag{12}$$

This may be interpreted as the surplus variation in a transition from marginal cost pricing p_0 to proportional pricing $\hat{p}_0$. Similarly, the trade-off T^{ij} in the second step of the reallocation is

$$T^{ij} = -q_1^{ij}\left[p_1^{ij} - \hat{p}_0^{ij}\right] \tag{13}$$

The concept of trade-offs that we introduced may therefore be looked upon as a measure of the distance between two pricing systems satisfying the same profit constraint: prices proportional to marginal costs, and actual prices. Since total telecommunications revenues are the same under each of these systems, the trade-offs must add up to zero:

$$\sum_{i,j} T^{ij} = 0 \quad . \tag{14}$$

The trade-offs between users alone can then be defined as

$$T^{i.} = \sum_j T^{ij} \quad ; \tag{15}$$

and between facilities alone as

$$T^{\cdot j} = \sum_i T^{ij} \quad , \tag{16}$$

with

$$\sum_i T^{i\cdot} = \sum_j T^{\cdot j} = 0 \quad . \tag{17}$$

A positive (negative) trade-off for usage (i,j), user $(i,.)$, or facility $(.,j)$ means that this market segment is better (worse) off under actual pricing than under proportional to marginal cost pricing; i.e., this segment yields to telecommunications a smaller (larger) part of its surplus.

Table 9.2 shows the trade-offs per subscriber in 1981, calculated as mentioned above. (A qualitative analysis of these trade-offs may be found in Curien and Pautrat, 1983, and a quantitative one in de la Brunetiere and Curien, 1984.)

Table 9.2 Revenue trade-offs T^{ij} between users and between outputs (thousands of French francs, 1981 value).

	Type of user, i						
Type of output, j	*House-holds*	*Profes-sions*	*Indus-stry*	*Adminis-tration and services*	*All busi-nesses*	*Coin tele-phones*	*Total, $T^{\cdot j}$*
Connections	487	487	487	487	487	1556	498
Line operations	120	120	120	120	120	4261	161
Local traffic	122	101	605	1324	1158	976	205
Toll traffic	−512	−707	−8439	−3476	−4617	−4383	−864
Total, $T^{i\cdot}$	217	1	−7227	−1545	−2852	2410	0

5 Interpretation of Results

Figure 9.1 summarizes the direction of cross-subsidization between different classes of outputs in the French telecommunications industry. Toll traffic subsidizes local traffic owing to two factors. The cost of a local call depends on its duration, whereas the tariff does not and the cost of the average call exceeds the tariff per call; whereas, the actual cost of long-distance calls does not rise with distance as rapidly as does the tariff, which means overcharging of long-distance calls. This is becoming more pronounced with the spread of digital technology, which tends to reduce the dependence of cost on distance.

Second, traffic as a whole subsidizes lines operations and connections. The rental charge does not make up for the costs of lines and the operation of subscribers' equipment, nor does the connection charge make up for the true costs of access. Moreover, the social policy of expanding the telephone stock is widening this gap by lowering connection charges and keeping rental charges constant in current money. This policy will be discussed below.

Finally, although this is not shown explicitly in the table, off-peak traffic subsidizes peak traffic, and short local calls subsidize long ones (they are priced the same and cost less). Conversely, long toll calls subsidize short ones, because through the periodic pulse system the over-tariffing of toll calls is proportional to their duration. In order to cancel all trade-offs between outputs, the rental and connection charges should be roughly *doubled*, and so should the local pulse rate, whereas the toll pulse rate should be more than halved.

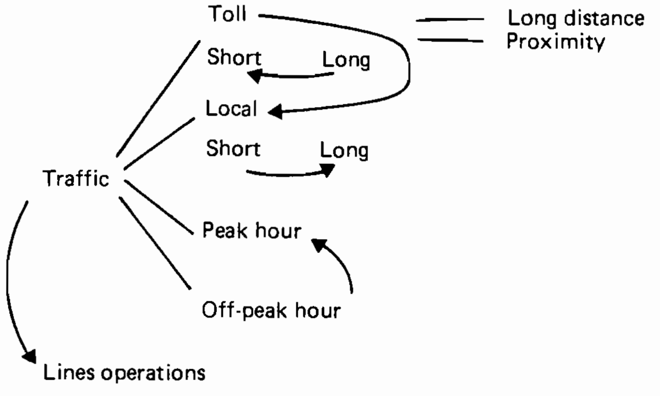

Figure 9.1

As for trade-offs between classes of users, the main feature is that businesses subsidize households (and coin telephones). More accurately, calculations show that under the present price structure, households are undercharged by about 15 percent and coin telephones by 30 percent, industry is over charged by 50 percent and services by 20 percent, tariffs being roughly fair for the professions. This trade-off from business to residential customers is the result of two parallel effects. On the one hand, these two categories account for a comparable share of the profits made on traffic, whereas households are the main source of new connections costs (in 1982 they accounted for 85 percent of the telephone stock increase). On the other hand, businesses make greater use of long-distance service, where profits are greatest.

In the future the traffic-to-access trade-off might be mitigated by the saturation of the residential market, and the development of telematics between businesses, which will tend to modify the relative contributions of households and businesses to network growth. Conversely, the intra-traffic cross-subsidies may become more important if an increase in network capacity in order to handle the residential evening traffic makes households rather than businesses responsible for peak demand.

Though not shown in Table 9.1, subsidies also occur from high-density areas to low-density ones, since installation and operating costs are greater in the latter case. However, this subsidy is partly offset by greater reliance on long-distance calls: due to sparser populations, a smaller number of subscribers can be called at the local rate than in urban areas.

6 Implications for policy

The above analysis of tariff–cost discrepancies and associated trade-offs can serve as a basis for designing a pricing policy for telecommunications. Such a policy has to seek for the "best" balance between two contradictory criteria. On the one hand, one wants to aim at economically efficient results – ideally through marginal pricing, more realistically through some form of second best, such as Ramsey pricing. On the other hand, it may be desirable to encourage the growth of certain outputs (such as nonvoice services) and to control the relative development of others. One might also wish to promote trade-offs between classes of users so as to favor access and usage by lower-income groups. The necessary compromise between economic pricing and trade-off control can be solved through two key options.

First, converging toward the relative structure of marginal costs in traffic facilities pricing, in order to guarantee a fair return on transmission and switching equipment and to steer demand towards efficient choices by reducing trade-offs between local and long-distance or peak and off-peak traffic. In this vein, time of day multilevel pricing and local measured service have just been introduced in France. Local measured service (for an economic analysis of local service pricing, see Mitchell, 1978, 1979, 1983; Brander and Spencer, 1983; Dansby, 1983) will remunerate the local network at proper time and cost, will allow the lowering of long-distance tariffs and thus reduce the effect of the distance factor on tariffs, and will up-grade the quality of service by shortening peak-hour calls. It will also encourage nonvoice uses of the network by charging in proportion to the volume of transmitted information. Time-of-day pricing, by offering attractive off-peak charges, could represent both a social advantage for households, as well as an incentive to the development of telematics, and by regulating the peaks, may improve the quality of supply without any increase in the average level of tariffs.

Second, subsidizing connections and lines operations from traffic revenues, while inconsistent with economic efficiency, may be partly justified by the positive externality of network growth – the new subscribers extending the potential available to those already subscribing (see Artle and Averous, 1973; Squire, 1973; Rohlfs, 1974; Littlechild, 1979; Taylor, 1980; Curien and Vilmin, 1983). Externalities associated with the growth of lines, both stock and usage, are mentioned by these authors as key factors in modeling telephone demand. Moreover, provided that the rate of return on investment in new connections exceeds the rate recommended to public investors by the central planning authority, the trade-off between traffic and access is a shift of revenue within the telecommunications industry which fits the macroeconomic social choice concerning the return on investment. Economic calculations show that the planned target of 23 million mainstations in 1985 (95 percent of households, 85 percent of workers equipped) can be reached by reducing connection and rental charges to about half of their 1980 values, which should preserve an acceptable rate of return on new connections (over 8 percent in real terms). An optimization of the French telephone network growth under tariff control and subject to profitability constraints is discussed in Curien *et al.*, 1981.

The evolution of the French telephone rate structure from its present state according to the options presented above (especially time-of-day pricing), will necessarily be a stepwise process in which the observation of consumer behavior and the measurement of price elasticities will allow a check on efficiency and show the direction for the next steps. Moreover, as mentioned in the Introduction, this evolution will have to cope with the development of nonvoice uses of the telephone network and with the growth of specialized networks parallel but interconnected to the basic one (datapack, satellite, cable TV, etc.). The overall purpose is to aim at a "consistent" pricing system that reconciles the various social and market development goals with economic efficiency.

References

Abraham, C. and Thomas, A. (1966) *Microeconomie: Decisions optimales dans l'entreprise et dans la nation* (Paris: Dunod).

Artle, R. and Averous, C. (1973) The telephone system as a public good: static and dynamic aspects. *Bell Journal of Economics and Management Science* 4.

Autin, C. and Leblanc, G. (1983) Empirical evaluation of cross-subsidy tests for Canadian interregional telecommunications network, in L. Courville, A. de Fontenay, and R. Dobel (Eds) *Economic Analysis of Telecommunications* (Amsterdam: North-Holland) pp 341–64.

Baumol, W.J. and Bradford, D.F. (1970) Optimal departures from marginal cost pricing. *American Economic Review* 60: 265–83.

Boiteux, M. (1971) On the management of public monopolies subject to budgetary constraints. *Journal of Economic Theory* 3: 219–40.

Brander, J.A. and Spencer, B.J. (1983) Local telephone pricing: two tariffs and price discrimination, in L. Courville *et al.* (Eds) *Economic Analysis of Telecommunications* (Amsterdam: North-Holland) pp 305–16.

de la Brunetiere, J. and Curien, N. (1984) Les transferts de revenus induits par la tarification telephonique entre categories d'abonnes et entre types de prestations, *Annales des Telecommunications* 39 (11–12).

Christensen, L.R., Cummings, D., and Schoech, P.E. (1983) Econometric estimation of scale economies in telecommunications, in L. Courville *et al.* (Eds) *Economic Analysis of Telecommunications* (Amsterdam, North-Holland) pp 27–53.

Curien, N., Gensollen, M., and Vilmin, E. (1981) Prévision de la demande téléphonique: besoins, instruments, procédures. La Documentation Française *Bulletin RCB*, 44: 2–23.

Curien, N. and Pautrat, C. (1983) An economic approach to telecommunications tariffs. *ITU Telecommunications Journal* 50: 183–94.

Curien, N. and Vilmin, E. (1983) Telephone demand and usage: a global residential model, in L. Courville *et al.* (Eds) *Economic Analysis of Telecommunications* (Amsterdam: North-Holland) pp 231–49.

Dansby, R.E. (1983) Economic analysis of a measured service option, in L. Courville *et al.* (Eds) *Economic Analysis of Telecommunications* (Amsterdam: North-Holland) pp 279–92.

Faulhaber, G.R. (1975) Cross-subsidization: pricing in public enterprises. *American Economic Review* 65: 966–77.

Fuss, M.A. (1983) A survey of recent results in the analysis of production conditions in telecommunications, in L. Courville *et al.* (Eds) *Economic Analysis of Telecommunications* (Amsterdam: North-Holland) pp 3–26.

Kiss, F., Karabadjian, S., and Lefebvre, B.J. (1983) Economies of scale and scope in Bell Canada, in L. Courville *et al.*, (Eds) *Economic Analysis of Telecommunications* (Amsterdam: North-Holland) pp 55–82.

Littlechild, S.C. (1970) A game theoretic approach to public utility pricing. *Western Economic Journal* 8(2): 162–66.

Littlechild, S.C. (1979) *Elements of Telecommunications Economics* (London: Institute of Electrical Engineers).

Mirman, L.J. and Tauman, Y. (1982) Demand compatible equitable cost sharing prices. *Mathematics of Operations Research* 7(1): 40–56.

Mitchell, B.M. (1978) Optimal pricing of local telephone service. *American Economic Review* 64(4): 517–37.

Mitchell, B.M. (1979) *Economic Aspects of Measured Service Telephone Pricing.* Rand Memorandum P.6321, March 1979 (Santa Monica, CA: Rand Corporation).

Mitchell, B.M. (1983) Local telephone costs and the design of rate structures, in L. Courville *et al.* (Eds) *Economic Analysis of Telecommunications* (Amsterdam: North-Holland) pp 293–304.

Ramsey, F.P. (1927) A contribution to the theory of taxation. *Economic Journal* 37: 47–61.

Rheaume, G.C. (1983) Welfare optimal, subsidy free prices under a regulated monopoly, in L. Courville *et al.* (Eds) *Economic Analysis of Telecommunications* (Amsterdam: North-Holland) pp 319–40.

Rohlfs, J. (1974) A theory of interdependent demand for a communication service. *Bell Journal of Economics and Management Science* 5(1): 16–37.

Squire, L. (1973) Some aspects of optimal pricing for telecommunications. *Bell Journal of Economics and Management Science* 4(2): 515–25.

Taylor, A.L. (1980) *Telecommunications Demand: A Survey and Critique* (Cambridge, MA: Ballinger).

Thomas, A.L. (1980) *A Behavioral Analysis of Joint Cost Allocation and Transfer Pricing.* (Urbana, IL: Stipes).

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 10

AUCTIONING LANDING RIGHTS AT CONGESTED AIRPORTS

Michel L. Balinski and Francis M. Sand

1 Introduction

The intent of the United States Airline Deregulation Act of 1978 was to open the industry to the usual market forces, and, in particular, to allow air carriers to initiate or terminate service wherever they liked. Airports, however, have a fixed capacity. Due to congestion the FAA (Federal Aviation Administration) established quotas in 1969 on the hourly number of scheduled landings and take-offs (called "slots"). The 1978 law was passed before anyone became aware of the need for a method to allocate the available slots among competing carriers that is in the spirit of the Act.

This chapter describes a mechanism that was developed in response to the need for finding an economically efficient allocation of slots that explicitly accounts for the complex interdependence among the slots that airlines require for scheduling flights. It is an auctioning procedure in which repeated bidding is done simultaneously across the many different hours and airports that constitute the "markets". It is also a concrete instance of an auctioning mechanism for allocating public rights of access that have joint benefits (and possibly joint costs).

This approach to allocation was developed and proposed in the context of the 1980 airline situation, so the facts reported below date from then and the perceptions of what was realistically acceptable are reflected in the work. In fact, scheduling committees (as described below) have continued to be the instance of allocation through October 1984, at least, although under increasing attack and the strain of prolonged scheduling deadlocks.

2 Background

A *slot* is the right to schedule either a landing or a take-off in a particular hourly period (e.g., 8 to 9 a.m.) at a particular airport. To schedule flights the airlines must have slots, but slots must not be confused with actual landings or take-offs which are, of course, afforded to planes in distress, emergencies, etc., and for which landing fees and other rental charges will exist.

Standard practice is for each airline to determine, in response to its perceived market opportunities, a half-year ***fleet schedule***, that is, a schedule for each aircraft in its fleet, in set cycles through different airports. These cycles are determined by customer demand and the need to assure periodic maintenance of planes. For example, each of United Airline's DC-10s had its own 37-day cycle involving about 410 hours in flight, including San Francisco at least once for an overnight stay to perform a major maintenance. Finding a good fleet schedule is an extremely complex combinatorial optimization problem that demands a major planning effort. Given a good schedule, that is, one that makes efficient use of the fleet in meeting expected demand for flights, an airline is loath to change it. Once determined, crew routing is fitted to the fleet schedule. The fleet schedules decide the slots each airline seeks to acquire. In a word, fleet schedules, and the flights that they assure, are the essential objects of importance and their values determine the values of individual slots. There is, therefore, an extremely strong interdependency among any prices that may become attached to slots.

Airline "scheduling committees" meet twice a year under the auspices of the Air Transport Association. They are composed of airline industry representatives operating under a special antitrust exemption sanctioned by the Civil Aeronautics Board. Each committee is concerned with a single airport and seeks to obtain agreement on an allocation of slots among the air carriers that satisfies set "quotas" at high-density airports. These are: 40 at Washington National, 48 at New York LaGuardia, 70 to 80 at New York JFK and 115 at Chicago O'Hare (depending on the time of day). Slots are allocated for a fixed period of 6 months. Prior to each meeting and based on their fleet schedules, the carriers submit their requests for slots to the scheduling committee staff. These requests are tabulated and distributed to members at the start of the meeting. Slots requested generally exceed slots available. Through negotiation, the committee first seeks to reduce total requests by all carriers to the daily quota (for example 640 at Washington National Airport). Then they try to limit hourly reservation requests to hourly quotas by having the carriers "slide" their operations from one hour to a neighboring one (airlines thereby hope to maintain the basic integrity of their fleet schedules). Under the guidelines set down in the CAB antitrust exemption, committee representatives are not permitted to discuss operations with respect to particular markets, routes or other airports. Should a committee fail to reach an agreement, the responsibility for a decision rests with the FAA. Although in the past the scheduling committees have typically reached a unanimous agreement, the flexibility airlines have acquired in entering new markets, due to the Deregulation Act, has considerably lengthened the process. Previously, the road to obtaining unanimous agreements was for all to agree to a proportional reduction in slot requests. Experience has shown that the new, smaller entrants are less willing than established carriers to agree to such proportional reductions. Since the rule for agreement is unanimity, it is not surprising that arriving at decisions in the deregulated environment has become much harder.

3 The Problem

The problem of the allocation of slots is fundamentally rooted in the efficient operation of an airport. An early paper concluded, "... the existing price system for airport services fails to allocate the existing capacity so as to maximize its value. It fails also to guide investment in airports so as to achieve the appropriate mix and level of output with a minimum investment of resources" (Levine, 1969). One might wish for an airport pricing system that adequately reflects the multiple attributes of airport capacities, such as different types of pollution, limits, rights of access, terminal space and the like, and also reflects user characteristics such as the type of plane, its weight, and the time of day. It is neither reasonable nor politically viable, however, to consider a comprehensive pricing policy.

Accordingly, we view slots at a given hour and airport as a homogeneous good representing, in effect, an option giving the holder the right to schedule an operation at that hour and airport for a six-month period. There are good reasons that such rights should be vested for sufficiently long periods of time, since the services provided by carriers require investments in support facilities, advertisement and the like which cannot be altered in the short run. In addition, since the demand for air travel, the financial positions of carriers, the general state of the industry and the condition of the economy as a whole may change, the holders of slot options in a deregulated environment should be allowed to trade – to buy and/or to sell – their options. The problem thus becomes: to design competitive and efficient mechanisms to first allot and then facilitate the trading of slots.

The various approaches to the initial allocation of slots that have been considered may be classified into administered and competitive schemes (see, for example, Geisinger, 1980). Administered schemes include: the current scheduling committees; first-come, first-served; proportional allotments by priority guidelines (depending upon such factors as historic shares of operations, potential public service provided); lotteries (with win probabilities a function of historic shares, estimated future passengers, or on the basis of purchased lots); solutions of mathematical programming problems that maximize the total "value" of all scheduled flights; and time-differentiated landing fees.

An extensive study of the scheduling committee (Grether *et al.*, 1979) demonstrates the economic inefficiency of its outcomes. The committee's requirement for unanimous decisions gives each participant, notably the least viable economically, the power of veto, so factors other than economic ones determine the outcomes.

Maximizing the total value of the carrier's total scheduled flight is a laudable aim, but how can such values be determined? The basic planning decisions involve fleet schedules, not flights. Moreover, even if precise values could be imputed, airlines would probably not be willing to divulge them.

Time-differentiated landing fees seem to be an attractive alternative (Carlin and Park, 1970). They have been used by the British Airports Authority and were applied, in limited form, by the Port Authority of New York beginning in 1968 to relieve congestion by discouraging smaller aircraft. Unfortunately it is impossible in practice to set such fees correctly: if prices are too low, excess demand results and the problem remains; if prices are too high, excess supply results and the airport is under utilized, contrary to the public welfare. If some prices are too high and others too low, both conditions obtain. It should be feasible to incorporate sufficiently modest time-differentiated fees to avoid causing underutilization, and then rely on another, competitive mechanism for solving the allocation problem.

Competitive allocation schemes connote auctions and/or market mechanisms. If auctions, what type should they be? If markets, how are trades to be effected? There are a host of specialized auction rules, each apparently tailored to the needs and institutional environment of the industry where they are used (Cassady, 1967). The English auction begins with low bids to buy and goes up, the last (highest) bidder winning. The Dutch auction begins with high offers to sell and goes down, the first bidder winning. Should bids be open (e.g., oral) or sealed? How should many homogeneous goods (40 slots per hour at Washington National) be auctioned as versus one specified item (e.g., a painting)? For example, it has been suggested by Vickrey (1972) that for the problem of a flight overbooked with passengers, each passenger should indicate on a sealed "bid" that amount of money he or she would be willing to accept for not flying; the passengers making the highest bids would be given seats, the remainder would receive compensation equal to the price on the highest rejected bid. This is a competitive sealed-bid auction in that all bids are filled at one price. Should the auction be discriminatory sealed-bid, with bids filled at the different full-bid prices? In an exchange market, should offers to buy and offers to sell be anonymous? Should direct negotiation be allowed between participants?

To define the terms of the discussion assume that each airline has a utility function and that given a set of prices each airline buys and/or sells slots to maximize its utility subject to its budgetary constraints. Let us call these "optimal individual plans". A ***competitive equilibrium*** solution is a set of prices and a set of corresponding optimal individual plans which are together feasible, that is, at which demand does not exceed the available of supply slots. A competitive equilibrium is "efficient" if there is no other solution whose utility is at least as great for all participants and strictly greater for at least one; that is, there is no solution that dominates in the sense that each airline is at least as well off according to its own evaluation and at least one is better off. The central problem is to elicit a set of equilibrium prices.

The difficulty with almost all auction mechanisms is that they ignore the interdependency problem. Acquiring one slot is of no use alone, since it takes at least two slots to link a pair of congested airports. One way to overcome this is a "first-choice" auction. In barest terms, it proceeds as follows: at any stage, an ordinary auction is held (say, of the English variety) with the winner having the right to choose any slot at any airport which is still available. The interdependency problem appears to be in part accommodated: the bidders can acquire in sequence the slots necessary to complete flights and fleet schedules. The solution is freely competitive, but it is not efficient. Moreover, the prices paid by bidders for slots in the same hour at the same airport may differ widely, and so do not constitute a set of nondiscriminatory equilibrium prices. In addition it is difficult to imagine the industry agreeing to pay significantly different prices for what it considers an identical good.

A direct approach, similar to that suggested by Vickrey (1972) for resolving overbooking problems, was proposed for slot allocation by Grether ***et al.*** Each airline interested in acquiring slots in some particular "market", for example the 5-6 p.m. period at Washington National, makes sealed bids for each of the slots it desires, perhaps at different prices. The sealed bids are collected, and the 40 highest bids are awarded the 40 available slots at the price of the lowest of those 40 bids. In general, in a market with an hourly quota of q, the q highest bids win at the price of the lowest of the q bids. If the total number of bids is lower than q, then all are awarded at no cost (there is excess supply in the market). This "one-time" auction is, in many ways, a reasonable approach. However, it fails to take into account the institutional problem faced by the industry.

Assume, for the purposes of concrete discussion, that the following situation obtains: 4 congested airports have hourly quotas which are in force for certain hours of the day. This means there are some 40 slot-markets. Either the many auctions operate simultaneously or they operate sequentially.

Suppose that the auctions are conducted simultaneously. Then each of the participating airlines must place values on each of the slots it desires over the entire network and at all times. What strategy should an airline follow? Suppose an airline knows what slots it wishes and knows the values it attaches to flights. How should the airline allocate the *values* it has on flights to the *prices* it bids for slots desired to realize those flights? For a flight needing only one slot, there is no difficulty. But for a flight needing two slots (or more) there are many ways in which the airline could distribute the value over the two (or more) slots. What it *should* bid depends entirely upon the total demand pressure in each of the markets. The "one-time" auction provides no information concerning this pressure since the airlines are obliged to bid "in the dark". The intertemporal and network dependencies so crucial to the industry's environment are unaccounted for. Nothing remotely resembling an efficient economic equilibrium is likely to be produced: the bids can only be pure guesses.

Suppose, then, that the auctions are conducted sequentially. Then, the first auction, say 7:00-8:00 a.m. at Washington National, is carried out "in the dark"; the second, say 8:00-9:00 a.m. at Washington National, is carried out with the outcome of the first known; ... the last, say 10:00-11:00 p.m. at Chicago O'Hare, is carried out with the outcomes of all the prior auctions known. The strategy of each participant may be increasingly clarified; but early mistakes cannot be recouped, and the later bids, and so outcomes, must depend upon the earlier results. As is well known to auctioneers, and to readers of the auction literature, "... both the allocations determined by auctions and the final vector of prices associated with these allocations are very sensitive to the sequence in which the goods are brought up for sale" (Schotter, 1976). Thus, beginning with Washington National and going on to Chicago O'Hare may yield very different results than *vice versa*; beginning with the morning hours and going on to the afternoon hours may give different prices and allocations than beginning with the afternoon periods and proceeding to the earlier hours of the day. Economically efficient allocations cannot be expected to result from sequential auctions either. Moreover, sequential auctions may invite the bidders to make threats concerning future bids that can lock them into inefficient strategies and uneconomic outcomes. Numerical examples can be constructed to show "prisoner's dilemma" phenomena in which no airline can adopt a more rational behavior alone because doing so would only worsen its situation. Declared threats could be outlawed as a rule of procedure, yet the fact of such possibilities points to an additional weakness.

In either case – simultaneous or sequential – inefficient allocations can result. In order to rectify this possibility and to facilitate the coordination of slots to meet the needs of flights, Grether *et al.* (1979) proposed an aftermarket. The potential unbalanced endowments of slots resulting from the initial "one-time" auctions would be corrected in an aftermarket of the NASDAQ, "open-book" type where carriers could sell or buy slots. These markets can, indeed, be expected to converge to competitive equilibrium prices: but competitive prices and trades which occur are directly dependent upon the initial endowments which the traders bring with them, and these initial endowments are the result of the previous rounds of auctions.

The one-time auction mechanism, in which airlines bid totally unaware of the demand pressures in one or another market, can lead to almost any outcome. The aftermarket produces a competitive equilibrium based upon the initial slot allocation. What of the joint outcome? It will not, in general, be an efficient competitive equilibrium. The outcomes of the auctions may so distort initial distributions that the aftermarkets are unable to achieve competitive equilibria efficient for the final allocation of slots. An example demonstrating how this may occur is given in the Appendix. The difficulty is that in the auction process each airline bids prices on slots whose totals equal (or approximate) the values of the corresponding flights it wishes to schedule, but, being ignorant of the remaining airline's demands and bids, it receives only some of the necessary slots. It – and the other airlines – enter the aftermarket to acquire needed slots and dispose of slots for which the prices of the mates they would need to complete a flight are too high (the total value of the flight would be surpassed). Economic self-interest results in some trades, but not in sufficient trades to drive out certain flights that should not belong to a competitive, efficient solution. The prices and allocations of the initial auction inhibit the aftermarket from producing an efficient solution.

4 Repeated Auctions

Instead of a "one-time" auction consider a "repeated" auction. The idea is to permit airlines to iteratively learn of the demand pressures that exist in each slot market and so to enable them to develop informed and rational strategies in view of the demands and the interdependencies; and to obtain auction allocations that are at least close to efficient competitive equilibria. The auction is followed by an aftermarket designed not only to permit trading in slots throughout the period of 6 months in which options are valid, but also to permit the minor marginal adjustments that may be necessary to improve the initial allocation reached by the repeated auction.

Each airline comes to the initial round of the repeated auction with its desired schedules in hand and some appreciation for the total expenditure (the "value") it is prepared to make to realize a flight or scheduling cycle. Each airline is requested to prepare sealed bids for the slots that it requires. This means that for each hour at each congested airport – at each "trading post" – it prepares bids for those slots it desires. These first bids are made "in the dark" just as in the "one-time" auction. The airlines should bid any prices for slots realizing a flight whose totals do not surpass the value it attaches to the flight. The individual demand curve of an airline is as given on the left of Figure 10.1. Accumulating the individual demands yields the aggregate demand of the carriers at each post. A typical aggregate demand curve is given on the right of Figure 10.1.

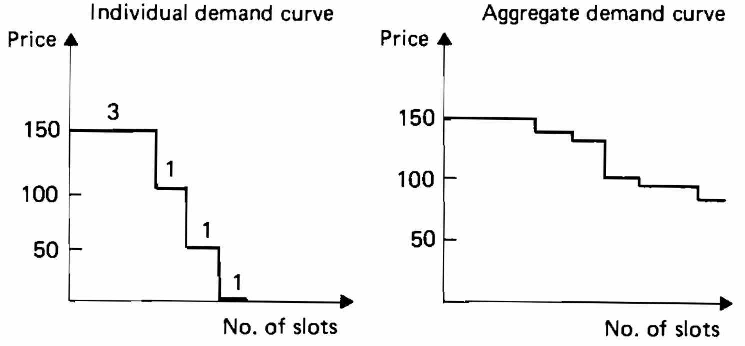

Figure 10.1

Either the total demand is less than the quota of that trading post (as on the left of Figure 10.2) or the total demand is at least as great as the quota (as on the right of Figure 10.2). The dotted line in effect represents the supply curve, so the trading post price should be determined by where demand and supply intersect. In the case of excess supply (Figure 10.2 left) the price should be 0; in the case of excess demand (Figure 10.2 right) the price should lie in the interval between the price of the qth highest bid and the $(q+1)$th highest bid. It may be that an aggregate demand curve is as shown in Figure 10.3, where there are several bids at the trading-post price.

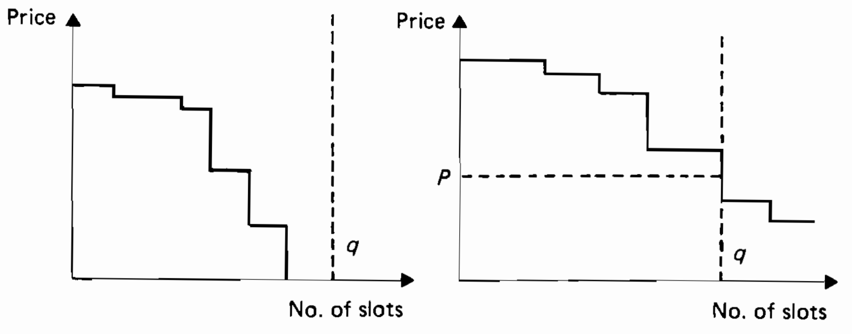

Figure 10.2

At this point the auction administrator reveals to all airlines the aggregate demand at each of the trading posts. Each of the airlines who made bids at or above the trading post price p (which makes supply equal to demand) is *conditionally* awarded those slots, and each of those having made bids lower than p is not. In the case of Figure 10.3, where more than one bid is made at the trading post price p but the quota is such that not all bids at that price or higher can be awarded, then some random allocation among those bidding p is made. If this were the final auction, then each airline winning a slot at a trading post would actually pay the price p at that post.

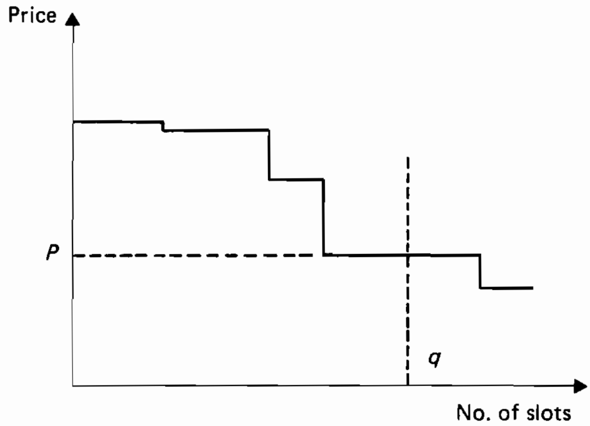

Figure 10.3

If the conditional allocations and prices are agreeable to all airlines then an efficient, competitive equilibrium allocation has been found. Typically – and certainly after the first round of bidding – many airlines will be dissatisfied with the conditional solution. The results of the first round are precisely equivalent to what is produced by the "one-time" auction and so its defects are known.

This is why the administrator announces *conditional* allocations and trading post prices, and the current expression of total demand at each trading post. The first round gives each bidder information concerning the demand pressures that exist at each trading post. On the basis of these, and of their individual needs, each airline prepares new bids in a second round of the auction. The administrator accumulates the sealed, secret individual bids, and by the same procedure, announces new conditional allocations, trading prices and total demands. If, at any step, no airline announces the wish to change its bid, then the process terminates. Since it is unrealistic to assume that convergence so perfect will obtain, other stopping rules may need to be invoked (see below).

The outcome of the slot exchange auction is that each airline is endowed with the ownership of a collection of slot options. Typically, the auction procedure would not result in a perfect equilibrium: some airlines would wish to acquire several slots, some to dispose of several. And, as the six-month period of vested rights elapses, the desire to acquire more slots or to dispose of more, could develop. Therefore, a NASDAQ "open-book" aftermarket would be maintained continuously until the expiration of the six-month period. An airline would be free to offer slots for sale at stated prices or to bid to purchase slots at stated prices at each of the trading posts at any time. Each trading post would have a total demand D and total supply S which could be in any one of the four different qualitative forms shown in Figure 10.4. In case (d) the offers to sell are all at prices above the offers to buy, so no exchanges would take place. Otherwise, in cases (a), (b) and (c), q^* slots are bought and sold at some trading post price p^* that lies between p_D^* and p_S^*, $p_S^* \leq p^* \leq p_D^*$. In these cases all sellers who announce a price higher than p^* would sell nothing. All buyers who quoted a lower price would buy nothing. If there was excess demand at p^* case (a) with, say,

$p^* = p_D^* = p_S^*$ or excess supply at p^* case (b) with, say, $p^* = p_S^* = p_D^*$ then, those who bid p would be rationed randomly.

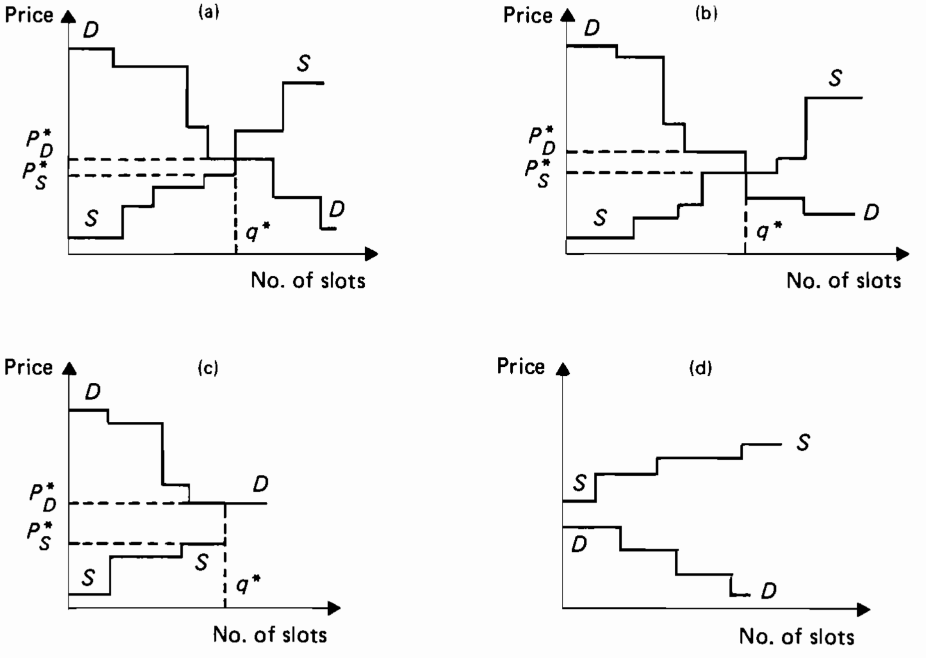

Figure 10.4

The repeated auction and aftermarket embody precisely the same economic mechanism: the law of supply and demand determines the equilibrium solutions in both cases. Thus, the same arguments sustain the relevance of both mechanisms. In fact, there is little controversy over the open-book NASDAQ-type slot exchange: this is a well practiced form of market.

Partial theoretical justification for repeated auctions may be based on a paper of Dubey (1982), which considers a Walrasian exchange economy with a finite number of traders and of commodities, with traders making contingent statements about the quantities they are prepared to buy or sell at different prices. He shows, under very general conditions, that the competitive equilibria are efficient and at the same time are strong Nash or noncooperative equilibria. A "Nash" or "noncooperative" equilibrium means that no airline acting alone could improve its position by changing its bid. A "strong" Nash equilibrium means that no cabal of airlines acting in concert could improve their positions by changing their bids, in the sense that at least one airline is better off and none are worse off. This implies that there would be no inducement for one or several airlines to posture.

Dubey's model explicitly includes the interdependency of the commodities, that is, of the slots at different trading posts. It does, however, make several assumptions that are not strictly met by the slot allocation problem. The marginal values of commodities are assumed to be non-increasing. This may be reasonable for slots, except for the temptation of cornering a market as more are acquired. The commodities are assumed to be divisible. Slots are not; however, there are sufficient numbers of them (e.g., 40) to justify the use of the results. The model is of an exchange economy so the initial "owners" of the slots – the airports or the FAA – should be considered as strategic players: setting an hourly quota q is equivalent to announcing a supply curve that offers up to q slots at price 0 but no more at any price. However, the repeated auction mechanism does not allow the airports to participate in the trading by announcing supply curves. The procedure itself is, of course, easily modified to include the airports as bidders. Airports would announce non-decreasing supply curves (such as in Figure 10.4) and thus adopt supply strategies, somewhat relaxing the absolutely fixed nature of the quota limitations by introducing extra slots at higher prices. This could serve to answer the airlines frequently voiced wish for the quotas to be increased. The difficulty is that strategic behavior – maximizing airport revenues – could be inimical to the public welfare, although that would depend upon many other factors.

The repeated auction is a *tâtonnement* process: the idea is to try solutions and have participants react by bidding higher for slots that are desired but so far denied and to drop from bidding for slots that are too dear. There can be no mathematical guarantee that this process will converge: examples prove that in some cases it will not. However, experimental work of Vernon Smith (1976) and others shows that, in many open-market bidding situations, remarkably fast convergence to a competitive equilibrium is obtained. This typically happens in the presence of very little information. "The experimental facts are that no double auction trader needs to know ***anything*** about the valuation conditions of other traders, or have ***any*** understanding or knowledge of market supply and demand conditions, or have any trading experience (although experience may speed convergence) ..." (Smith, 1976). Smith's laboratory experiments concern very simple trading situations. The repeated slot auctions are considerably more complex, and there have been to date no experiments that adequately treat the possibility of interdependence in the goods that are traded. However, the repeated auctions build information concerning the values others put on goods and cumulatively reveal the total demand pressures at the various trading posts. One would expect that prices and allocations would begin to settle down over successive outcomes with changes occurring only at the margins. However, the indivisible nature of slots could well cause the prices of certain ones to be chased up to the point that many bidders would drop out together, only to bid up other prices. Such a cyclic phenomenon, due to the indivisibility of the commodity, could warrant the introduction of "threshold" prices or "entrance fees". Other candidates for rules to induce convergence are to stop auctioning when prices change by little, or when allocations change by little, or when only a few of the bidders wish to change their bids. In any case the threat of no further auctions after several rounds should always be present to prevent posturing or other disruptive behavior. It could also be useful to charge bidders for the privilege of changing their bids.

While a carefully designed experiment using auctions to obtain an initial slot allocation remains to be performed, the procedure has been tested for manageability and convenience in an institutional setting (Balinski and Sand, 1980; Bechhofer, 1980). The test was conducted for the FAA over the course of one week with airline representatives as players, using an airline management game (Elias, 1979), which provided a quasi-realistic environment for the participants. Slot values were established in terms of the short-term profitability of airline operations with a fixed fleet of aircraft and simulated markets defined in terms of seat-mile capacity, schedule convenience and frequency, and a generalized passenger demand function. Five "airline teams" competed for market shares in the game.

The results were far from conclusive for several reasons. First, the game was complicated and required a significant effort to learn. Within the time constraints, it was not possible to fully distinguish between learning effects and the convergence of the auction procedure. Second, there were no controls in the experimental design. Third, the two repeated auctions were terminated after three rounds of bidding in each, in spite of the obvious need for further rounds, as seen from the volatile price movements. However, the experiment did show that airlines could use the repeated auction procedure to obtain a reasonable initial slot allocation. Most of the slots which the "airlines" required for flight scheduling were in fact obtained, and at prices which diminished but did not destroy profits. In both auctions there was a remarkable degree of convergence in allocations: the third rounds produced almost the same allocations as the second rounds, despite considerable price variability. Moreover, the second auction – which concerned the same simulated problem – showed significantly better convergence than the first. Learning clearly made for considerable improvement. Some of this may be due to the fact that the participants could easily guess that the third round would be the last, and so would wish to assure their "current" positions. It is fair to conclude that the experiment contributed to showing the practical interest of repeated auctions but at the same time confirmed the need for further trials and study.

5 Criticisms

The criticisms leveled at repeated auctions by industry representatives and others seem to fall into four major headings: what to do with the proceeds, how to implement the auctions in practice, how to obtain convergence, and how equitable they are to small as opposed to large airlines.

The predominant worry had to do with the monies generated. As one observer of the experiment remarked, "Several representatives of the airports attended the experiments, and their presence compounded the scheduler's fears. The airport men watched eagerly as total revenues from slots climbed higher and higher" (Bechhofer, 1980). This is a question inherent in any pricing system, not to auctions alone, which raises economic and political issues as to who controls the funds and for what purpose. While this issue is beyond the scope of our study, it may well be that any pricing system for the allocation of slots will founder on this point.

The fear that small carriers may not be treated equitably is founded on the idea that the large carriers' financial might could lock them out of any slots. This seems to be something of a red herring since it is an issue in any market whatsoever. Moreover, the behavior of scheduling committees seems to indicate that the large airlines have bowed to the few demands for slots made by the small to avoid trouble.

Time was the next major worry. Preparing bids in successive rounds in response to information gleaned in previous rounds was difficult even in the simplified experimental setting. Computerized communications to facilitate the physical realization of auctions would help, but one would have to count on learning and experience to eventually allow bidders to respond relatively quickly. Further testing would clearly be necessary.

The time needed to converge to a "near" equilibrium solution is a practical as well as a theoretical issue. The practical evidence in many other domains encourages a faith in convergence despite the existence of theoretical examples that shows it does not necessarily obtain. Further testing of different stopping rules and devices that promote convergence is needed.

Appendix. Example of an inefficient outcome with "one-time" auction and aftermarket.

There are five airlines (A, B, C, D and E) and one congested airport. The quota of the airport is 4 slots. Each airline wishes to route flights through the airport at particular times of the day, first landing then taking off. Each airline knows the "value" of each flight, which represents the total dollar amount it is willing to pay to acquire the necessary slots. The data of the problem are as follows. The notation X indicates the slots needed to realize the flights in question.

			Time (a.m.)		
Airline	*Flight*	*Value*	*9–10*	*10–11*	*11–12*
A	A1	305	X	X	
	A2	305		X	X
B	B1	320	XX		
	B2	305	X	X	
	B3	295		X	X
C	C1	302	XX		
	C2	320		XX	
	C3	300		X	X
D	D1	300	X	X	
	D2	305		X	X
E	E1	324			XX

C2, for example, is to land and take-off in the 10–11 a.m. period; D2 to land in the same period and take off in the 11–12 a.m. period.

Suppose that the (simultaneous) "one-time" bidding is as shown in the next table.

Then, the price in the first period (9–10 a.m.) is $157, in the second period $145, and in the third period is $162. The winners of slots are starred (*). Thus, A wins one slot in the first period, and one in the third; B wins two in the first and two in the second; C wins two in the second and one in the third; D wins one in the first and one in the third; and E wins one in the third period. The initial endowments are unbalanced. Most flights cannot be realized.

Consider A. It paid $157 for one slot at 9 a.m. for A1, and so is prepared to pay at most $305–$157 = $148 for a slot at 10 a.m. B paid $145 for a slot at 10 a.m.

			Time (a.m.)		
Airline	*Flight*	*Value*	*9–10*	*10–11*	*11–12*
A	A1	305	$170*	$135	
	A2	305		$140	$165*
B	B1	320	$160*,$160*		
	B2	305	$150	$155*	
	B3	295		$145*	$150
C	C1	302	$151,$151		
	C2	320		$160*,$160*	
	C3	300		$136	$164*
D	D1	300	$157*	$143	
	D2	305		$135	$170*
E	E1	324			$162*,$162

needs a slot at 11 a.m. to accommodate its flight B3, but notes the price, $162, is too high since $162 + $145 = $307, which is more than its value. So, B is willing to sell one of its second-period slots for at least $145. B, therefore, sells one of these slots to A at some price between $145 and $148, say $146.50. Similarly, B is willing to offer up to $160 for a first-period slot and D to sell a first-period slot for at least $157. So B purchases such a slot from D for, say, $158.50. Similarly E purchases a third-period slot from A at $162.

At this point, A has acquired the slots for A1; B has acquired the slots for B1 and B2; C has acquired the slots for C2; and E has acquired the slots for E1. C and D each hold one 11–12 a.m. slot. No airline is willing to make any departures from its present holdings: the solution is a Nash equilibrium.

However, the solution is inefficient. C and D cannot use their slots since they do not possess the corresponding slots necessary to complete the flights. The markets for first- and second-period are saturated, but the third period has only 2 flights, not the quota of 4. The total "value" of the 5 realized flights in $1574. The total paid at the auctions is $1832.

There does exist an efficient, competitive equilibrium determined by the following prices: $151 in the 9 a.m. market; $155 in the 10 a.m. market; and $150 in the 11 a.m. market. At these prices, A obtains the slots for A2; B obtains the slots for B1; C obtains the slots for C1 and C2; D obtains the slots for D2; and E obtains the slots for E1. No other flight is economically viable. The quota of each period is saturated. The total value of the 6 realized flights is $1874. The total sum paid at the prices named is $1824.

Acknowledgments

This paper summarizes work (Balinski and Sand, 1980) that was done for the Office of Aviation Policy of the Federal Aviation Administration by Econ, Inc., under contract DOT FA79WA-4374.

We are pleased to acknowledge the participation of David Bechhofer, Pradeep Dubey, Jai Jaikumar, Peter Stevenson and Robert Weber in this project. It is also a pleasure to express our thanks to John Rodgers of the Office of Aviation Policy for his help and interest. Finally, thanks are due to H.P. Young for asking us to write an account of the idea and the work to be included in this volume.

References

Balinski, M.L. and Sand, F.M. (1980) *The Allocation of Runway Slots by Auction*. Report FAA-AVP-80-3 (Washington, DC: Federal Aviation Administration).

Bechhofer, D. (1980) *The Airport Runway Slot Auction: Examination and Evaluation*. Yale University, Senior Project.

Carlin, A. and Park, R.E. (1970) Marginal cost pricing of airport runway capacity. *American Economic Review* 60(3): 310–19.

Cassady, R. Jr. (1967) *Auctions and Auctioneering* (Berkeley: University of California Press).

Dubey, P. (1982) Price-quantity strategic market games. *Econometrica* 50: 111–26.

Elias, A.L. (1979) *The Development of an Operational Game for the US Domestic Airline Industry*. Flight Transportation Laboratory, FTL Report RT8-5 (Cambridge, MA: Massachusetts Institute of Technology).

Federal Aviation Administration (1976) *Airport Quotas and Peak-Hour Pricing: Theory and Practice*. Report FAA-AVP-5 (Washington, DC: FAA).

Federal Aviation Administration (1979) *Airport Deregulation and Airport Access: Definition of Issues and Impacts of Regulatory Reform*. Report AVP–200 (Washington, DC: Office of Aviation Policy).

Federal Aviation Administration (1979) *Issue Paper: Allocation of Runway Capacity, Washington National Airport* (Washington, DC: Office of Aviation Policy).

Geisinger, K.E. (1980) *A Method for Administrative Assignment of Airport Runway Slots*. Report FAA-AVP-80-5 (Washington, DC: FAA).

Grether, D.M., Isaac, R.M., and Plott, C.R. (1979) *Alternative Methods of Allocating Airport Slots: Performance and Estimation* (Polinomics Research Laboratories, Inc).

Levine, M.E. (1969) Landing fees and the airport congestion problem. *Journal of Law and Economics* 12: 79–108.

Plott, C.R. and Smith, V.L. (1975) *An Experimental Examination of Two Exchange Institutions* (Pasadena, CA: California Institute of Technology).

Schotter, A. (1976) Auctions and economic theory, in Y. Amihud (Ed) *Bidding and Auctioning for Procurement and Allocation* (New York: New York University Press).

Smith, V.L. (1976) Bidding and auctioning institutions: experimental results, in Y. Amihud (Ed) *Bidding and Auctioning for Procurement and Allocation* (New York: New York University Press).

Vickrey, W. (1972) Airline overbooking: some further solutions. *Journal of Transport Economics and Policy* Sept: 257–83.

COST ALLOCATION: Methods, Principles, Applications
H. Peyton Young (Editor)
Elsevier Science Publishers B.V. (North-Holland)

CHAPTER 11

COST ALLOCATION IN MULTIPURPOSE RESERVOIR DEVELOPMENT: THE JAPANESE EXPERIENCE

Norio Okada

1 Introduction

It was not until 1953 that an official cost allocation procedure was established for multipurpose reservoir development in Japan. In that year the Energy Power Development Promotion Act was passed and the so-called alternative justifiable expenditure (AJE) method was officially adopted as the procedure for allocating costs. An increasing number of multipurpose reservoir developments, involving principally three sectors – power, river management (chiefly flood control) and irrigation – demanded a new law specific to this type of water resources development. In 1957 the Specified Multipurpose Reservoir Development Act was enacted, which authorized the Ministry of Construction exclusively to plan, implement and maintain particular types of multipurpose reservoir developments. "Dam utilization rights" were established under which users were regarded as "tenants" of a dam and were entitled to "utilize" the dam instead of "owning" the facility. The 1957 Act carried forward the AJE method as the official cost allocation procedure.

In the early 1960s a remarkable upsurge in the economy caused domestic and industrial water demands to pick up sharply. This economic growth was accompanied by accelerated urbanization, which was gradually encroaching on floodplains and exposing residents to flood damage. All of this meant an ever-growing demand for multipurpose reservoir developments, including flood control, hydroelectric power, and increased industrial and municipal water supplies.

In practice the AJE method was found not to be appropriate in many of these cases. Typically the AJE method tended to favor the power sector, in that its justifiable cost (i.e., the alternative cost of supply) was assumed to be low (meaning that it was allocated a relatively small share of the joint costs), whereas its inclusion in the development often caused total costs to increase sharply, which then had to be borne in part by the other water-using sectors. This treatment was

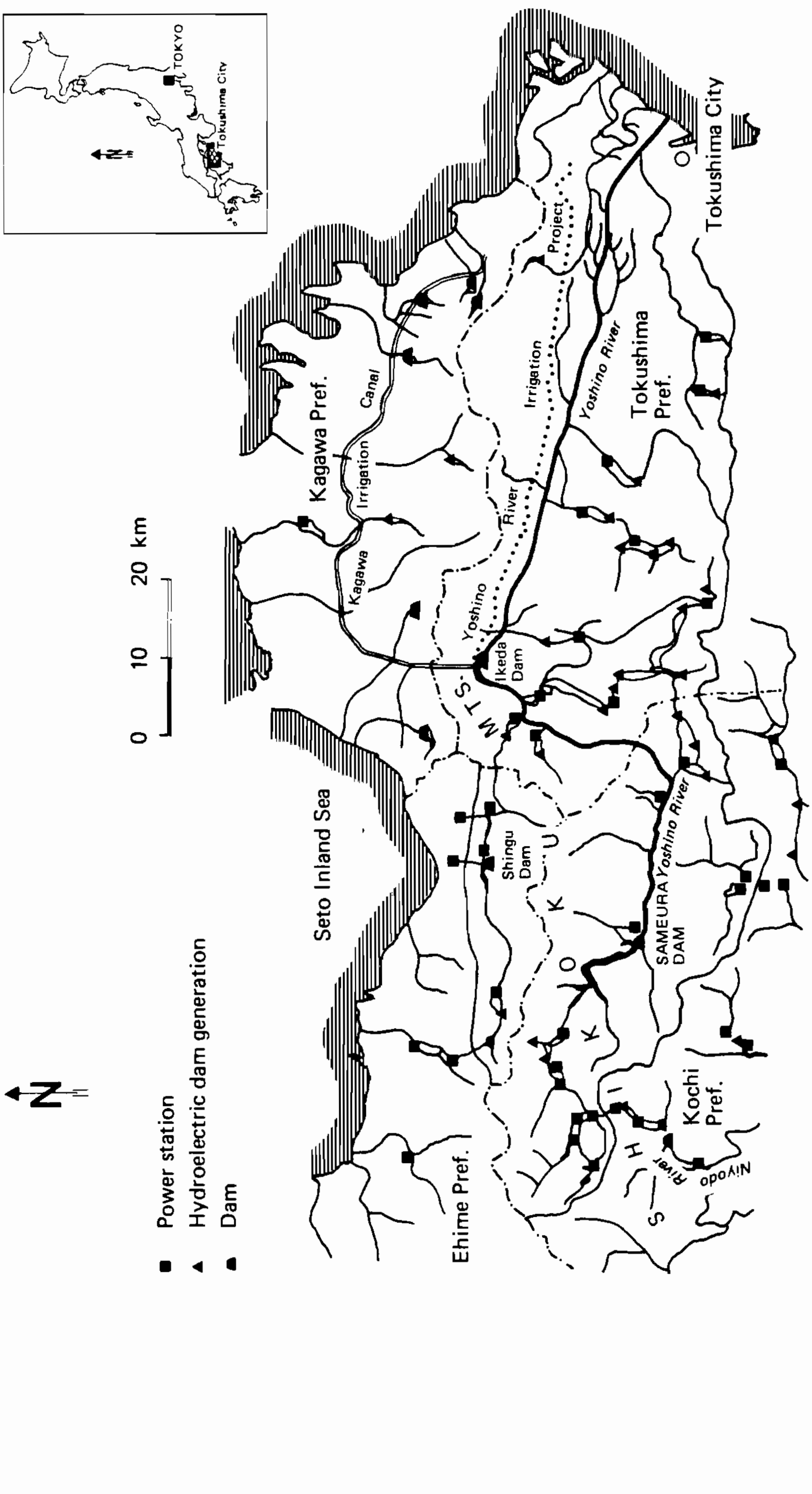
Power station
Hydroelectric dam generation
Dam
0 10 20 km
TOKYO
Tokushima City
Ehime Pref.
Seto Inland Sea
Kagawa Pref.
Kagawa
Irrigation
Canal
Tokushima Pref.
Kochi Pref.
Tokushima City
Yoshino River
Yoshino
River
Irrigation
Project
Ikeda Dam
Shingu Dam
SAMEURA DAM
Niyodo River
S H I K O K U M T S.

Figure 11.1

considered unfair by the others, and led to its abandonment in favor of the so-called separable costs remaining benefits (SCRB) method, which we describe in detail below. This method is still in use in Japan for most multipurpose reservoir developments, and has proved to be fairly viable – despite various inconsistencies and the ***ad hoc*** definitions it employs (for a detailed critique of the SCRB method see Young ***et al.***, 1982).

Our purpose is to show, through a detailed example, how the SCRB method is actually applied in Japan. We conclude with a discussion of some of the methodological complications encountered in allocating the costs of a multipurpose reservoir.

2 Case Study: The Sameura Dam Development

The Sameura Dam is on the Yoshino River, one of the major rivers in Japan, which originates in the western part of the Shikoku mountain range and runs from East to West through Shikoku Island. It then flows eastward as far as the border between Kochi and Tokushima prefectures, then turns northward to meet the mountain range, and runs eastward again before it flows into the sea at Tokushima City, some 200 km from its origin (see Figure 11.1). The Yoshino basin has an area of 3,750 km^2 (some 20 percent of the total area of the island) and covers parts of four prefectures: Kochi, Tokushima, Kagawa, and Ehime.

The Sameura Dam Development Project was started in 1963 and completed in 1970. It is a multipurpose reservoir project involving river management (flood control and minimum streamflow maintenance), irrigation, domestic and industrial water supplies, and hydroelectric power. A significant complication is that the users are distributed over four prefectures, which have substantially different interests in the project. Tokushima prefecture, which is located downstream and covers more than half of the basin area, looked for flood control as well as the development of water supplies for irrigators and cities. Tokushima City, the seat of the prefecture, had planned a large-scale industrial development and therefore needed increased water supplies. It had, however, been reluctant to allow the other prefectures to gain access to the water of the Yoshino River. Kochi prefecture, south of the mountain range, was eager to develop water for power, industrial and domestic uses, and claimed a special position as the prefecture commanding the river source. Ehime and Kagawa prefectures had suffered water shortages over a long period and were therefore interested in gaining access to the Yoshino River for industrial and domestic water supplies, power, and irrigation.

As the four prefectures began discussing the possible allocation of the project costs, it became clear that the mechanical application of the AJE method gave unsatisfactory results. Various proposals were developed and tested by the Ministry of Construction with regard to the requirements and claims from the prefectures. After much wrangling – including threats of withdrawal from the project – an alternative allocation method, combined partially with the SCRB method, emerged as a compromise solution. This marked the turning point in Japanese cost allocation practices in multipurpose reservoir development; since then the SCRB has been used exclusively in this field.

3 The SCRB Cost Allocation Procedure

The Specified Multipurpose Reservoir Development Act (as amended in 1967) provides that the SCRB method be employed in the allocation of costs of any multipurpose reservoir development covered by the Act – with the exception that in any case in which the straightforward application of the SCRB is found "unjustifiable" the so-called priority expenditure method (PEM) or some other (unspecified) method be employed. In fact the PEM method has rarely been used and the SCRB method is applied to almost all multipurpose reservoir developments.[1]

The SCRB method employs essentially four elements:

(1) "Direct" costs directly attributable to each specific use.
(2) The "individual" costs of providing each use by some other means.
(3) "Benefits" accruing from each use.
(4) The "separable" or "incremental" costs of including each use in the project last.

The calculation of each of these items proceeds by a definite method prescribed by the Act, and the terms are combined to give the allocation of costs. To illustrate how such a method works in practice, we will show how these prescriptions apply to the specific case of the Sameura Dam project.

3.1 *Storage Allocation Design*

The Act provides that the storage allocation design precede the cost allocation for multipurpose reservoir development. The storage allocation design physically assigns the storage capacity to different uses. This work is based on the hydrological and geophysical conditions of the construction site and on the amount of water claimed by each use (see Figure 11.2).

A distinction is made between flood and non-flood time. This means simply that the usable capacity of the reservoir is different during flood time and non-flood time. No specific assignment of storage is made to the power sector; it is assumed to share storage with the rest of the uses since its use of water is non-consumptive.

It is assumed in Figure 11.2 that out of a total of 85 million m^3 in flood time, 41.49 million m^3 is assigned to municipal (domestic) water supply, 5.26 million m^3 to industrial water supply, and 38.25 million m^3 to minimum streamflow maintenance. On top of this, 35 million m^3 of storage is reserved for accommodating the peak flow of a flood. During non-flood time the operation of the reservoir is adjusted to that of the increment of 35 million m^3, of which 17.085 million m^3 is allocated to municipal water supply, 2.165 million m^3 to industrial water supply and 15.75 million m^3 to minimum streamflow maintenance. Thus, the total storage amounts to 130 million m^3, including 10 million m^3 dead storage.

Next a storage cost curve is estimated as in Figure 11.3. This serves as the basis for calculating the various cost components of the SCRB method.

3.2 *Direct and Alternative Costs*

Only the power sector needs direct facilities attached to the reservoir in the form of generating and distribution equipment. These direct costs are calculated for the Sameura Project as $DC(\mathrm{P}) = 3.911$ billion yen.

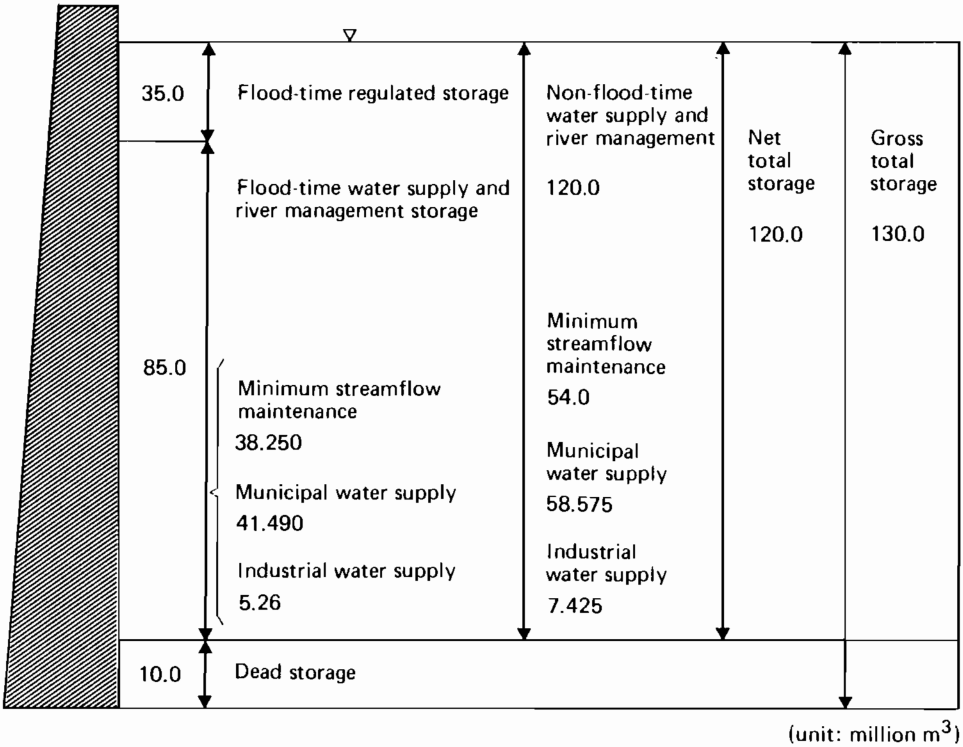

Figure 11.2

As defined by the Act, the individual alternative cost for a particular use is the estimated cost of an alternative facility offering the same level of service as that provided by the joint facility. The location of the alternative facility is not necessarily limited to the site where the joint facility is to be constructed. Common practices show, however, that the site is usually presumed to be the same. It is also true that the alternative facility is, in practice, assumed to be identical to a dam designed exclusively for the particular use.

The estimation of individual costs utilizes the "storage cost curve" of Figure 11.3 which is obtained by plotting the cost of constructing a reservoir, be it single- or multipurpose, against different designs. Based on the storage cost curve estimated for our example, individual costs for the different uses are calculated as follows:

(1) *River Management*. The storage needed: SR = 35.000 (for flood control) + 38.250 (for minimum streamflow maintenance) + 10.000 (for dead storage) = 83.250 million m^3. The individual cost: $C(R)$ = 12.600 billion yen.

(2) *Municipal Water Supply*. The storage needed: SM = 58.575 (for supply) + 10.000 (for dead storage) = 68.575 million m^3. The individual cost: $C(M)$ = 11.300 billion yen.

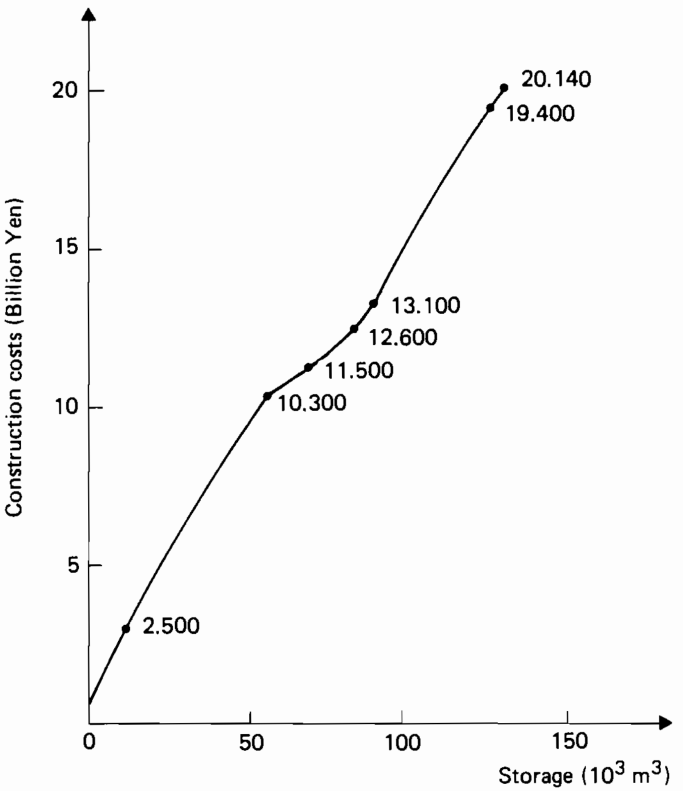

Figure 11.3

(3) *Industrial Water Supply*. The storage needed: SI = 7.425 (for supply) + 5.000 (for dead storage) = 12.425 million m^3. The individual cost: $C(I)$ = 2.500 billion yen.

(4) *Power*. Since the individual cost of power is generally far greater than its benefit (to be calculated later), its estimation is omitted. (The claim that power's benefits generally exceed its alternative costs is rather a crucial and tricky point which favors the power sector, but may not always hold, as we shall see later.)

3.3 *Benefits*

The incorporation of benefit terms as distinct from individual opportunity costs differentiates the SCRB method from the previous methods such as the AJE. It is defined by law that the benefit from a particular use. u, is to be calculated by the following formula:

$$B_u = \frac{U_u - O_u}{f_u} \quad ,$$

where U_u represents the estimated annual benefit (in money terms), O_u is the estimated annual cost of operation and maintenance, and f_u is the capital recovery factor. By definition the estimated annual utility U_u is the worth of the services, measured in monetary terms, which the joint project is designed to provide for a particular use, u. If there are direct facilities that exclusively serve a particular use, the worth of these facilities must also be included in the utility of the particular use. The estimated annual operation and maintenance costs for a particular use, O_u, merely refer to that portion of the costs of operating, maintaining and repairing (OMR) those elements of the joint facility associated with a particular use.

Note that OMR costs appear in the benefit terms and not in the cost terms. The implication is that the cost terms merely concern the costs of implementing fixed costs of the project and that the annual OMR costs that come out after project implementation are treated as benefits. According to law the annual OMR costs refer to those elements of the joint facility associated with a particular use. The difficulties with isolating those elements from the entire joint facility are supposed to be overcome by simply allocating the joint OMR costs to uses ***in proportion to the share of the joint project costs*** assigned to the respective uses. This, however, involves a circular definition: the calculus of OMR costs requires the knowledge of the share of the joint project costs assigned to the respective uses, which in turn is what is needed as input. This adds to the difficulty and inconsistency in bringing benefits directly into the mechanism of cost allocation as the SCRB method attempts to do.

The capital recovery factors f_u prescribed by law are listed in Table 11.1. Differences in the value of f_u for different uses reflect the differences in the interest rates available, durability (physical depreciation), property tax rates, etc.

Table 11.1 Different discount factors adopted for the SCRB procedure.

Use		*Interest rate r_u*	*Durability (lifetime)*	*Property tax rate*	*Capital recovery factor f_u*
Flood control		0.045	80 years	–	0.0464
Irrigation		0.055	45 years	–	0.0604
	Type A*	0.08	45 (0.1)†	0.014	0.0932
Power	Type B*				
	Type C*	0.07	45 (0.1)†	0.017	0.0785

* Type A – nine major private power enterprises; type B – the Dengenkaihatsu Co. Ltd. (a semi-governmental power development enterprise); type C – public power development enterprises.
† The figures in parentheses show the leftover rate.

The interest rate for construction investment, r_u, is introduced only when interest is to be paid out over the construction period for the accumulated investments related to a given use, u. Law prescribes that in this case the benefit of this use be discounted by this interest; that is, the benefit formula is modified:

$$B_u = (U_u - O_u) \,/\, [f_u(1 + r_u)] \quad .$$

The empirical formula for determining the value of r_u for power P is

$$r_P = 0.4\, i_P T \quad \text{for } i_P = 0.08 \text{ or } 0.07 \quad ,$$

where i_p is the interest rate used to obtain the capital recovery factor for power, and T is the construction period in years.

Another empirical formula is prescribed for use by irrigation. Letting A stand for irrigation, $r_A = 0.25 \times 0.42\, i_u T$, where $i_u = 0.065$, and 0.25 is a factor representing the average share of the farming sector of the costs of an irrigation project; the rest is borne partly by the local governments and partly by the Ministry of Agriculture.

The interest rate for construction investment is not taken into account for flood control, because such works are considered to be entirely of a public nature.

Municipal and industrial water supplies also do not rely on an interest rate. According to law, the benefit of a water supply, whether industrial or municipal, is assumed to be identical to its individual cost. Accordingly, the benefit of a water supply can be estimated with no regard to capital recovery or interest factor for construction investment.

This completes the development of formulas for estimating the benefits of different uses. Based on these, the benefits are computed as follows:

1. *River Management (R).*
   (a) Flood Control. The annual utility: U_{RF} = expected annual savings from decreased damages = 0.758 billion yen. The annual OMR costs, O_{RF} = 0.016 billion yen. The benefit $B_{RF} = (0.758 - 0.016) / 0.0464 = 15.991$ billion yen.
   (b) Minimum Streamflow Maintenance. The benefit B_{RS} = its alternative cost = 4.023 billion yen.

   It is simply postulated that the benefit of minimum streamflow maintenance is identical to its alternative cost, i.e. the individual cost for that particular purpose only. Combining (a) and (b), the benefit of river management, $B(R)$, is calculated to be 20.014 billion yen.

2. *Industrial and Municipal Water Supply (IM).* A joint alternative facility (a dam) is assumed for both industrial and municipal water supplies (this assumption is notably inconsistent with the former assumption that an alternative joint facility be contemplated only by all uses except for a particular use). The storage needed for the two used together is calculated as 58.575 (for municipal water supply) + 7.425 (for industrial water supply) + 10.000 (for dead storage) = 76.000 million m^3. Based on the storage cost curve (see Figure 11.3) the alternative cost for the two joint uses is identified as 12.000 billion yen.

   It is specifically prescribed by law that in the above case the joint alternative cost be divided as follows to obtain the individual benefits. The benefit of municipal water supply is obtained as $B(M) = 12.000 \times \frac{1}{3}$ (demand share ratio + storage share ratio + individual cost ratio)

$$= 12.000 \times \frac{1}{3}\left(\frac{14.2}{16.0} + \frac{58.575}{66.000} + \frac{11.300}{13.800}\right) = 10.380 \text{ billion yen} \quad .$$

In a similar manner, the benefit of industrial water supply, $B(I)$, is obtained as:

$$B(I) = 12.000 \times \frac{1}{3} \left[\frac{1.8}{16.0} + \frac{7.425}{66.000} + \frac{2.5000}{13.8000} \right] = 1.620 \text{ billion yen} \quad .$$

This implies that the joint benefit is divided in proportion to the average over the three ratios, i.e., demand share, storage share, and individual costs.

This is another example of an ***ad hoc*** assumption whose justification is highly debatable; it also contradicts the former assumption that the benefit of a water supply, either industrial or municipal, equals its individual cost. There seems to be no particular reason why estimates of the benefits of water supply should differ case by case. It amounts to an ***ad hoc*** manipulation to make the allocation more favorable for the water supply sector.

3. *Power (P).* The law prescribes that the annual utility of power generation, U_P, be estimated by the formula: U_P = effective kW × kW price + effective kWh × kWh price. In our example:

$$\begin{aligned} U_p &= 12{,}434 \text{ kW} \times 10{,}658 \text{ yen/kW} + 59{,}213{,}000 \text{ kWh} \times 4.84 \text{ yen/kWh} \\ &= 0.419 \text{ billion yen} \quad . \end{aligned}$$

The first term refers to the utility of power plant capacity; and the second to the utility of plant operation.

The annual operating and maintenance costs for power, O_P, are calculated to be 0.049 billion yen. Accordingly, the benefit of power is estimated as:

$$\begin{aligned} B_P &= (0.419 - 0.049) / [0.0785 \times (1 + 0.4 \times 0.07 \times 3)] \\ &= 4.353 \text{ billion yen} \quad . \end{aligned}$$

It should be observed that the interest rate for construction investment is employed to discount the benefit.

3.4 *Separable Costs*

It is defined by law that the separable cost is the amount that will be saved if a particular use is eliminated from the multipurpose project. The separable cost of a particular use, u, is calculated as follows.

(1) Specify that portion of storage which would have to be added if a particular use enters the joint project at the margin.
(2) Deduct that amount of storage from the total storage to obtain the remaining portion serving the rest of the uses.
(3) Identify the separable cost with the cost corresponding to the remaining portion of storage on the storage cost curve.

In our example separable costs are calculated as follows:

(1) *River Management (R).* The storage needed for the rest of uses: S_{PIM} = 130.000 (gross total) − 35.000 (for flood control) − 38.250 (for minimum streamflow maintenance) = 56.750 million m^3. The construction cost of the joint project without river management, $C(PIM)$ = 10.300 billion yen. The separable cost, $SC(R) = C(RPIM) - C(PIM) = 20.140 - 10.300 = 9.840$ billion yen.

(2) *Municipal Water Supply (M).* The storage needed for the remaining uses: S_{RPI} = 130.000 (gross total) − 41.490 (for municipal water supply) = 88.510 million m^3. The separable cost, $SC(M) = C(RPIM) - C(RPI) = 20.140 - 13.10 = 7.04$ billion yen.

(3) *Industrial Water Supply (I).* The storage needed for the rest of uses: S_{RPM} = 130.000 (gross total) − 5.260 (industrial water supply) = 124.740 million m^3. The construction cost of the joint project without industrial water supply, $C(RPM)$ = 19.400 billion yen. The separable cost, $SC(I) = C(RPIM) - C(RPM) = 20.140 - 19.400 = 0.740$ billion yen.

(4) *Power (P).* There is often some difficulty in specifying the separable cost for power since inclusion of the power sector into the joint project as the last participant would not necessarily require any substantial increase in storage if the amount of storage designed for the other uses and its design head are larger than or as large as the hydropower plant necessitates in generating the planned power output.

This is particularly true of our example. Nevertheless, the present procedure adheres to bringing in the "incremental storage", which may not have a reasonable physical implication. The pitfall is in basing the estimation only upon the storage allocation, while disregarding the fact that power's water use is not consumption, but mostly utilizes the kinetic energy of falling water. The legislation attempts to overcome this difficulty by axiomatically specifying the incremental storage attributable to power as follows.

The amount of water, q, as defined below is regarded as the incremental storage needed exclusively for power:

$$q = 3{,}600t\,(q_{\text{max}} - q_{\text{mean}})\ \text{m}^3 \quad ,$$

where

q_{max} is the peak-time water discharge (m^3/s),
q_{mean} is the average water discharge (m^3/s), and
t is the duration of peak-time (hr).

The cost of the joint project without power, C(RIM) is estimated by locating the cost versus total storage minus q on the storage cost curve. The separable cost of power, $SC(P)$ is the difference between the cost of the joint project, $C(RPIM)$ and $C(RIM)$. By applying the formula to our example, we have

$$q = 3{,}600 \times 6 \times (10.0 - 2.35) = 0.170 \text{ million m}^3 \quad .$$

The storage needed for the rest of uses is 129.83 million m^3. Accordingly, the construction of the joint project without power, $C(RIM)$ = 20.116. Therefore the separable cost of power, $SC(P) = C(RPIM) - C(RIM) = 0.024$ billion yen.

This completes the calculation of all the parameters needed. Application of the SCRB formula yields the cost allocation as shown in Table 11.2.

Table 11.2 Cost allocation (billion yen) based on SCRB.

	Use				
Calculation items	*River management*	*Municipal water supply*	*Industrial water supply*	*Power*	*Subtotal*
Alternative cost (C)	12.6	11.3	2.5	–	–
Justifiable cost (B)	20.014	10.38	1.62	4.353	–
Min (B,C)	12.6	10.38	1.62	4.353	–
Direct cost (DC)	0	0	0	3.911	–
Min (B,C)–DC	12.6	10.38	0	0.442	–
Separable cost (SC)	9.84	7.04	0.74	0.024	17.644
Remaining benefit (RB), Min (B,C)–DC–SC	2.76	3.34	0.88	0.418	7.398
Ratio thereof %	37.3	45.1	11.9	5.7	100.0
Allocation of nonseparable costs	0.931	1.126	0.297	0.142	2.496
Net allocation	10.771	8.166	1.037	0.166	20.14
Ratio thereof %	53.5	40.5	5.1	0.8	100.0
Gross allocation	10.771	8.166	1.037	4.077	24.05
Ratio thereof %	44.8	33.9	4.3	17.0	100.0

4 Problems in Cost Allocation: Conclusion

To sum up, based upon the above discussion an attempt will be made to specify the major problems of cost allocation in the field of Japanese multipurpose reservoir development and to examine how they should be approached.

The purpose of this example is to illustrate how an apparently simple method such as the SCRB involves a host of complications and *ad hoc* assumptions in its implementation. In particular, the manner of including benefits opens the door to an array of possible manipulations. Specifically, for the river management sector (flood control and minimum streamflow maintenance) and domestic and industrial water supplies, it is assumed that the "benefit" (the minimum of "justifiable" cost and alternative cost) coincides with the alternative costs, since the justifiable cost is estimated to be higher than its alternative cost for the river sector, and both terms are designed to be equal for the water supply sectors. On the other hand, for the irrigation and power sectors the situation is reversed and alternative costs are assumed to exceed justifiable costs.

There seem to be many causes involved. One of the primary ones is that the river management sector is not a profit-making entity, whereas the irrigation and power sectors are. Any profit-making enterprise is naturally inclined to underestimate its justifiable expenditure for participating in a particular project way below its alternative cost, as long as its cost sharing is determined by the minimum of either of the two terms. On the other hand, a public enterprise which, by nature, should not and cannot make profits tends to be rather generous about cost sharing compared with profit-making enterprises. This is particularly true when public financing conditions are not tight. As is always the case with public works, a project may not be funded if its benefits are expected to be below its costs, and this naturally leads to the outcome that for any justifiable project its minimum term should coincide with its alternative cost.

In the case of the Sameura Dam development project, the existence of subsidies by the national government compounds the problem of cost allocation. That is to say, an allocation to a use based on an officially adopted method is not necessarily what it has to pay in all. For instance, the domestic water supply sector gets one-half to one-third of its (allocated) cost subsidized by the national government. An exception is the autonomous Okinawa Island, which had been under US control after the war and was designated by the Japanese government as a special region for promoting regional developments with top priority when it was returned to Japan. All of its allocated costs are subsidized by the national government; the industrial sector commonly gets twenty to forty percent of its allocated costs subsidized. As far as the irrigation sector is concerned, its subsidy ratio is ninety percent, which means that it is charged only ten percent of the totals.

It is very common that the entire cost allocation is structured in hierarchical form. Namely, at the first level of cost allocation the total costs are allocated to each sector. Then at the second level the costs allocated at the first level are further allocated to subsectors. This is followed by the third level at which the costs allocated to each subsector are again allocated to sub-subsectors. This continues until the lowest level of cost allocation is reached. There seems to be a tendency that, as we come down to lower levels of cost allocation, the allocation methods become simpler and more straightforward. For instance, we start off with the SCRB method and terminate the process by splitting the costs in proportion to the amount of water withdrawals. This suggests that there is a practical need for introducing the notion of hierarchical levels in discussing any type of cost allocation.

Notes

1. An exception is the Kuzuryu Dam construction project in the early 1960s. A power company had started the construction of the dam, when the flood sector joined the project, which required the structure to be elevated by 4 meters. It was reasoned that since the power company carried the original costs on its own, priority should be given to the power sector. The flood sector (the latecomer) was requested to pay the additional costs.
2. The first term is called "kW value" and the second term "kWh value". It should be noted that both kW value and kWh value are measured in terms of the opportunity cost that a thermal plant would have to spend if the same service was produced by an alternative facility. That is, the kW value refers to the annual fixed costs of a thermal plant with a given kW capacity and the kWh value to the annual variable

costs of the thermal plant with a given kWh operation ability. What is tricky with this assessment is that no evaluation is made of what really characterizes the value of service of a hydroelectric power plant, which is far more responsive and adaptive to fluctuations in power demand than is a thermal plant. Though the construction of a hydropower plant is several times more expensive than that of a thermal plant, its operation is far less expensive. Moreover, the lifetime of a hydropower plant is generally three times that of a thermal plant and it may even be the case that a hydropower plant continues to be used even after its design lifetime is over. Additionally, in the face of ever-increasing prices for fuel oil, hydropower is becoming increasingly cost-effective. In light of these considerations, it may be said that the assessment formula for the benefits of a hydropower plant, as prescribed by law, grossly underestimates the worth of its service, thus almost always resulting in estimated benefits that are far less than its individual cost. This, in turn favors the power sector. The ***ad hoc*** way in which benefits are estimated (and in a sense manipulated) immediately casts doubt upon the SCRB method.

Bibliography

Archives and Documents Section (1981) *Construction Compendium* (Tokyo Horei Shuppan Publishing Co.; in Japanese).

Hanayama, K. and Fuse, T. (1977) *The City and Water Resources – Political Economics of Water* (Kajima Shuppan Co.; in Japanese).

Inter-Agency Committee on Water Resources (1958) *Proposed Practices for Economic Analysis of River Basin Projects* (Washington, DC: US Government Printing Office).

James, L.D., and Lee, R.R. (1971) *Economics of Water Resources Planning* (New York: McGraw-Hill).

Kazuo Hori (1971) *Multipurpose Reservoir – Planning and Investigation* (Tokyo, Zen'nihonkensetsugijutsu-kyokai Publishing Co.; in Japanese).

Niizawa, K. (1962) *Theory of Hydraulic Controls of Rivers* (Iwanami Publishing Co.; in Japanese).

River Bureau (1977) *Construction of Multipurpose Reservoirs* (Tokyo: Japanese Ministry of Construction; in Japanese).

River Bureau (1980) *Multipurpose Reservoirs of Japan – Planning* (Tokyo: Japanese Ministry of Construction; in Japanese).

Young, H.P., Okada, N., and Hashimoto, T. (1982) Cost allocation in water resources development. *Water Resources Research* 18(3): 463–475.

ISBN: 0 444 87863 7